Lodz University of Technology

Institute of Applied Computer Science

Proceedings
of

INTERNATIONAL INTERDISCIPLINARY PHD WORKSHOP

IIPHDW 2017

SEPTEMBER 9-11, 2017
ŁÓDŹ, POLAND

https://iiphdw2017.p.lodz.pl/

PROCEEDINGS

International Interdisciplinary PHD Workshop 2017

Organised by:

Lodz University of Technology

Institute of Applied Computer Science

In cooperation with:

	Lublin University of Technology **Institute of Electronics and Information Technology**
	Research & Development Centre Netrix S.A.
	Brno University of Technology **Faculty of Electrical Engineering and Communication**
	Electrotechnical Institute in Warsaw

September 9-11, 2017

Łódź, Poland

International Interdisciplinary PhD Workshop 2017
Proceedings

Editor: Robert Banasiak, Radosław Wajman, Jacek Kucharski, Dominik Sankowski

Publisher: Lodz University of Technology Press
Wólczańska 223, 90-924 Łódź, Poland

ISBN **978-83-7283-858-2**

International Interdisciplinary PhD Workshop 2017

Conference Chairman

Robert Banasiak, Lodz University of Technology, Poland

Honorary Chairman

Dominik Sankowski, Lodz University of Technology, Poland
Krzysztof Kluszczyński, Silesian University of Technology, Poland
Jan Sikora, Lublin University of Technology, Poland

Vice-chairman

Tomasz Rymarczyk, Netrix S.A. / WSEI
Paweł Komada, Lublin University of Technology, Poland
Adam Pelikant, Lodz University of Technology, Poland

Technical Program Committee

Jacek Kucharski, Lodz University of Technology, Poland
Radosław Wajman, Lodz University of Technology, Poland
Aleksandra Kowalska, Lodz University of Technology, Poland

Scientific Committee

Adamiak K., University of Western Ontario, Canada
Adamkiewicz P., NETRIX S.A., Poland
Ahrens A., Hochschule Wismar, University of Technology, Business and Design, Germany
Babout L., Lodz University of Technology, Poland
Banasiak R., Lodz University of Technology, Poland
Bieniecki W., Lodz University of Technology, Poland
Binkowski T., Rzeszow University of Technology, Poland
Brauner H., Technische Universität Ilmenau, Germany
Brzeziński D., Lodz University of Technology, Poland
Butryło B., Bialystok University of Technology, Poland
Chaniecki Z., Lodz University of Technology, Poland
Cichocki A., Riken Brain Institute, Japan
Cieplak T., Lublin University of Technology, Poland
Drałus G., Rzeszow University of Technology, Poland
Drexler P., Brno University of Technology, Czech Republic
Fabijańska A., Young Scientist Academy PAN, Lodz University of Technology, Poland
Fiala P., Brno University of Technology, Czech Republic
Fjeld M., Chalmers Univ. of Technology, Sweden
Frączyk A., Lodz University of Technology, Poland
Grabowski S., Lodz University of Technology, Poland
Grudzień K., Lodz University of Technology, Poland
Hotra Z., Lviv Polytechnic National University, Ukraine
Jackowska-Strumiłło L., Lodz University of Technology, Poland
Kapusta P., Lodz University of Technology, Poland
Kciuk M., Silesian University of Technology, Poland
Kłosowski G., Lublin University of Technology, Poland
Komada P., Lublin University of Technology, Poland
Korbel P., Lodz University of Technology, Poland
Kotyra A., Lublin University of Technology, Poland
Kowalski T., Lodz University of Technology, Poland
Kucharski J., Lodz University of Technology, Poland
Majchrowicz M., Lodz University of Technology, Poland
Malska W., Rzeszow University of Technology, Poland
Mikulka J., Brno University of Technology, Czech Republic
Mosorow W., Lodz University of Technology, Poland
Mylvaganam S., University of Southeast Norway, Norway
Nowakowski J., Lodz University of Technology, Poland
Ostalczyk P., Lodz University of Technology, Poland
Pańczyk M., Lublin University of Technology, Poland
Pilch Z., Silesian University of Technology, Poland
Polakowski K., Warsaw University of Technology, Poland
Romanowski A., Lodz University of Technology, Poland
Roshani G., Kermanshah University of Technology, Iran
Rymarczyk T., Netrix S.A., Poland

Rząsa M., Opole University of Technology, Poland
Sankowski D., Lodz University of Technology, Poland
Shamma F., Kermanshah University of Technology, Iran
Sierszeń A., Lodz University of Technology, Poland
Sikora J., Lublin University of Technology, Poland
Sobczyński D., Rzeszow University of Technology, Poland
Sokołowski J., Université de Lorraine, France
Sowa M., Silesian University of Technology (Gliwice), Poland
Strumiłło P., Lodz University of Technology, Poland
Stryczewska H., Lublin University of Technology, Poland
Strzelecki M., Lodz University of Technology, Poland
Sykulski J., University of Southampton, United Kingdom
Szeląg A., Warsaw University of Technology, Poland
Ślot K., Lodz University of Technology, Poland
Urbanek P., Lodz University of Technology, Poland
Wac-Włodarczyk A., Lublin University of Technology, Poland
Wajman R., Lodz University of Technology, Poland
Ziółkowski M., Technische Universität Ilmenau, Germany

International Interdisciplinary PHD Workshop

IIPhDW 2017

September 9-11 Łódź, Poland

Contents

Mateusz Krzysztoń[1,2]

The Nodes Cooperation for Maintaining Connectivity of Mobile Ad Hoc Network during Phenomena Cloud Monitoring

MANETs are increasingly applied in case of a disaster to support an emergency team. During a rescue operation the external communication infrastructure is often overloaded or even destroyed. Thus, it is necessary to ensure communication between nodes to provide a possibility for exchanging information about an environment. In this paper a new method for maintaining connectivity of the network is presented. The main advantage of the method is reduced impact on a performance of MANET for monitoring the environment. The method was verified in scenario of a heavy gas leakage simulated with the SLAB simulator.

1. Introduction

Mobile ad hoc network (MANET) is a wireless network characterized by decentralized structure. The network is composed of autonomous mobile platforms equipped with a wireless communication device and sensors. The MANET can selforganize in decentralized manner to adopt to the environment in which it was deployed. Each network node can freely change its location and role played within the network depending on the locations of other nodes and its knowledge about the environment.

One of the key areas of MANET application is to support a rescue team coping with a disaster situation. The examples of such situations are toxic heavy gas leakage, oil spill, flood or forest fire. In all those situations the object of interest can be defined as an object covering significant area, characterized by nondeterministic, dynamic variations of shape, size, speed, and direction of motion along multiple axes [7]. Such object is called *phenomena cloud*. In case of one of the aforementioned disasters the rescue team is interested in current size, shape and movement direction of the phenomena cloud. Such knowledge can be used to surround and neutralize the disaster (in case of a heavy gas cloud or oil spill) or to predict the area that should be evacuated first (in case of a forest fire).

During a disaster situation fast information exchange within network is required. At the same time the external communication infrastructure can be damaged or overloaded due to the disaster. Thus, providing an internal communication between each pair of nodes in the network is necessary. In the MANET each device can communicate with the other devices that are located within its transmission range or with other devices using intermediate nodes to relay messages hop by hop (*multi-hop* communication). Due to the limited transmission range of devices coordination of devices motion is required to maintain connectivity within the network. At the same time requirement to maintain connectivity should not affect the quality of task realization by MANET significantly.

[1]Research Academic Computer Network (NASK), Kolska 12, 01-045 Warsaw, Poland, e-mail: mateuszkr@nask.pl

[2]Institute of Control and Computation Engineering, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warsaw, Poland, e-mail: mateusz.krzyszton@gmail.com

The connectivity maintenance is the subject of a lot of research. In [2] is presented a method for the network reconfiguration if the target topology is known. During the whole process connectivity is guaranteed. An interesting method based on constraints is presented in [8]. The method allow to set a lower and an upper limit of the number of communication links that each node takes part in. Decentralized scheme for detecting redundant links by each node based only on local information was introduced in [1]. However, with only local information identification of all redundant links is impossible. Another decentralized algorithm for detecting redundant links was described in [4]. However, an additional auction-based mechanism (involving communication between all nodes) is required to decide if the given redundant link can be deleted.

To the best of the author's knowledge none of the connectivity maintenance methods proposed in the literature takes into account its negative influence on MANET performance in terms of task realization quality. In this work a new method for connectivity maintenance is proposed. The main novelty of the method is additional cooperation between nodes that reduces aforementioned negative impact on MANET performance. The proposed cooperation scheme can be easily adapted for the most of other methods for connectivity maintenance. The proposed method is verified in a task of heavy gas cloud monitoring. The research is conducted with the simulator of mobile ad hoc network MobASim [6].

The article is organized as follows. First, a scheme for constructing MANET for a phenomena boundary detection is introduced and the connectivity maintenance problem is defined. Then, the method for connectivity maintenance is extended with cooperation. The method is verified in a scenario of a heavy gas dispersion. The paper is concluded in section 5.

2. Problem Formulation

In this work the MANET network is comprised of a set V of N unmanned devices D_i, $i = 1,\ldots,N$. Each of the devices is equipped with a GPS receiver, a proper punctual phenomena sensor and a radio transceivers that enables communication between devices. The location of D_i is $o_i = [x_i, y_i]$. The network in time t can be seen as a graph $G(t) = (V, E(t))$, where $E(t)$ is a set of bidirectional communication links (D_i, D_j), $D_i, D_j \in V$, $i \neq j$ that exist in time t. The link exists between D_i and D_j if the Euclidean distance between those nodes d_j^i in time t is smaller than the transmission range r_t. In this work it is assumed that communication range of each node D_i is circle with a radius r_t centered in o_i.

Two nodes D_i and D_j can communicate with each other in time t if there exists path (single or multi hop connection) composed of links from $E(t)$ starting at D_i and ending at D_j. If every pair of nodes can communicate with each other in time t then the network (and the graph $G(t)$) is *connected*. In this work it is assumed that in the time step t_0 the network is connected. In practice the assumption is easy to satisfy.

The strategy for constructing the MANET to cover a boundary of the phenomena cloud is described in details in [3]. The aim of the strategy is to distribute network nodes evenly on the boundary. In this approach the network is divided into K separated clusters V_k, $k = 1,\ldots,K$ of devices:

$$V_1 \cup V_2 \cup \ldots \cup V_K = V;\ V_1 \cap V_2 \cap \ldots \cap V_K = \emptyset. \tag{1}$$

The device $H_k \in V_k$ is assigned a role of the kth cluster head, $k = 1,\ldots,K$. One of the cluster heads H is selected to be a head of the whole network, $H \in H_1,\ldots,H_K$. The communication links can be establish only within cluster and between cluster heads. Thus, the kth cluster devices create subnetwork $G_k(t) = (V_k, E_k(t))$, where $E_k(t)$ is a set of bidirectional communication

links (D_i, D_j), $D_i, D_j \in V_k, i \neq j$ existing in time t. The set of cluster heads $V_H = \{H_1, \ldots, H_K\}$ is treated as a separate cluster, which analogously creates subnetwork $G_H(t) = (V_H, E_H(t))$, where $E_H(t)$ is set of links between cluster heads in time t. Thus, the problem of maintaining connectivity of the whole network is equivalent to the problem of maintaining connectivity of all subnetworks G_k, $k \in 1, \ldots, K$ and the subnetwork G_H.

The method described in [3] guarantees connectivity of the whole network, but does not address negative effect on the quality of the boundary detection caused by this requirement. In this work the extended version of the connectivity maintenance method is presented. The proposed method can be easily adopted to maintain connectivity in any MANET regardless of the task being performed and network clusterization.

3. Connectivity Maintenance with Cooperation method

The CMC (Connectivity Maintenance with Cooperation) method is adjusted to the PFM mobility model, described in details in [6]. The PFM is an example of an artificial potential based mobility model. In this model the trajectory of a device movement is a result of repulsive and attractive potentials. The CMC method modifies list of potentials of each device to maintain connectivity within the network. The concept of CMC method can be easily adjusted to other mobility models. The PFM mobility model was used in aforementioned work on the boundary detection as well [3].

The CMC method is based on identifying critical communication links in a network or, if clusterization is applied, a subnetwork. A critical link is a link that is essential for the network to be connected. Hence, the devices that create the critical link are attracted to each other. However, thanks to the additional cooperation mechanism, some of nodes are released from the link maintenance duty and may continue performing their task.

Connectivity maintenance

The algorithm for links that need to be maintained (i.e. critical links) identification is based on works [1] and [4] and consists of two phases. In the first phase links that are candidates for critical ones are identified by each node D_i. The node D_i considers only links that it takes part in. To decide which links are potentially critical two regions within the communication range of D_i are defined (Fig. 1a) — safe (cov_s^i) and critical (cov_c^i):

- cov_s^i denotes a disc of a radius r_t^s, $r_t^s < r_t$ centered at o_i.
- cov_c^i denotes a set of points o_j with distance d_j^i to o_i satisfying the condition $r_t^s \leq d_j^i \leq r_t$.

All links of D_i with nodes located in the region cov_c^i are considered to be at risk of breaking due to nodes movement. Thus, the decision has to be made which of those links has to be maintain as critical for the network connectivity.

In the second phase each node $D_i \in V_m$ sends to the head of its cluster H_m a set of nodes located in its critical region cov_c^i in time t — a set $C_i(t)$. The H_m responds with updated set $C_i^{'}(t)$ containing nodes with which communication link has to be preserved to maintain the network connected. The cluster head H_m decides whether the link between the node D_i and $D_j \in C_i(t)$ can be removed by simulation of the link (D_i, D_j) removal from the graph $G_m(t)$. For the resultant graph $G_m^{'}(t)$ the Laplacian matrix $L^*(t)$ is determined: $L^*(t) = \Delta^*(t) - A^*(t)$,

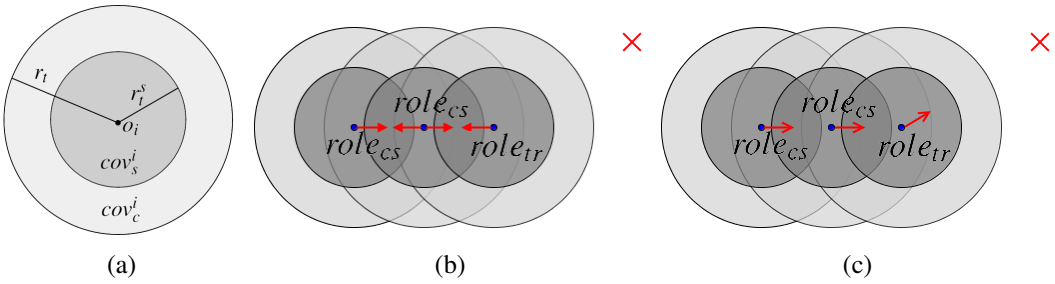


Figure 1: (a) cov_s^i and cov_c^i regions of D_i; (b) a direction of nodes motion if cooperation mechanism is disabled and (c) enabled — due to cooperation mechanism the node with $role_{tr}$ can move towards its target location (marked with a red cross).

where $A^*(t) = (a_{ij}^*(t))$ denotes the adjacency matrix of the graph $G_m^{'}(t)$. $a_{ij}^*(t) = 1$ if $(D_i, D_j) \in E_m(t)$ and 0 otherwise, $\Delta^*(t) = diag(\sum_{j=1}^{|V_m|} a_{ij}^*(t))$. The eigenvalues λ_i^* of the matrix $L^*(t)$ are calculated and sorted in an ascending order: $\lambda_1^* \leq \lambda_2^* \leq \ldots \leq \lambda_{|V_m|}^*$. $G_m^{'}(t)$ is connected if $\lambda_2^* \geq 0$ — link (D_i, D_j) is redundant and the node D_j can be removed from the set $C_i(t)$ [4].

The node D_i has to preserve communication links with all nodes in $C_i^{'}(t)$ to maintain connectivity within its cluster. In the aforementioned work [3] all nodes $D_j \in C_i^{'}(t)$ were attracted to the node D_i and forced to move to its safe region cov_s^i. It easy to observe that if $D_j \in C_i^{'}(t)$ then $D_i \in C_j^{'}(t)$. Hence, both nodes D_i and D_j get closer to each other instead of performing their task, which obviously reduces efficiency of the network. In the next subsection this efficiency problem is addressed by introducing an additional cooperation between nodes.

Nodes Cooperation

The aim of a cooperation between nodes is to release some nodes from the duty to preserve their critical communication links. To implement the cooperation each node is given one of two roles: $role_{tr}$ and $role_{cs}$. The main goal of a node with the role $role_{tr}$ is to perform the task assigned to the MANET (e.g. cloud monitoring), whereas the most important goal of the node with the role $role_{cs}$ is to maintain connectivity of the network. The node with the role $role_{cs}$ can perform other tasks as long as it is not necessary to preserve some communication links.

Three types of messages are defined: *NEEDS_HELP*, *CAN_HELP* and *CANNOT_HELP*. The *CAN_HELP* and *CANNOT_HELP* messages are responses for the message *NEEDS_HELP*. To the message *NEEDS_HELP* list *cooperation_nodes* is attached. The list contains list of nodes which already take part in cooperation. The node D_j can response for the message *NEEDS_HELP* from the node D_i with the message *CAN_HELP* if all following conditions are satisfied:

- has role $role_{cs}$;
- none of the nodes from the set $C_j^{'}(t)$ (except D_i) is on the *cooperation_nodes* list;
- does not have any other critical node than the node D_i or all its critical nodes D_k other than D_i ($D_k \in C_j^{'}(t), k \neq i$) respond to message *NEEDS_HELP* sent by node D_j with the message *CAN_HELP*.

Otherwise the node D_i responses with the message $CANNOT\_HELP$.

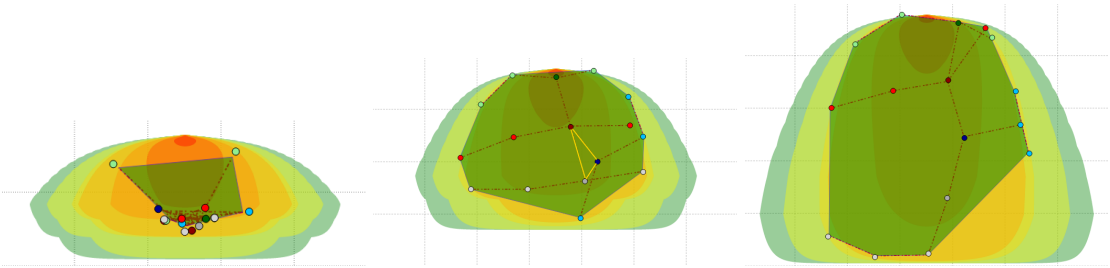
In the cooperation phase each node D_i with non-empty set $C_i^{'}(t)$ sends to each node $D_j \in C_i^{'}(t)$ message $NEEDS\_HELP$. Each node D_j that receives message $NEEDS\_HELP$ add to the $cooperation\_nodes$ list its id j and sends this message further to all nodes from $C_j^{'}(t)$. The node D_i that receives the response message CAN_HELP from the node D_j removes the node D_j from the set $C_i^{'}(t)$ of its critical nodes. Otherwise, if the message $CANNOT\_HELP$ is received the node D_i cannot remove that node from the set $C_i^{'}(t)$ and is responsible for maintaining the communication link with that node. As the result of the cooperation process each node D_i has the updated set $C_i^{'}(t)$. If the set is empty then the node can continue performing the task that is consistent with the task of the whole network. Otherwise, the priority of node is to maintain communication links with all its critical nodes.

In case of $D_j \in C_i^{'}(t)$ (the connection link (D_i, D_j) is critical one and D_i is responsible for that link maintenance) the node D_i is attracted to that node (i.e. by potential source located in o_j). Hence, according to the PFM model, a new position of the node D_i is calculated solving the optimization problem $\min_{o_i} \sum_{D_j \in C_i^{'}(t)} \left(\frac{\overline{d_j^i}}{d_j^i} - 1 \right)^2$, where $\overline{d_j^i} \leq r_t^s$ is an expected distance between nodes D_i and D_j. As a result of solving the optimization problem the nodes move to safe regions of their critical nodes. The directions of nodes motion in an exemplary situation with disabled and enabled cooperation mechanism is compared on Fig 1b and Fig 1c. The list $cooperation\_nodes$ and corresponding condition for replying with CAN_HELP were introduced to avoid a cyclic cooperation.

4. Experimental verification

The performance of the CMC method was verified through a simulation experiment. The task was to create a connected sensing MANET for monitoring uncontrolled instantaneous release of chlorine gas. During the whole deployment process the connectivity of the network has to be maintained. Additionally, the CMC method should not prevent the nodes from distributing evenly on the boundary.

The simulation parameters are as follows: the chlorine gas source area was $0.1m^2$, the mass source rate was $33.33\frac{kg}{s}$, the temperature of an ambient environment was $276K$ and the wind speed was $3\frac{m}{s}$. The dispersion of the cloud was simulated using the SLAB simulator [5]. The sensing network was built by $N = 16$ devices divided into $K = 4$ clusters. In order to maintain



(a) (b) (c)

Figure 2: Topology of the sensing network:(a) t = 65, (b) t = 90 and (c) t = 135. Nodes of the same cluster are marked with the same color, the cluster head is marked with darker hue. With red lines existing links are presented.

connectivity of the whole network the connectivity of each subnetwork G_k, $k \in 1,\ldots,K$ and the subnetwork G_H was maintained with the CMC method. To perform cooperation between nodes the role $role_{cs}$ was assigned to all cluster heads H_k, $k \in 1,\ldots,K$. Other nodes had the role $role_{tr}$.

Fig. 2 shows the process of forming the network topology for cloud boundary monitoring. At time $t = 65$ nodes arrive to the area covered with gas (Fig. 2a) — the connectivity is maintained and multiple redundant communication links exist as network covers little area. With the network expansion some redundant links are identified correctly and removed (Fig. 2b, $t = 90$). At that moment redundant link exists only within the subnetwork G_H — removing one of the three links marked with yellow color (Fig. 2b) is safe in terms of connectivity maintenance. The redundant link is removed correctly as clusters spread over the area covered with gas (Fig. 2c).

Thanks to introducing cooperation between nodes each cluster head follows its cluster. As the result nodes are placed on the boundary of a heavy gas cloud quite evenly. The subnetwork G_H serves as internal communication network and each pair of nodes can communicate. During the whole simulation process the connectivity within the network was maintained.

5. Conclusions

MANETs can support a rescue team efficiently in tracking a phenomena cloud. Thanks to internal communication they are independent on external communication systems. In this paper novel method for connectivity maintenance with cooperation was introduced. Due to cooperation a part of network serves as communication layer, whereas the other part can perform task efficiently. The CMC method can be used to maintain connectivity of any MANET, regardless of the application. The quality of the method should be evaluated in other scenarios.

References

1. P. D. Hung, M.-T. Pham, T. Q. Vinh, and T. D. Ngo. Self-deployment strategy for a swarm of robots with global network preservation to assist rescuers in hazardous environments. In *Robotics and Biomimetics (ROBIO), 2014 IEEE International Conference on*, pages 2655–2660, Dec 2014.

2. Z. Kan, L. Navaravong, J. M. Shea, E. L. Pasiliao, and W. E. Dixon. Graph matching-based formation reconfiguration of networked agents with connectivity maintenance. *Control of Network Systems, IEEE Transactions on*, 2(1):24–35, 2015.

3. M. Krzyszton and E. Niewiadomska-Szynkiewicz. Heavy gas cloud boundary estimation and tracking using mobile sensors. *Journal of Telecommunications and Information Technology*, (3):38, 2016.

4. N. Michael, M. M. Zavlanos, V. Kumar, and G. J. Pappas. Maintaining connectivity in mobile robot networks. In *Experimental Robotics*, pages 117–126. Springer, 2009.

5. D. Morgan Jr, L. K. Morris, and D. L. Ermak. Slab: a time-dependent computer model for the dispersion of heavy gases released in the atmosphere. Technical report, Lawrence Livermore National Lab., CA (USA), 1983.

6. A. Sikora and E. Niewiadomska-Szynkiewicz. Mobility model for self-organizing and cooperative msn and manet systems. *Sensors & Transducers*, 14(2):164, 2012.

7. M. T. Thai, R. Tiwari, R. Bose, and A. Helal. On detection and tracking of variant phenomena clouds. *ACM Trans. Sen. Netw.*, 10(2):34:1–34:33, Jan. 2014.

8. R. Williams and G. Sukhatme. Constrained interaction and coordination in proximity-limited multiagent systems. *Robotics, IEEE Transactions on*, 29(4):930–944, 2013.

Michał MAJCHROWICZ[1], Paweł KAPUSTA[2]

Rooting Smart TVs by exploiting design flaws in application framework

Each year an increasing number of devices gain Internet access capabilities, which can be convenient, but is also a big risk if done poorly. When a device is constantly connected to the Internet, security becomes a very important concern. This is especially true for devices, that were considered safe until now, as the security layer is often underdeveloped or even considered an afterthought. In this article we will show that is possible, and relatively easy, to circumvent security mechanisms present in Samsung Smart TVs to gain root privileges and unrestricted access to underlying file system.

1. Introduction

We live in the era of Internet of Things, many of the devices we use everyday are connected to the global network. Moreover those devices have access to our personal data so any breaches of their security should have high significance for their users. About 5-10 years ago we mostly cared about the security of our PCs, the safety of mobile devices was a new thing. Nowadays many people are aware of viruses on these platforms and know that their personal data, banking details etc. are at risk. However people are still not informed about the dangers of malware or hackers taking control of their embedded devices. Smart TVs are a good example of a device, which has Internet access, full operating system and are even often equipped with a camera and microphone.

Many problems with these devices have been reported by other researchers [1][6][7], however knowledge of this is not common. Nevertheless, we hope that this situation changes when people finally realize, that their devices are more than simply TVs and that somebody might take advantage of that in order to spy on them. Those risks can be especially significant when such device is a part of any Smart Home system [3].

In this paper we will focus on research performed by the Authors on abusing built-in features to remove all security mechanisms from a Smart TV. We will illustrate the process on Samsung Smart TVs because they are one of the largest supplier of such devices, reaching over 30% of market share in Europe [12].

In preparing this article we have decided to follow the guidelines of responsible disclosure, which means we will not be discussing in details techniques that can be used on newest devices, as to not endanger the user security in significant manner. Because the manufacturer has recently changed their operating system from Orsay to Tizen we will be describing the process and issues found in previous operating system, specifically vulnerabilities found on 2014 (H) model. However it should be noted that the problems we illustrate, as well as techniques described in this paper, can be adapted for use with newest operating system provided by the

[1]Ph.D., Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, tel: (+48) 42 631 27 50, fax (+48) 42 631 27 55, e-mail: mmajchr@iis.p.lodz.pl

[2]Ph.D., Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, tel: (+48) 42 631 27 50, fax (+48) 42 631 27 55, e-mail: pkapust@iis.p.lodz.pl

vendor. The significance of this research can be emphasized by the fact that both US and British intelligence have already used results of Authors research [13]. CIA in cooperation with the United Kingdom's MI5/BTSS have adapted techniques described in this paper in order to develop the Weeping Angel, which is a tool that allows spying trough Samsung Smart TVs [14].

It is important to note, that parts of the of the work have been described at the SamyGO forum (https://www.samygo.tv), which has been already cited by multiple researchers [2][9] as the main hub for information about hacking and rooting of Samsung Smart TVs. Parts of the research was performed with the help of the forum users, which made it possible to test the developed exploits on multiple devices from the same series. This confirmed the vulnerability for all of the tested Smart TVs.

2. Smart TV Applications

Smart TV is a term that defines a television set with integrated Internet support. One of the features of this platform is support for some form of Application Store, sometimes called Marketplace, from which users can download games and programs. Usually there are two kinds of applications in this marketplace: web and native. However word "web" shouldn't be interpreted in a classical way. Here it simply means that the application is written using SDK (Software Development Kit) that offers mostly "Web 2.0" technologies like HTML (HyperText Markup Language), JavaScript, CSS (Cascading Style Sheets) and some TV specific APIs (Application Programming Interface). Such applications are processed by a specialized HTML and JavaScript engine, and are often written acording to the Web design guidelines [8]. On the other hand native applications are written using some form of NDK (Native Development Kit) and have a form of binary files that are executed directly by the device CPU (Central Processing Unit). It should also be noted that in case of majority of vendors access to SDK is free whereas access to NDK is paid and restricted. As a result usually only big companies can afford it or are allowed to use it.

2.1. Samsung Smart TV SDK

Samsung distributes SDK that allows developing HTML5 based applications which are called widgets. Developers are allowed to freely test them on both emulator, which is provided by SDK, as well on actual devices. Applications can be tested on the device in one of two ways. First requires setting up a web server with widget list and each widget files packed in a zip format. Thenceforth it is necessary to switch the TV to developer mode and synchronize widgets with selected server. If all steps were done correctly TV will download the application and install it, which enables launching it as any other widget. However there is a much simpler way of installing new software on the TV. It's enough to unpack zip file mentioned in previous step to a USB memory stick and, after connecting this device to the TV, new application will appear in Smart Hub. As a result, due to limitations in NDK usage described above as well as relative ease with which USB method can be used, SDK is the obvious target for abuse. This method is also used by the Weeping Angel [13]

In this work we have decided to concentrate our efforts specifically on researching any potential vulnerabilities in the official SDK based on its public availability. The issue of developing good and secure public JavsScript APIs has been thoroughly researched, even to the point of preparing automatic tools for testing and finding vulnerabilities [11]. However, as with every

public API, it is hard to make it secure and easy to use at the same time. It is also, potentially, the easiest attack vector for accessing the TVs capabilities. There are of course other techniques, that can give even better results, but those usually require extensive knowledge about inner workings of the device firmware [9] or physical assess to the device, which can enable full forensic analysis [2]. However, attack vector through sanctioned market apps and public APIs are still preferable as they can allow for fully remote code execution without the user's knowledge or permission.

2.2. Smart TV applications sandbox

In computer security the term sandbox is a safeguard measure, that separates the currently running application from the OS. As a result if security of this piece of software is breached it does not immediately compromise whole system. Such approach is often used for programs that are untested, untrusted or were made by third party developers. Sandboxing isn't very popular on PC market tough for past few years it has been used by Apple for it's Mac App Store applications, as well as Microsoft for Store apps. However, it has became a standard for mobile applications and it's also commonly used on Smart TVs. Very often sandbox represents a single folder or a set of folders to which application has access to. Therefore widgets will not be able to read nor modify system files. In order to enforce that applications can only access the filesystem using a specific set of APIs. Privilege restriction is a good way of enhancing OS security [10], however it has to be done correctly and with proper granularity for it to work, which is not the case for Samsung Smart TV web apps, which we will show in this article.

2.3. Public APIs for accessing the file system

The first step of exploiting the vulnerabilities of the public Store APIs was to research the possibility of accessing the file system. Normaly access to Samsung Smart TVs file system is provided by FilePlugin and information about attached USB devices can be obtained via StoragePlugin API. In order to use those plugins, the developer first needs to put the following piece of code in the app HTML file:

```
<object id="pluginFileSystem" border=0
classid="clsid:SAMSUNG-INFOLINK-FILESYSTEM"></object>
<object id="pluginStorage" border=0
classid="clsid:SAMSUNG-INFOLINK-STORAGE"></object>
```

Listing 1: Piece of HTML code needed for FilePlugin and StoragePlugin APIs

After attaching FilePlugin object to HTML file [Listing 1] developer can access it using following piece of code:

```
var FilePlugin = document.getElementById("pluginFileSystem");
var StoragePlugin = document.getElementById("pluginStorage");
```

Listing 2: Piece of JavaScript code needed for FilePlugin and StoragePlugin APIs

After obtaining access to FilePlugin [Listing 2] it is possible to get access to a file system by executing one of following functions: `SetAppInfo`, `GetPluginInfo`, `Unzip`, `Copy`, `Move`, `Delete`, `IsExistedPath`, `GetType`, `GetTotalSize`, `GetUsedSize`, `GetFreeSize`, `GetFolderSize`,

UnzipAsync, DeleteAsync. We can also set specific callbacks: OnDeleteAsyncComplete, OnCopyAsyncComplete, OnMoveAsyncComplete and OnUnzipAsyncComplete.

StoragePlugin gives access to a lot useful informations about attached USB devices through functions: GetUSBListSize, GetUSBDeviceID, GetUSBPartitionNum, GetUSBVendorName, GetUSBModelName, GetUSBMountPath, GetUSBAvailSize, GetUSBTotalSize. We can also make our widget react to plugging or unplugging of USB devices by using SubscribeEvent function to set a callback for USB_CONNECTED_EVENT and USB_DISCONNECTED_EVENT events.

3. Abusing FilePlugin API to escape the sandbox

After researching the topic we have discovered, that on Orsay there are only two partitions that are mounted with write access: /mtd_rwarea and /mtd_rwcommon, whereas everything else is mounted as read only. Of course we also have write access to USB devices but that's part of what is allowed by the public SDK. Other than that application should be able to write data only into it's own folder, which in our case is also located on USB. As a result it is not possible to execute the following piece of code:

```
var fromUSB = "/dtv/usb/sda1/binaryForRooting.so";
var toOutsideSandbox = "/mtd_rwcommon/";
FilePlugin.Copy(fromUSB, toOutsideSandbox);
```

Listing 3: Failed attempt of sandbox escape with FilePlugin API

Therefore at first glance everything looks to be as intended. Every application is kept inside it's own sandbox and can't interact with other parts of file system. All calls to it are handled trough FilePlugin API which properly restricts the access. However, after further analysis of the FilePlugin API behavior and how it is used throughout the system we were able to determine where this check is taking place. After careful analysis of Orsay internals we can clearly see that many of system application like Samsung Store also use FilePlugin API, for example to install new applications. In consequence it is obvious that the check ensuring sandbox integration has to be implemented at language engine level. Such approach has its pros and cons. It allows for reuse of existing code by applications of different types and levels of security. However main con is that JavaScript is a dynamically typed language which often is executed after performing JIT (Just In Time) compilation.

```
var fromUSB = "/dtv/usb/sda1/binaryForRooting.so";
var toOutside = "/mtd_rwcommon/";
eval("FilePlugin.Copy('" + fromUSB + "','" + toOutside+ "')";
```

Listing 4: Successful sandbox escape with FilePlugin API

As a result many checks performed at the engine level can be bypassed by exploiting language features. One of them is eval function which performs JIT compilation of piece of code passed to it by string and immediately executes it. Therefore if we modify our small piece of code [Listing 3] and instead of executing the code directly we will construct a string with paths outside of sandbox [Listing 4]. As a result we will be able to bypass all security checks and use FilePlugin API to copy, remove or modify files in any writeable location. With such access level it's pretty easy to replace even system binaries with our code and completely root selected device [Figure 1].

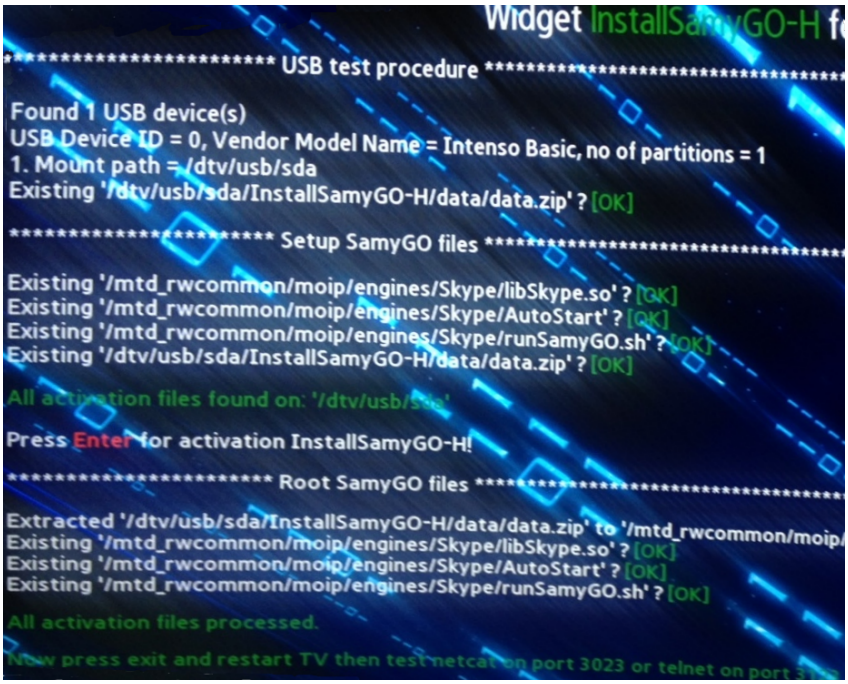


Figure 1: Screenshot of sample application using FilePlugin and StoragePlugin API for rooting.

3.1. Achieving direct command execution trough FilePlugin API and replacing system executables

On some firmware versions it is even possible to execute commands directly trough FilePlugin API. Strings passed to Copy function are passed directly to `/bin/cp` shell command tough before doing that both are sanitized. However presence of some characters like `$`, `(` and `)` is not checked. As a result it's possible to execute custom shell command [Listing 5].

```
var fromUSB = "/proc/self/cmdline";
var toTmp = "$(sh␣/mtd_rwcommon/SamyGO/run.sh)/tmp/cmdline";
eval("FilePlugin.Copy('" + fromUSB + "','" + toTmp+ "')";
```

Listing 5: Successful execution of custom shell command with FilePlugin API

Command presented on listing 5 will not only copy `/proc/self/cmdline` to `/tmp/cmdline` but at the same time execute script located at `/mtd_rwcommon/SamyGO/run.sh`.

As it was mentioned before most of system binaries, libraries and scripts are mounted on read-only partition which checksum is verified on every boot. Therefore we are not allowed to modify them. However TV has to keep user settings, applications and other data somewhere. That's precisely why `/mtd_rwarea` and `/mtd_rwcommon` are writable. One of the things that TV writes there is settings needed for wireless connection. On 2014 H models script `/mtd_rwarea/network/insmod_qca.sh` is used for devices with Atheros wireless card whereas script `/mtd_rwarea/network/insmod_bcm.sh` serves the same purpose on those with Broadcom chips. This script is executed directly by `system()` function so it has to have executable bit set.

For security reason the operating system of the TV mounts USBs as noexec, which marks everything as not executable, so if we simply tried to copy the script from USB memory stick to `/mtd_rwarea/network/` we wouldn't be able to achieve code execution. However the Authors

have noticed that FilePlugins `Unzip()` API does preserve executable bit. It behaves this way because it's used by Samsung Store application and some widgets available there have custom binaries as part of zip files. As a result Authors were also able to abuse FilePlugin API to replace scripts used by system to setup wireless connection [Listing 6]. Usually on 2014 H models commands and scripts are executed with `app` user privilege however since network scripts need to insert kernel module they are executed by `root`. As a result we were able to get full access to the device and bypass both sandboxing as well as privilege separation only by abusing built-in JavaScript API.

```
var fromUSB = "/dtv/usb/sda1/data.zip";
var toNetwork = "/mtd_rwarea/network/";
eval("FilePlugin.Unzip('" + fromUSB + "','" + toNetwork+ "')";
```

Listing 6: Successful replacement of builtin network script with FilePlugin API

4. Conclusions

Smart TVs are becomming again a hub for entertainment in a home. With apps like Facebook and Twitter one can interact with social media, Netflix, Hulu, YouYube and Spotify give access to vast multimedia collections and Playstation Now can stream games directly on a big screen. With this even people that are not tech savvy can use this technology. This is a big concern, as people become much less security conscious in such environment. Which is why research about any potential vulnerabilities in every internet connected device is so important.

As was presented in the article even at first glance harmless decision in designing application framework can result in a significant security breach down the road. Sometimes even simple and presumably innocuous APIs can be used to bypass multiple layers of security like sandboxing or even allowing unauthorized command execution. Market for Smart TVs is growing constantly and so is the risk of attack and leakage of sensitive data. Even devices, that have been until now considered relatively safe can, and will be, attacked and subsequently hacked.

In this article the Authors wanted raise the awareness of this issue, especially now, when the devices are changing so rapidly that manufacturers don't always take the issues of safety as seriously as they should.

References

1. Y. Bachy et al., "Smart-TV Security Analysis: Practical Experiments", *2015 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks*, Rio de Janeiro, 2015, pp. 497-504.

2. A. Boztas, A.R.J. Riethoven, M. Roeloffs, "Smart TV forensics: Digital traces on televisions", *Digital Investigation, Volume 12, Supplement 1*, 2015, pp. S72-S80

3. E. Fernandes, J. Jung and A. Prakash, "Security Analysis of Emerging Smart Home Applications", *2016 IEEE Symposium on Security and Privacy (SP)*, San Jose, CA, 2016, pp. 636-654.

4. A. Grattafiori, J. Yavor, "The outer limits: hacking the Samsung Smart TV", *Blackhat Briefing 2013*, Las Vegas, July 2013.

5. A. Grau, "Can you trust your fridge?", *IEEE Spectrum, vol. 52, no. 3*, 2015, pp. 50-56

6. S. Landau, ”What Was Samsung Thinking?”, *IEEE Security & Privacy, vol. 13, no. 3*, pp. 3-4, May-June 2015.

7. M. Niemietz, J. Somorovsky, C. Mainka and J. Schwenk, ”Not so Smart: On Smart TV Apps”, *2015 International Workshop on Secure Internet of Things (SIoT)*, Vienna, 2015, pp. 72-81.

8. E. Perakakis and G. Ghinea, ”Responsive web design for the internet connected TV: The answer to more smart TV content?”, *2015 IEEE 5th International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*, Berlin, 2015, pp. 38-42.

9. L. SeungJin, K. Seungjoo, ”Hacking, Surveiling and Deceiving Victims on Smart TV”, *Black Hat Briefings*, Las Vegas, July 2013

10. H. Shukla, V. Singh, Y. H. Choi, J. Kwon and C. h. Hahm, ”Enhance OS security by restricting privileges of vulnerable application”, *2013 IEEE 2nd Global Conference on Consumer Electronics (GCCE)*, Tokyo, 2013, pp. 207-211.

11. A. Taly, U. Erlingsson, J.C. Mitchell, M.S. Miller, J. Nagra, ”Automated Analysis of Security-Critical JavaScript APIs”, *SP ’11 Proceedings of the 2011 IEEE Symposium on Security and Privacy*, 2011, pp. 363-378

12. ”Share of smart TV sales in Europe by operating system in 2016” *https://www.statista.com/statistics/660850/smart-tv-operating-system-market-share-in-europe*

13. ”Vault 7: CIA Hacking Tools Revealed” *https://wikileaks.org/ciav7p1/*

14. ”Weeping Angel: The latest surveillance tool, that can turn your smart TV into a bug TV” *https://mastersofmedia.hum.uva.nl/blog/2017/09/25/weeping-angel-cia-bug-smart-tv/*

Marcin SOWA[1]

SubIval computation time assessment

The paper presents a part of a study aiming at the improvement of the efficiency of SubIval (the subinterval-based method for fractional derivative computation in initial value problems). The current paper concerns the computation time in a constant step size time stepping solver basing on SubIval. The basics of SubIval are briefly given and some already applied improvements have been listed. A computational problem has been introduced for which the computation time was studied.

1. Introduction

Fractional calculus is a field with many applications e.g. in control theory and its applications (due to properties of fractional PID controllers [1, 2]), circuit theory [3, 4, 5], electric and magnetic field analyses [6, 7, 8], viscoelasticity [9, 10] and the design of fractional order filters [11, 12].

Because of this, it is important to find new possibilities of dealing with elements of fractional calculus appearing in the mathematical descriptions of computational problems. Out of the many elements of fractional calculus (fractional derivative and integral definitions [13] with respect to both space [14] and time) the paper concerns only the Caputo definition [15] of the fractional time derivative of order $\alpha \in (0,1]$:

$$ {}_{t_\mathrm{a}}\mathrm{D}^{\alpha}_{t_\mathrm{b}} x(t) = \frac{1}{\Gamma(1-\alpha)} \int_{t_\mathrm{a}}^{t_\mathrm{b}} \frac{x^{(1)}(\tau)}{(t-\tau)^{\alpha}} \mathrm{d}\tau. \tag{1} $$

Lately the author of this paper has been working on a numerical method, now known by its acronym – SubIval (first introduced as the subinterval-based method for fractional derivative computations in initial value problems in [16]). It has been described with great detail in [17] for adaptive time stepping solvers. There is a DLL available in [18] (along with examples of its usage), which allows it to be used in popular computer algebra systems (it has been tested in Matlab, Mathematica and FreeMat 32-bit) and programs supporting the .NET Framework (the DLL has been written in C#).

In the previous papers the author has shown that the method can be very accurate – this has been proven through a comparison of the results with evaluations of analytical solutions (for selected problems).

This paper concerns the issue of the computation time of the discussed numerical method.

2. SubIval basics

SubIval is designed for time stepping solvers where in each time step the solution for selected variables is computed only for one time instance (denoted by $t = t_{\mathrm{now}}$). The approximation of

[1]PhD (EEng), Silesian University of Technology, Akademicka 2A, 44-100 Gliwice, POLAND, e-mail: marcin.sowa@polsl.pl

the fractional derivative that results from the application of SubIval has the implicit form:

$$ {}_{t_0}\mathrm{D}^{\alpha}_{t_{\mathrm{now}}} x(t) \approx a\, x(t_{\mathrm{now}}) + b, \tag{2} $$

where a and b result from the internal computations of SubIval. This applies for all variables under fractional time derivatives of order $\alpha \in (0,1]$ (further on referred to as state variables).

When dealing with typical computations – one does not need to know how it exactly works if one applies the available DLL [18]. The method is convenient enough that what one needs to do every time a new time instance is selected is to inform the library of that t_{now} value and of the value of x computed at the previously selected time instance. The library then provides the a and b values, which allow to substitute the fractional derivative into an implicit dependence on the variable.

This paper addresses the issue of the time required to perform the computations in a single time step. For the purpose of this analysis it is worthwhile to briefly recall some of the properties of SubIval.

During each time step a new time instance is proposed (either adaptively [17] or using a constant step size) for which the solution is obtained. These time instances form nodes on the time axis. The nodes are then used for local polynomial approximations (denoted by $\widetilde{x}_s(t)$, with $s = 1,2,\ldots M$) defined in local subintervals (further on denoted by Θ_s) covering the nodes of interest. M denotes the number of approximating polynomials (at the same time it is the number of the Θ_s intervals). An example of how the Θ_s intervals are set up is presented in Fig. 1.

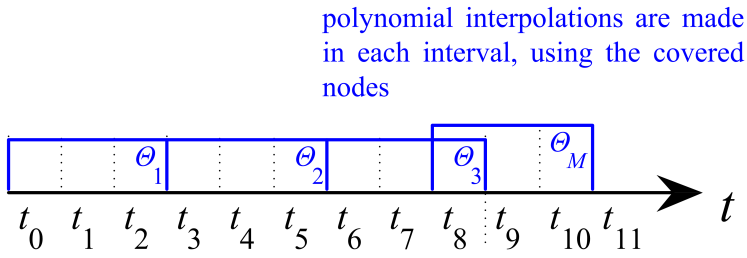


Figure 1: Example of a setup of Θ_s subintervals in SubIval. The order of all $\widetilde{x}_s(t)$ polynomials is 3.

The local polynomial approximations are used in the base equation of SubIval, i.e. its partition into local integrodifferentiations on subintervals $\Xi_s = [t_{s\,\mathrm{start}}, t_{s\,\mathrm{end}}]$:

$$ {}_{t_0}\mathrm{D}^{\alpha}_{t_{\mathrm{now}}} x(t) \approx \sum_{s=1}^{M} {}_{t_{s\,\mathrm{start}}}\mathrm{D}^{\alpha}_{t_{s\,\mathrm{end}}} \widetilde{x}_s(t), \tag{3} $$

where:

- $t_{1\,\mathrm{start}} = t_0$ (the initial time instance) and $t_{M\,\mathrm{end}} = t_{\mathrm{now}}$ (always denoting the final time instance, being the one for which the solution is sought),
- the subintervals form a continuous set i.e. $t_{s\,\mathrm{end}} = t_{s+1\,\mathrm{start}}$ for $s = 1,2,\ldots M-1$ (how the Ξ_s subintervals are set up in the case given previously given as an example in Fig. 1 is presented in Fig. 2),
- each $\widetilde{x}_s$ is defined using nodes denoted by $t_{s,k}$ with $k = 1,2,\ldots n_s$ (hence $q_s = n_s - 1$ is the order of the polynomial denoted by $\widetilde{x}_s$),

- $\Xi_s \subseteq \Theta_s$ for each $s = 1, 2, \ldots M$.

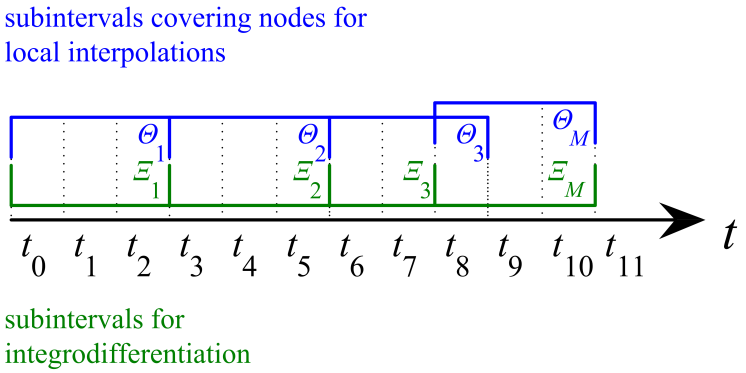


Figure 2: Example of a setup Θ_s and Ξ_s subinterval pairs.

The manner in which each subinterval pair (Ξ_s with Θ_s) is formulated is determined by an original subinterval dynamics algorithm explained (with great detail) in [17].

Each polynomial can be given in terms of Lagrange basis polynomials:

$$\widetilde{x}_s = \sum_{k=1}^{n_s} x_{s,k} L_{s,k}(t) \tag{4}$$

where $x_{s,k}$ denotes the numerical solution for the $t_{s,k}$ node and the Lagrange basis polynomial:

$$L_{s,k}(t_{s,j}) = \begin{cases} 0, & \text{if } j \neq k, \\ 1, & \text{if } j = k. \end{cases} \tag{5}$$

SubIval is designed for typical time stepping solvers, in which the solution is computed for one time instance at a time, meaning that for a time instance t_j variables for previous time instances $t_1, t_2, \ldots t_{j-1}$ are known. Only the $\widetilde{x}_M$ polynomial takes into account the final node on the time axis (for which the state variables' values are still unknown). This finally leads to Eq. (2), with:

$$a = {}_{t_{M\,\mathrm{start}}}\mathrm{D}^{\alpha}_{t_{M\,\mathrm{end}}} L_{M,n_M}(t), \tag{6}$$

and:

$$b = {}_{t_{M\,\mathrm{start}}}\mathrm{D}^{\alpha}_{t_{M\,\mathrm{end}}} \sum_{k=1}^{n_M-1} x_{M,k} L_{M,k}(t) + \sum_{s=1}^{M-1} {}_{t_{s\,\mathrm{start}}}\mathrm{D}^{\alpha}_{t_{s\,\mathrm{end}}} \sum_{k=1}^{n_s} x_{s,k} L_{s,k}(t). \tag{7}$$

In order to deal with the local integrodifferentiations, SubIval uses analytical formulae for each monomial $c\,t_{\mathrm{loc}}^k$ (with $t_{\mathrm{loc}} = t - t_{s,1}$, $c \in \mathbb{R}$ and $k \in \mathbb{Z}_+$). The following formula has been derived in a previous study [19]:

$${}_{t_{\mathrm{a}}}\mathrm{D}^{\alpha}_{t_{\mathrm{b}}}\, c\, t_{\mathrm{loc}}^k = \frac{kc}{\Gamma(1-\alpha)} t_{\mathrm{loc}}^{k-\alpha} \left(\mathrm{B}_{\frac{t_{\mathrm{loc\,b}}}{t_{\mathrm{loc}}}}(k, 1-\alpha) - \mathrm{B}_{\frac{t_{\mathrm{loc\,a}}}{t_{\mathrm{loc}}}}(k, 1-\alpha)\right), \tag{8}$$

with $t_{\text{loc a}} = t_{\text{a}} - t_{s,1}$, $t_{\text{loc b}} = t_{\text{b}} - t_{s,1}$ and [20]:

$$\mathrm{B}_{\rho}(k, 1-\alpha) = \frac{\Gamma(k)\Gamma(1-\alpha)}{\Gamma(k+1-\alpha)} \left(1 - (1-\rho)^{1-\alpha} \sum_{j=0}^{k-1} \frac{\rho^j \prod_{i=0}^{j-1}(1-\alpha+i)}{j!} \right). \tag{9}$$

Taking into account the fact that with each new time step the number of dependencies on previously computed values, hence also – basic computations, rises – the computation time will, naturally, rise for later time steps. This issue is discussed in further sections of the paper.

3. Simple improvements

A few observations can be made, which lead to improvements in the time required for the SubIval internal computations. These improvements are implemented in version 1.02 of the SubIval DLL (some of the improvements have already been applied in previous versions). The observations are such that for each new time instance being introduced:

- only the Lagrange basis polynomials for $\widetilde{x}_M$ need to be established anew,
- each polynomial $\widetilde{x}_s$ with $s < M$ does not need to be updated,
- the gamma function computations in (9) can be done initially for each distinct α and $k = 1, 2, \ldots q_{\max}$ (with $q_{\max}$ being the maximum polynomial order),
- the same goes for the $\frac{\prod_{i=0}^{j-1}(1-\alpha+i)}{j!}$ terms in (9).

Naturally, a lot of factors will influence the computation time of the time stepping solver apart from the internal SubIval computations. For example, if dealing with nonlinear problems the computation times can vary greatly. If an adaptive step size algorithm is used then additional computations need to be made per time step (like error estimation).

The current paper only raises the issue of the computation time of a solver (basing on SubIval) for the simplest case, i.e. for linear problems and a constant step size Δt. This is because the computation time of SubIval has not yet been discussed in any earlier paper, hence the choice of the most basic case in this initial study.

4. Computational example

In this section the *RC* ladder circuit problem presented in Fig. 3 is studied. The example has been brought forward to study the computation time of a SubIval solver. The example has been chosen because of its known analytical solution [3] for any number of fractional capacitors (this number is denoted n). Moreover, for a significant number of capacitors, in the time stepping solver – a significant amount of equations can appear that need to be solved in each time step.

For the computations of this paper the number of capacitors was set to $n = 15$. The time functions obtained by means of the numerical solver are presented in Fig. 4. The domain for the solution was set to $t \in [0, T_{\max}]$ with $T_{\max} = 20\ \mu$s. The time step for the numerical solver was set to $\Delta t = T_{\max}/2000$.

The solution of SubIval has been verified through the computation of a relative error given

by:

$$e_{\text{an vs num}} = 100 \cdot \frac{|x(t_j) - x_{\text{num } j}|}{\max\limits_{j=1,2,\ldots m} |x(t_j)|} \ \%. \tag{10}$$

The above has been checked for each capacitor voltage being the x variable. The $t = t_1, t_2, \ldots t_m$ nodes are the ones selected in the numerical solution. $x(t_j)$ denotes the value of the variable for the $t = t_j$ node, which resulted in the evaluation of the analytical solution. $x_{\text{num } j}$ is the value at that same node but computed through the numerical method.

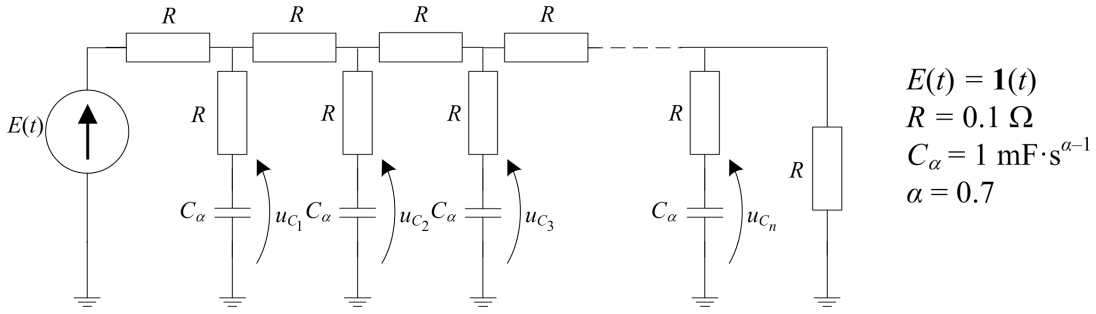


Figure 3: Exemplary problem: RC ladder circuit (all capacitors are of order α).

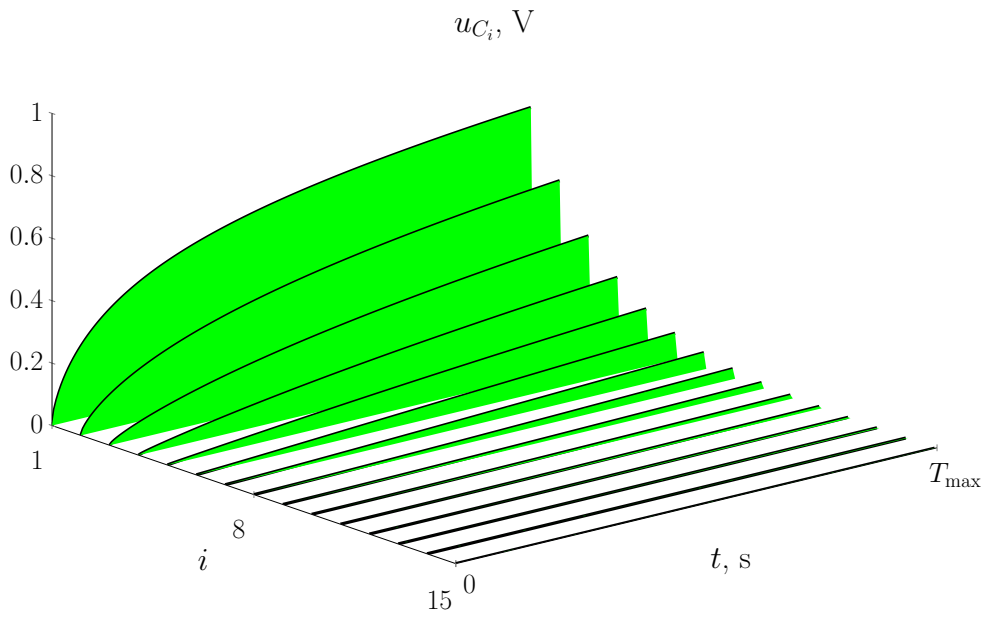


Figure 4: Time functions of the capacitor voltages obtained by the numerical solver based on SubIval.

The differences between the analytical and numerical solutions have been very small (the maximum relative error from all the obtained voltage time functions has been around 0.3 %). As it has been numerously proven [16, 17, 19] a solver basing on SubIval can be very accurate. If an even more accurate solution is required – one can apply a smaller Δt value. Another way is to apply adaptive step size techniques [17] though this still generally leads to smaller step sizes. The choice of smaller step sizes, in turn, leads to more nodes being taken into account in the SubIval computations. This leads to increased computation times. A chart of the computation

times for the studied problem is given in Fig. 5. The plot actually shows minimum computation times from 20 attempts (to eliminate occurrences where there have been delays caused by other programs running on the computer [2]).

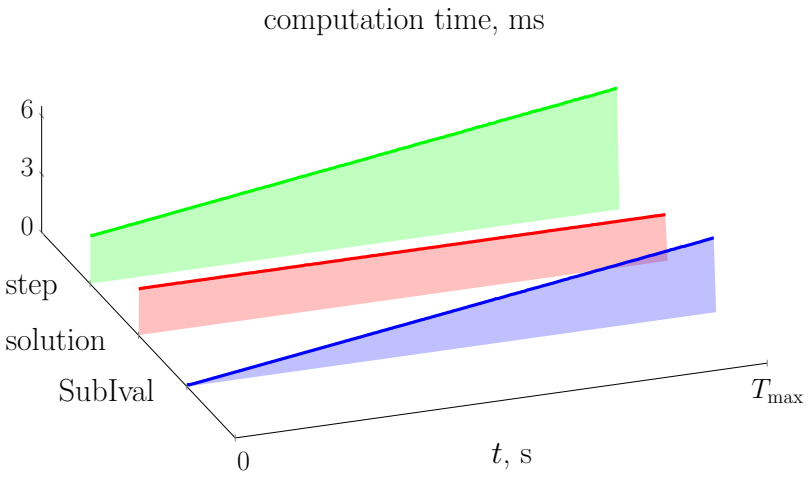


Figure 5: Computation times for the subsequent steps of the solver, the solutions in each step and the SubIval computations.

Obviously the solution of a system of equations requires very similar computation times in each time step when a linear system is solved. The time required for the SubIval computations rises linearly. The problem and the options for the solver have been chosen deliberately so that the computation of the solution and the SubIval computations have comparable durations. However, there are cases where this is not true and e.g. if a large system is studied where there are not many fractional elements. In such a case the SubIval computations could be insignificant and, therefore, do not need to be optimized.

5. Conclusions and remarks for future research

The computation time has been studied for a solver based on SubIval (the subinterval-based method for fractional derivative computations in initial value problems). The SubIval basics have been briefly described and some simple improvements (that have already been implemented) have been listed. A circuit problem (with a known analytical solution) has been introduced as a test example to ascertain the computation time for each time step of the SubIval solver.

Some improvements to consider for future research relate to the problem of increasing computation time for the internal SubIval computations. These take into account increasing amounts of terms with every new time step. One of the possibilities that will be considered is to exclude subintervals (in the computations of the b coefficients) that contribute below a certain threshold. Another possibility is to simplify the computations of more distant nodes (e.g. as it has been done in [21] for a different method).

[2] the computations have been made with a C# program running on Windows 7 on a 2.26 GHz dual-core processor

References

1. Spałek D.: "Synchronous Generator Model with Fractional Order Voltage Regulator PI^{b}D^a", *Acta Energetica*, 2015, Vol. 23 nr 2, 78–84

2. Mercorelli P.: "A Discrete-Time Fractional Order PI Controller for a Three Phase Synchronous Motor Using an Optimal Loop Shaping Approach", *Theory and Applications of Non-Integer Order Systems*, Springer, 2017, 477–487

3. Mitkowski W., Skruch P.: "Fractional-order models of the supercapacitors in the form of RC ladder networks", *Bull. Pol. Ac.: Tech.*, 2013, Vol. 61, nr 3, 580–587

4. Jakubowska-Ciszek A., Walczak J.: "Analysis of the transient state in a parallel circuit of the class $RL_\beta C_\alpha$", *Applied Mathematics and Computation*, 2017 (in press)

5. Schäfer I., Krüger K.: "Modelling of lossy coils using fractional derivatives", *Phys. D: Appl. Phys.*, 2008, Vol. 41, 1–8

6. Garrappa R., Maione G.: "Fractional Prabhakar Derivative and Applications in Anomalous Dielectrics: A Numerical Approach", *Theory and Applications of Non-Integer Order Systems*, Springer, 2017, 429–439

7. Mescia L., Bia P., Caratelli D., "Fractional Derivative Based FDTD Modeling of Transient Wave Propagation in Havriliak-Negami media", *IEEE Transactions on Microwave Theory and Techniques*, 2014, Vol. 62, nr 9, 1920–1929

8. Litak G., Ducharne B., Sebald G., Guyomar D., "Dynamics of magnetic field penetration into soft ferromagnets", *Journal of Applied Physics*, 2015, Vol. 117, nr 24, 243907

9. Faraji Oskouie M., Ansari R.: "Linear and nonlinear vibrations of fractional viscoelastic Timoshenko nanobeams considering surface energy effects", *Applied Mathematical Modelling*, 2007, Vol. 43, 337–350

10. Ezzat M.A., El-Karamany A.S., El-Bary A.A.: "Thermo-viscoelastic materials with fractional relaxation operators", *Applied Mathematical Modelling* 2015, Vol. 39, Issues 23–24, 7499–7512

11. Bauer W., Kawala-Janik A., "Implementation of Bi-fractional Filtering on the Arduino Uno Hardware Platform", *Theory and Applications of Non-Integer Order Systems*, 2017, 419–428

12. Kawala-Janik A., Podpora M., Gardecki A., Czuczwara W., Baranowski J., Bauer W.: "Game controller based on biomedical signals", *Methods and Models in Automation and Robotics (MMAR) 2015 20th International Conference*, 2015, 934–939

13. Katugampola U.N.: "Mellin transforms of generalized fractional integrals and derivatives", *Applied Mathematics and Computation*, 2015, Vol. 257, 566–580

14. Gómez Aguilar J.F., Hernández M.M.: "Space-Time Fractional Diffusion-Advection Equation with Caputo Derivative", *Abstract and Applied Analysis*, 2014, 8 pages.

15. Caputo M.: "Linear models of dissipation whose Q is almost frequency independent – II", *Geophysical Journal International*, 1967, Vol. 13, nr 5, 529–539

16. Sowa M., "A subinterval-based method for circuits with fractional order elements", *Bull. Pol. Ac.: Tech.*, 2014, Vol. 62, nr 3, 449–454

17. Sowa M.: "Application of SubIval in solving initial value problems with fractional derivatives", *Applied Mathematics and Computation*, 2017 (in press).

18. http://msowascience.com

19. Sowa M.: "Application of SubIval, a Method for Fractional-Order Derivative Computations in IVPs", *Theory and Applications of Non-Integer Order Systems*, Springer, 2017, 405–418

20. http://functions.wolfram.com/GammaBetaErf/Beta3/03/01/02/0003/

21. Ostalczyk P.: "On simplified forms of the fractional-order backward difference and related fractional-order linear discrete-time system description", *Bull. Pol. Ac.: Tech.*, 2015, Vol. 63, nr 2, 423–433

Paweł MACIĄG[1], Paweł MALCZYK[2], Janusz FRĄCZEK[3]

Kinematic Synthesis of Planar Robotic Systems

The tools for dimensional synthesis of robotic linkages that are modeled as multi-rigid-body systems are often used in the design process. Many robotic devices such as, e.g.: robotic manipulators or mechanical hands, are designed to provide an appropriate workspace for a link or end-effector. The objective of the kinematic synthesis is to determine the dimensions of links that help to achieve a specific task by taking into account the complexity of a multibody system (MBS). This paper presents a general method for kinematic optimal synthesis of planar MBS. The solutions of such optimization problems require evaluation of a gradient of an objective function. The technique described herein is used for the kinematic analysis and synthesis of planar human-like robotic finger. Sample simulation scenarios and numerical results are reported in the text.

1. Introduction to kinematic synthesis

Kinematic synthesis, also known as geometric design of mechanisms, is the process of finding such bodies' dimensions that provide desired motion of its links [5]. This problem is particularly important in mechanical design of various robotic systems including serial or parallel topologies [1], [2], [3], [6], [9], [10]. The concept of designing most mechanisms is to transfer simple linear or rotary motion into more complex movements. When a multibody system contains many bodies, it is convenient to introduce absolute coordinates for the MBS state description. Such coordinates are commonly used in many general formulations of kinematics and dynamics of constrained multibody systems [7], [8]. Therefore, position of each body in the system is defined by a set of absolute coordinates, which for planar i-th body can be expressed as: $\mathbf{q}_i = [x_i, y_i, \varphi_i]^T$, where x_i, y_i denote position of coordinate frame fixed with i-th body (usually located at its center of mass), and φ_i is an angle between local and global coordinate systems. The existence of joints and imposed kinematic motion introduce n constraint equations into the system description:

$$\boldsymbol{\Phi}(t,\mathbf{q}) = \left[\left[\boldsymbol{\Phi}^K\right]^T \quad \left[\boldsymbol{\Phi}^D\right]^T\right]^T = \mathbf{0}_{n\times 1}, \tag{1}$$

where $\boldsymbol{\Phi}^K$ is a vector of kinematic constraints that define rotational or translational joints, whereas $\boldsymbol{\Phi}^D$ is a vector of driving constraints that defines desired motion in certain joints. Solution of (1) yields position vector $\mathbf{q}$ for a given time interval. Since equations (1) are algebraic and usually highly nonlinear, the Newton - Raphson method is employed to solve the problem of kinematic analysis. Velocity and acceleration level constraint equations are obtained by differentiating (1) once and twice with respect to time. Let us denote $\mathbf{z}$ as design variables (e.g. link dimensions), then, the constraint equations may be written in the following form:

$$\boldsymbol{\Phi}(t,\mathbf{z},\mathbf{q}) = \mathbf{0}, \qquad \boldsymbol{\Phi}_\mathbf{q}(\mathbf{z},\mathbf{q})\cdot\dot{\mathbf{q}} - \boldsymbol{\Phi}_t(t) = \mathbf{0}, \qquad \boldsymbol{\Phi}_\mathbf{q}(\mathbf{z},\mathbf{q})\cdot\ddot{\mathbf{q}} - \boldsymbol{\Gamma}(t,\mathbf{z},\mathbf{q},\dot{\mathbf{q}}) = \mathbf{0}, \tag{2}$$

[1]M.Sc., Division of Theory of Machines and Robots, Faculty of Power and Aeronautical Engineering, Institute of Aeronautics and Applied Mechanics, Warsaw University of Technology, Nowowiejska 24, 00-665 Warsaw, POLAND, e-mail: pmaciag@meil.pw.edu.pl

[2]Ph.D., see above, e-mail: pmalczyk@meil.pw.edu.pl

[3]Professor, see above, e-mail: jfraczek@meil.pw.edu.pl

where $\mathbf{\Phi_q}$ is a Jacobian matrix of constraint equations with respect to absolute coordinates and $\mathbf{\Phi}_t$ is a vector of partial derivatives of $\mathbf{\Phi}$ with respect to time. The vector $\mathbf{\Gamma}$ denotes a position and velocity dependent term of the second derivative of constraint equations.

The optimal kinematic synthesis of MBS can be formulated as an optimization problem. The purpose is to find a set of design variables $\mathbf{z} \in \mathscr{Z} \subset \mathscr{R}^m$ that will minimize a given cost (objective) function $f(\mathbf{z})$ under imposed constraints that may have a form of equality and non-equality relations:

$$\min_{\mathbf{z}\in\mathscr{Z}} f(\mathbf{z}) \quad \text{s.t.} \qquad \mathbf{h}(\mathbf{z}) = \mathbf{0}, \quad \mathbf{g}(\mathbf{z}) \leqslant \mathbf{0}. \tag{3}$$

As far as kinematic synthesis is concerned, design variables usually represent dimensions or configuration of linkages. The system of equations (2) is thus a set of algebraic equality constraints that may be explicitly written in the form: $\mathbf{h}(\mathbf{z},\mathbf{q},\dot{\mathbf{q}},\ddot{\mathbf{q}}) = \mathbf{0}$. The non-equality constraints $\mathbf{g}(\mathbf{z}) \leqslant \mathbf{0}$ may appear when there is need to limit values of certain design variables in terms of other design variables (e.g. Grashof condition for a 4-bar linkage), or - more generally - as a user defined inequality constraint function.
For geometric design problems, especially for the tasks of achieving a specific configuration or trajectory of a robotic tool, the cost function may be defined as a scalar function of deviations (residuals) between designed and reference trajectories. One popular approach to formulate optimal kinematic synthesis problem is to limit the upper bound of the deviation. Therefore, it is convenient to rewrite (3) in the form:

$$\min_{\mathbf{z}\in\mathscr{Z}} f(\mathbf{z},\mathbf{q},\dot{\mathbf{q}},\ddot{\mathbf{q}}) = \min_{\mathbf{z}\in\mathscr{Z}} \left[\max_{\boldsymbol{\alpha}} f(\mathbf{z},\mathbf{q},\dot{\mathbf{q}},\ddot{\mathbf{q}})\right] \quad \text{s.t.} \qquad \mathbf{h}(\mathbf{z}) = \mathbf{0}, \quad \mathbf{g}(\mathbf{z}) \leqslant \mathbf{0}. \tag{4}$$

The formulation (4) expresses the goal of the kinematic synthesis, which is to find a set of design variables $\mathbf{z}$ that minimizes the maximum of the objective function with respect to the set of input variables $\boldsymbol{\alpha}$. The variables $\boldsymbol{\alpha}$ may be intuitively defined as motion (driving) constraints for a given system. The process of searching for optimal design is an iterative process, during which, motion of the system is generated for the current value of $\mathbf{z}^j$ according to driving constraints. Driving constraints can be translational or angular positions that are sampled on a discrete grid. We identify $\mathbf{\Phi}^D$ as a vector of input parameters $\alpha = \alpha_k$, where $k = 1, 2 \cdots, \tau$ is a set of motion's grid points.

One way of solving optimization problem is called upper bound technique [4]. The main idea is to replace an objective function with a new design variable z_{m+1} under the assumption that

$$z_{m+1} \geqslant \max_{\boldsymbol{\alpha}} f(\mathbf{z},\mathbf{q},\dot{\mathbf{q}},\ddot{\mathbf{q}}). \tag{5}$$

Moreover, additional inequality constraint must be included in order to supplement (5). The min-max problem (4) can thus be transformed into the following one:

$$\min_{\mathbf{z}\in\mathscr{Z}} f = \min_{\mathbf{z}\in\mathscr{Z}} z_{m+1} \quad \text{subjected to} \tag{6a}$$

$$g(\mathbf{z}) = \max_{\boldsymbol{\alpha}} \left[f(\mathbf{z},\mathbf{q},\dot{\mathbf{q}},\ddot{\mathbf{q}})\right] - z_{m+1} \leqslant 0. \tag{6b}$$

Many optimization algorithms require from a designer to provide gradient of an objective function. In the case of MBS optimization, it is possible to efficiently implement general expressions that facilitate the calculation of a gradient. Gradient of an objective function contains information on how a change of each design variable affects its value. Thus, design sensitivity analysis can be performed. The design sensitivity vector is a quantity describing how an objective func-

tion varies with a variation of design variables. By reformulating the problem from Eq. (4) into Eq. (6b), one can notice that the design sensitivity vector is identical to the gradient of inequality constraints $g(\mathbf{z})$. In order to calculate it, Leibniz chain rule of differentiation is applied to Eq. (6b). For the j-th grid point, we get:

$$\left[\left(\nabla_{\mathbf{z}} g\right)\right]^j = \left(g_{\mathbf{z}}^T\right)^j = \left[g_{\mathbf{q}}\mathbf{q}_{\mathbf{z}} + g_{\dot{\mathbf{q}}}\dot{\mathbf{q}}_{\mathbf{z}} + g_{\ddot{\mathbf{q}}}\ddot{\mathbf{q}}_{\mathbf{z}} + \frac{\partial g}{\partial \mathbf{z}}\right]^j_{1\times m}. \tag{7}$$

Equation (7) can be rewritten as:

$$\left(g_{\mathbf{z}}^T\right)^j = \left[g_{\mathbf{q}}, g_{\dot{\mathbf{q}}}, g_{\ddot{\mathbf{q}}}\right]^j_{1\times 3n} \cdot \begin{bmatrix}\mathbf{q}_{\mathbf{z}} \\ \dot{\mathbf{q}}_{\mathbf{z}} \\ \ddot{\mathbf{q}}_{\mathbf{z}}\end{bmatrix}^j_{3n\times m} + \left[\frac{\partial g}{\partial \mathbf{z}}\right]^j. \tag{8}$$

According to equation (8), one requires partial derivatives of g with respect to absolute coordinates, its derivatives, and design variables in order to calculate the design sensitivity vector. Moreover, there is a dependence of kinematic variables on the vector of design variables that can be expressed as: $\mathbf{q} = \mathbf{q}(\mathbf{z}), \dot{\mathbf{q}} = \dot{\mathbf{q}}(\mathbf{z}), \ddot{\mathbf{q}} = \ddot{\mathbf{q}}(\mathbf{z})$. Since such dependence arises from constraint equations (2), it is possible to find the derivatives $\mathbf{q}_{\mathbf{z}}$, $\dot{\mathbf{q}}_{\mathbf{z}}$, $\ddot{\mathbf{q}}_{\mathbf{z}}$, to get:

$$\begin{bmatrix} \boldsymbol{\Phi}_{\mathbf{q}} & \mathbf{0} & \mathbf{0} \\ (\boldsymbol{\Phi}_{\mathbf{q}}\dot{\mathbf{q}})_{\mathbf{q}} & \boldsymbol{\Phi}_{\mathbf{q}} & \mathbf{0} \\ (\boldsymbol{\Phi}_{\mathbf{q}}\ddot{\mathbf{q}})_{\mathbf{q}} + [(\boldsymbol{\Phi}_{\mathbf{q}}\dot{\mathbf{q}})_{\mathbf{q}}\dot{\mathbf{q}}]_{\mathbf{q}} & [(\boldsymbol{\Phi}_{\mathbf{q}}\dot{\mathbf{q}})_{\mathbf{q}}\dot{\mathbf{q}}]_{\dot{\mathbf{q}}} & \boldsymbol{\Phi}_{\mathbf{q}} \end{bmatrix} \cdot \begin{bmatrix}\mathbf{q}_{\mathbf{z}} \\ \dot{\mathbf{q}}_{\mathbf{z}} \\ \ddot{\mathbf{q}}_{\mathbf{z}}\end{bmatrix} = \begin{bmatrix} -\boldsymbol{\Phi}_{\mathbf{z}} \\ -(\boldsymbol{\Phi}_{\mathbf{q}}\dot{\mathbf{q}})_{\mathbf{z}} \\ -[(\boldsymbol{\Phi}_{\mathbf{q}}\dot{\mathbf{q}})_{\mathbf{q}}\dot{\mathbf{q}}]_{\mathbf{z}} - (\boldsymbol{\Phi}_{\mathbf{q}}\ddot{\mathbf{q}})_{\mathbf{z}} \end{bmatrix}. \tag{9}$$

The expression (9) is a linear set of algebraic equations. The solution of the system yields absolute derivatives of displacement, velocity, and acceleration with respect to design variables. The solution of (9) can be inserted back into Eq. (8) to give the desired partial derivative $g_{\mathbf{z}}$.

2. Kinematic synthesis of human-like robotic finger

2.1. Kinematic model and optimization problem

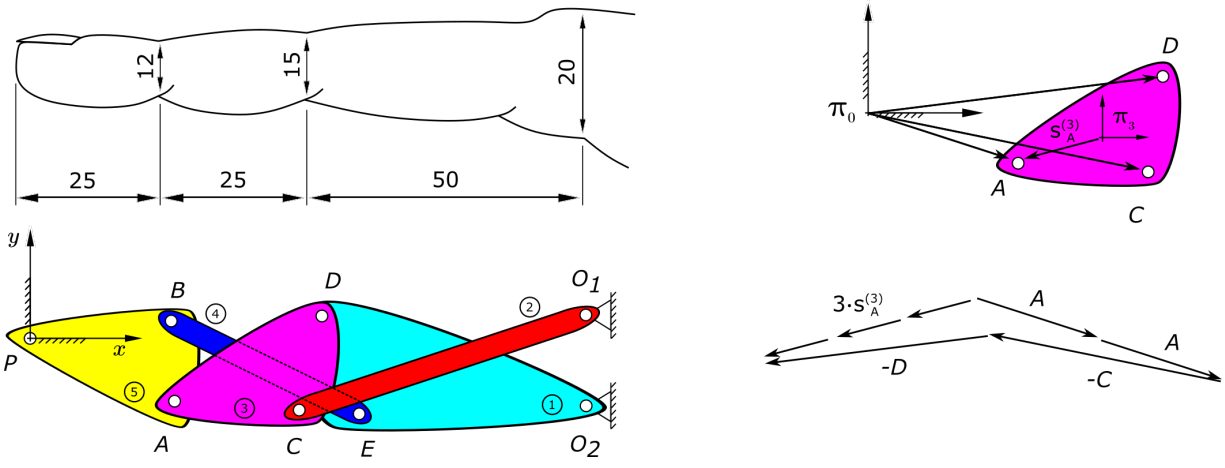


Figure 1: Human-like robotic finger modeled as a planar 6-bar linkage .

The method described in the previous section will be used in order to design human-like robotic finger depicted in Fig. 1. The motion of the finger can be reproduced by employing the *Watt-I* 6-bar linkage that consists of 5 moving bodies: 3 ternary and 2 binary (navy blue and red). Each body is associated with own coordinate frame that is located at its geometric center. Figure 1 shows the mechanism in initial position, which is defined by 8 points: $\boldsymbol{P},\boldsymbol{A}-\boldsymbol{E},\boldsymbol{O}_1,\boldsymbol{O}_2$. The Cartesian coordinates of these points are taken as design variables. Since the analyzed system consists of revolute (pin) joints, the constraint equations for each joint can be expressed as:

$$\boldsymbol{\Phi}^K_{rot}(\mathbf{q}) = \mathbf{r}_i + \mathbf{R}_i \cdot \mathbf{s}^{(i)}_A - \mathbf{r}_j - \mathbf{R}_j \cdot \mathbf{s}^{(j)}_A = \mathbf{0}_{2\times 1}, \tag{10}$$

where $\mathbf{r}_i = [x_i, y_i]^T$ is a vector of Cartesian coordinates of body-fixed coordinate frame with respect to the global reference frame. The quantity $\mathbf{R}_i$ is a rotation matrix associated with absolute angle of revolution φ_i of the i-th body with respect to the global reference frame. The vectors $\mathbf{s}^{(i)}_A, \mathbf{s}^{(j)}_A$ indicate the location of a joint with respect to body-fixed frame. These coordinates are constant for a given design, however the coordinates have to be calculated for each new set of design variables according to the following equations:

$$\mathbf{s}^{(i)}_F = \frac{\boldsymbol{F}-\boldsymbol{G}}{2} - \text{ for binary parts,} \quad \mathbf{s}^{(i)}_F = \frac{2\cdot\boldsymbol{F}-\boldsymbol{G}-\boldsymbol{H}}{3} \text{ - for ternary parts,} \tag{11}$$

where the symbols $\boldsymbol{F},\boldsymbol{G},\boldsymbol{H}$ denote any set of points belonging to certain part, e.g. $\boldsymbol{A},\boldsymbol{C},\boldsymbol{D}$ for the third body or $\boldsymbol{C},\boldsymbol{O}_1$ for the second body. The relation (11) is graphically interpreted in Fig. 1 for ternary body.

Since the analyzed linkage consists of 7 revolute joints, the vector of kinematic constraint equations $\boldsymbol{\Phi}^K$ includes 14 scalar equations. On the other hand the system has 1 degree of freedom, thus it is required to add one driving constraints in the form:

$$\Phi^D = P_x - P_x^{given} = \left[\mathbf{r}_5 + \mathbf{R}_5 \cdot \mathbf{s}^{(5)}_P\right]_x - P_x^{given} = x_5 + s^{(5)}_{Px}\cos(\varphi_5) - s^{(5)}_{Py}\sin(\varphi_5) - P_x^{given} = 0, \tag{12}$$

where P_x is a horizontal Cartesian coordinate of the end-link of the robotic finger. The driving constraint (12) supplements the kinematic constraints (10). Forward kinematic analysis and optimal synthesis of MBS may be thus performed. In the following optimization problem the tip of the finger should pass through a given set of points. Each point represents different configuration of the finger. The coordinates of characteristic points are given as:

- $\boldsymbol{P}_1^{given} = [0,0]^T$ *mm* - fully straighten finger,
- $\boldsymbol{P}_2^{given} = [14.75,-56.5]^T$ *mm* - finger slightly bent,
- $\boldsymbol{P}_3^{given} = [53.5,-81]^T$ *mm* - finger and thumb slightly touch,
- $\boldsymbol{P}_4^{given} = [103.5,-66.5]^T$ *mm* - finger clenched,

whereas the initial values of design variables are given as:

$$\boldsymbol{P}^0 = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}^0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \boldsymbol{A}^0 = \begin{bmatrix} z_3 \\ z_4 \end{bmatrix}^0 = \begin{bmatrix} 25 \\ -5 \end{bmatrix} \quad \boldsymbol{B}^0 = \begin{bmatrix} z_5 \\ z_6 \end{bmatrix}^0 = \begin{bmatrix} 30 \\ 5 \end{bmatrix} \quad \boldsymbol{C}^0 = \begin{bmatrix} z_7 \\ z_8 \end{bmatrix}^0 = \begin{bmatrix} 50 \\ -5 \end{bmatrix}$$

$$\boldsymbol{D}^0 = \begin{bmatrix} z_9 \\ z_{10} \end{bmatrix}^0 = \begin{bmatrix} 55 \\ 5 \end{bmatrix} \quad \boldsymbol{E}^0 = \begin{bmatrix} z_{11} \\ z_{12} \end{bmatrix}^0 = \begin{bmatrix} 60 \\ -5 \end{bmatrix} \quad \boldsymbol{O}_1^0 = \begin{bmatrix} z_{13} \\ z_{14} \end{bmatrix}^0 = \begin{bmatrix} 95 \\ 5 \end{bmatrix} \quad \boldsymbol{O}_2^0 = \begin{bmatrix} z_{15} \\ z_{16} \end{bmatrix}^0 = \begin{bmatrix} 95 \\ -5 \end{bmatrix}$$

There are several bound constraints imposed on the values of design variables to provide satisfactory convergence of the Newton - Raphson procedure. The imposed bound constraints do not, however, guarantee that the MBS is movable. If such a case occurs, penalty function is applied for a given set of design variables, that forces an optimization algorithm to discard current set of design variables and search for the new set that allows the mechanism to move.

The objective function is defined as a maximum quadratic deviation between the vertical coordinate P_y^k of the tip of the finger and the actual vertical coordinate y_P^k of that point calculated from the kinematic model. Since there are 4 characteristic points, the grid is defined as: $\alpha_k = P_x^k$, where $k = \{1,2,3,4\}$, then:

$$\min_{\mathbf{z}\in\mathscr{Z}} f(\mathbf{z},\mathbf{q}) = \min_{\mathbf{z}\in\mathscr{Z}} \left[\max_{\alpha_k} \left(y_P^k - P_y^k \right)^2 \right]. \tag{13}$$

In order to get rid of the complex min-max problem (13), new design variable z_{17} and additional inequality constraint are introduced:

$$g(\mathbf{z}) = \left[\max_{\alpha_k} \left(y_P^k - P_y^k \right)^2 \right] - z_{17} \leqslant 0 \ , \tag{14}$$

where y_P coordinate of point $\boldsymbol{P}$ is calculated as:

$$y_P = y_5 + s_{Px}^{(5)} \sin(\varphi_5) + s_{Py}^{(5)} \cos(\varphi_5).$$

There are 14 kinematic constraints imposed on the system together with one driving constraint. The constraint equations can be written in the form:

$$\boldsymbol{\Phi}(t,\mathbf{q}) = \begin{bmatrix} \boldsymbol{\Phi}_{14\times 1}^{rot}(\mathbf{q}) \\ \\ \Phi^D(t,\mathbf{q}) \end{bmatrix} = \mathbf{0}_{15\times 1}, \tag{15}$$

Gradient of inequality constraint (16) can be calculated according to equation (7):

$$\nabla_{\mathbf{z}} g = \frac{\partial g}{\partial \mathbf{q}} \cdot \left[-\left(\frac{\partial \boldsymbol{\Phi}}{\partial \mathbf{q}} \right)^{-1} \frac{\partial \boldsymbol{\Phi}}{\partial \mathbf{z}} \right] + \frac{\partial g}{\partial \mathbf{z}} \ . \tag{16}$$

The terms $g_{\dot{\mathbf{q}}}, g_{\ddot{\mathbf{q}}}$ are equal to zero, since there is no dependency of g on velocity or acceleration. Let us denote y_P^{max}, P_y^{max} as maximum values of y_P^k, P_y^k taken over the set $\{k = 1 \cdots 4\}$ for a given design. The partial derivatives of g are expressed as:

$$\frac{\partial g}{\partial \mathbf{q}} = 2\left(y_P^{max} - P_y^{max} \right) \cdot \begin{bmatrix} \mathbf{0}_{13\times 1} \\ 1 \\ s_{Px}^{(5)} \cos(\varphi_5) - s_{Py}^{(5)} \sin(\varphi_5) \end{bmatrix}_{1\times 15}^{T} \tag{17}$$

$$\begin{aligned} \frac{\partial g}{\partial \mathbf{z}} = \Big[& \frac{2}{3}\sin(\varphi_5)\cdot p, \frac{2}{3}\cos(\varphi_5)\cdot p, -\frac{1}{3}\sin(\varphi_5)\cdot p, -\frac{1}{3}\cos(\varphi_5)\cdot p, \\ & -\frac{1}{3}\sin(\varphi_5)\cdot p, -\frac{1}{3}\cos(\varphi_5)\cdot p, \mathbf{0}_{1\times 11}, -1 \Big]_{1\times 17} \ , \end{aligned} \tag{18}$$

where an additional variable p is defined as $p = 2\left(y_P^{max} - P_y^{max}\right)$. The Jacobian matrix $\mathbf{\Phi_q}$ can be automatically calculated in a straigthforward manner for a given MBS. The calculation of the crucial Jacobian matrix $\mathbf{\Phi_z}$, however, will be briefly explained. The vector of constraint equations $\mathbf{\Phi}$ involves seven 2×1 entries in the form of Eq. (10) (due to seven joints in the system). The derivative of $\mathbf{\Phi}$ with respect to the vector of design variables $\mathbf{z}$ can be written as:

$$\mathbf{\Phi}_{\mathbf{z}}^{i,j} = \mathbf{R}_i \cdot \frac{\partial \mathbf{s}_F^{(i)}}{\partial \mathbf{z}} - \mathbf{R}_j \cdot \frac{\partial \mathbf{s}_F^{(j)}}{\partial \mathbf{z}}. \tag{19}$$

Kinematic constraint vector $\mathbf{\Phi}$ is herein a linear function of design variables, and $\frac{\partial \mathbf{s}_F^{(i,j)}}{\partial \mathbf{z}}$ denote a matrix of coefficients obtained by differentiating equation (11) with respect to $\mathbf{z}$. For binary parts, we have:

$$\frac{\partial \mathbf{s}_F^{(i)}}{\partial \mathbf{z}} = \begin{bmatrix} \mathbf{0}_{2\times a} & \frac{1}{2} \cdot \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times b} & -\frac{1}{2} \cdot \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times c} \end{bmatrix}_{2\times 17} \tag{20}$$

and for ternary parts, we get:

$$\frac{\partial \mathbf{s}_F^{(i)}}{\partial \mathbf{z}} = \begin{bmatrix} \mathbf{0}_{2\times a} & \frac{2}{3} \cdot \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times b} & -\frac{1}{3} \cdot \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times c} & -\frac{1}{3} \cdot \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times d} \end{bmatrix}_{2\times 17} , \tag{21}$$

where $\mathbf{I}$ is a unity matrix and F denotes any of revolute joint definition points (see Fig. 1). The derivative $\Phi_{\mathbf{z}}^D$ is obtained using the same approach as for $\mathbf{\Phi}_{\mathbf{z}}^K$, however, let us recall that driving constraint is a scalar equation and only x component of the derivative needs to be calculated (see Eq. (12)). The indices $a - d$ determine non-zero entries in equations (20) and (21), which vary for each considered point. Equations (20) and (21) allow us to calculate $\mathbf{\Phi_z}$, which together with the original constraint Jacobian matrix $\mathbf{\Phi_q}$ can be substituted back into (16) to yield demanded design sensitivity vector.

2.2. Numerical results

Multiple numerical experiments have been performed in Matlab Optimization Toolbox. The design sensitivity vector has been compared with the same quantity calculated by using finite difference method provided by the software. Both solutions and their convergence rates were almost identical, however the proposed method found an optimal point in a slightly less number of iterations as indicated in Fig. 2. The finite difference method has also the tendency to produce poorly conditioned gradients at certain iterations as it is observed in Fig. 3.

Moreover, different optimization algorithms were tested, i.e.: *Active Set (AS), Interior Point (IP)* and *SQP*. Figure 4 shows trajectories of point $\boldsymbol{P}$ for each optimization method in the solution process as well as for initial design $\mathbf{z}_0$. Maximum absolute deviation between the location of characteristic points $\boldsymbol{P}_1$, $\boldsymbol{P}_2$, $\boldsymbol{P}_3$, and $\boldsymbol{P}_4$ and the calculated trajectories can be evaluated for each solution as demonstrated in table 1. Figure 5 shows the robotic finger at different configurations. The results were produced by SQP solver and demonstrate the verification of the method proposed in the paper.

Table 1: Absolute deviations $\Delta_{Solution} = |y_P^i - P_y^i|$.

Design	$\Delta_{Solution}$	Active point (i)
$\mathbf{z}_0$	14.07	4
$\mathbf{z}_{SQP}$	5.5	2
$\mathbf{z}_{IP}$	4.27	2
$\mathbf{z}_{AS}$	4.27	2

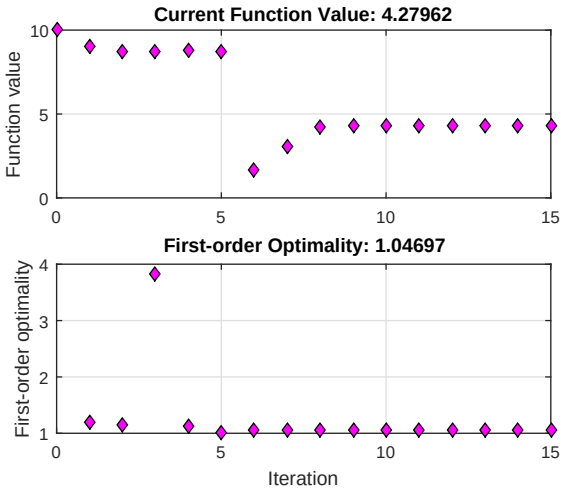


Figure 2: The performance of the SQP algorithm with analytical gradient supplied by the designer.

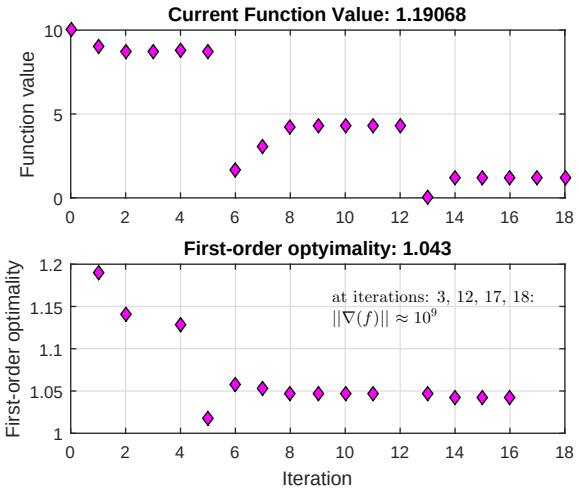


Figure 3: The performance of the SQP algorithm with gradient calculated by finite difference method.

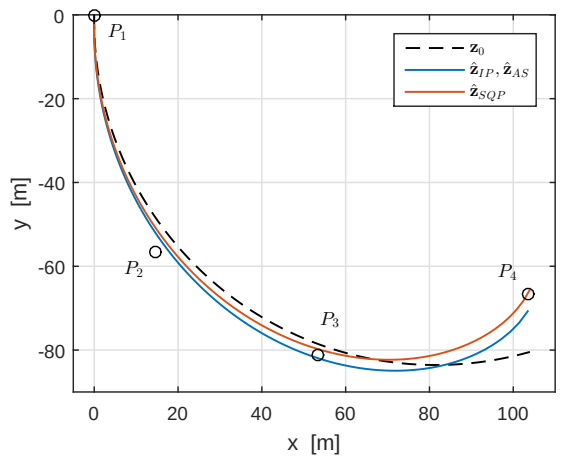


Figure 4: Trajectories of point ***P*** for initial design and optimized cases.

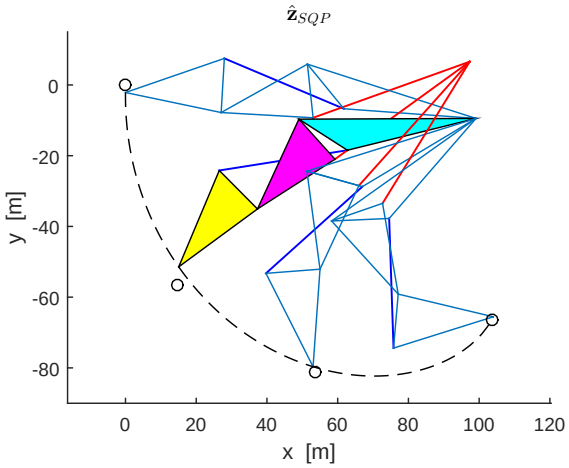


Figure 5: The snapshot from the simulation of the optimized mechanism (SQP method employed).

3. Conclusions and recommendations for futher research

In this paper we give a brief insight to multibody systems' synthesis, as well as appropriate formulation of the optimization problem for such task. Design sensitivity analysis has been addressed and systematically introduced with the prospective aim to develop general software tools for optimal synthesis of mechanical systems. Upper bound technique is presented and subsequently utilized throughout the solution of optimization problem. Numerical tests are

presented for the case of human-like robotic finger. The results from the optimization procedure are delivered in the text. The initial design of the robotic finger has been successfully improved to the extent possible by employing the method described herein.

The developed optimization technique may be a powerful tool for kinematic synthesis and design of multibody systems. There are a number of areas that should be further investigated in order to exploit the potential of the method. One recommendation is to make the techniques described herein more robust for the case of singular configurations and lock-up positions, which may severely limit the workspace of the mechanism. The second recommendation is to extend and implement the theory for fully three-dimensional mechanical systems that would further generalize the presented optimization technique.

Acknowledgements. This work has been supported by the Faculty of Power and Aeronautical Engineering at the Warsaw University of Technology through the statutory funds for young scientists.

References

1. Chadaj, K., Malczyk, P., Frączek, J.: "A Parallel Recursive Hamiltonian Algorithm for Forward Dynamics of Serial Kinematic Chains", *IEEE Transactions on Robotics*, 2017, Vol. 33, no. 3, February, DOI: 10.1109/TRO.2017.2654507

2. Frączek, J., Surowiec, M., Wojtyra, M.: "Multibody Modelling of a Tracked Robot's Actuation System", *Aerospace Robotics*, Book Series: *GeoPlanet-Earth and Planetary Sciences*, *I Conference on Robotics in Aeronautics and Astronautics*, Warsaw, March, 2013, DOI: 10.1007/978-3-642-34020-8_7

3. Frączek J., Wojtyra M.: "Application of General Multibody Methods to Robotics", *Multibody Dynamics: Computational Methods And Applications*, Book Series: *Computational Methods in Applied Sciences*, *Multibody Dynamics 2009*, Warsaw, June, 2009, DOI: 10.1007/978-90-481-9971- 6_7

4. Haug E.:, "Computer Aided Analysis and Optimization of Mechanical System Dynamics", Springer-Verlag Berlin Heidelberg, *Berlin*, 1984

5. McCarthy J., Soh G.: "Geometric design of linkages", Springer Science & Business Media, *London*, 2010

6. Mianowski K., Wojtyra M.: "Virtual prototype of a 6-DOF parallel robot", *11th World Congress In Mechanism And Machine Science*, Tianjin (China), April, 2004

7. Pękal M., Frączek J.: "Comparison of selected formulations for multibody system dynamics with redundant constraints", *Archive of Mechanical Engineering*, 2016, Vol. 63, no. 1, April, DOI: 10.1515/meceng-2016-0005

8. Pękal M., Frączek J.: "Comparison of natural complement formulations for multibody dynamics", *Journal of Theoretical and Applied Mechanics*, 2016, vol. 54, no. 4, January, DOI: 10.15632/jtam-pl.54.4.1391

9. Soh G., McCarthy J.: "The synthesis of six-bar linkages as constrained planar 3R chains", *Mechanism and Machine Theory* , 2008, vol. 43, no. 1, February

10. Woliński, Ł., Malczyk, P.: "Dynamic Modeling and Analysis Of a Lightweight Robotic Manipulator in Joint Space", *The Archive of Mechanical Engineeering*, 2015, Vol. 62, no. 2, January

Rastislav MOTUZ[1], Petr DREXLER[2]

Theoretical investigation of the novel model dedicated for linear birefringence compensation in fiber optic current sensors

In this paper we propose novel model of linear birefringence compensation in fiber optic current sensors, on theoretical basis. As a tool of theoretical analysis, Jones calculus was used. It has been shown that the proper concatenation and orientation of half wave plates allows the avoidance the usage of Faraday rotator.

1. Introduction

In fiber optic current sensors (FOCS), the presence and effect of linear birefringence must be considered. This can be observed in the fiber as a result of its original anisotropy (so-called intrinsic birefringence), or it can be created in the fiber by external influences (induced birefringence). Intrinsic and induced linear birefringence can cause undesired change in the polarization state of the light wave, which in case of a polarimetric evaluation of the light wave will cause sensitivity to decrease [1].

Linear birefringence is often the most critical parameter for sensor functionality in practical applications, so the appropriate attention must be given to it. So far, several techniques have been developed to suppress or compensate the effect of linear birefringence. The paper is focused on the linear birefringence compensation via orthogonal conjugation of modes.

2. Compensation by orthogonal conjugation

The linear birefringence in optical fiber is created when one of the optical modes spreads along so-called fast axis (refractive index n_x), and the other mode spreads along slow axis (refractive index n_y) of the optical fiber. The orthogonal conjugation of the modes represents the exchange of modes in the sense of orientation of their polarization and the re-propagation of the optical wave by the same optical fiber section. As it follows from the name of the compensation method, to achieve orthogonal conjugation of the modes is necessary to twist the polarization plane by the angle $\alpha = 90^\circ$ [2].

The authors in [3] describe a compensating method of linear birefringence using a loop that includes a polarization beam splitter (PBS) and a polarization maintaining fiber (PMF). PBS defines linear states of polarization of the optical wave in the loop. In the loop, the Faraday rotator 90° (FR 90°) is positioned as a compensatory element of a linear birefringence, ensuring orthogonal conjugation of modes. The situation is illustrated by the figure 1.

Input state of polarization (SOP) is divided into two orthogonal SOPs by PBS. FR 90° causes twisting of each SOP under the angle of 90°. Then the fractions of the light wave pass again

[1]Ph.D. student, Brno University of Technology, Technická 12, 616 00 Brno, CZECH REPUBLIC, tel. +420 778 723 429, e-mail: xmotuz00@feec.vutbr.cz

[2]Ph.D., Brno University of Technology, Technická 12, 616 00 Brno, CZECH REPUBLIC, tel. +420 777 497 184, e-mail: drexler@feec.vutbr.cz

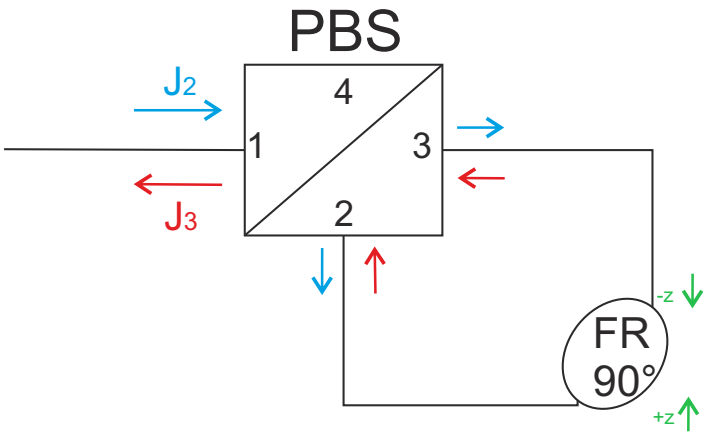


Figure 1: Linear birefringence compensation method using loop, including PBS and FR90° [3].

through the PBS and the light wave is being re-composed. PBS is described by the Jones calculus according to [3] subsequently:

$$\mathrm{P}_{12} = \mathrm{P}_{21} = \mathrm{P}_{34} = \mathrm{P}_{43} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \mathrm{P}_{13} = \mathrm{P}_{31} = \mathrm{P}_{24} = \mathrm{P}_{42} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \tag{1}$$

where P_{xy} is a sign for ports of PBS; index x means input and y output ports.

An overall Jones matrix of FR 90° and PBS according to [3] is:

$$\mathrm{FR}_{90^\circ} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{2}$$

Another and more often used technique of linear birefringence compensation uses Faraday mirror (FM), which is the concantenation of the Faraday rotator 45° and a plane mirror. This system can be seen in the figure 2.

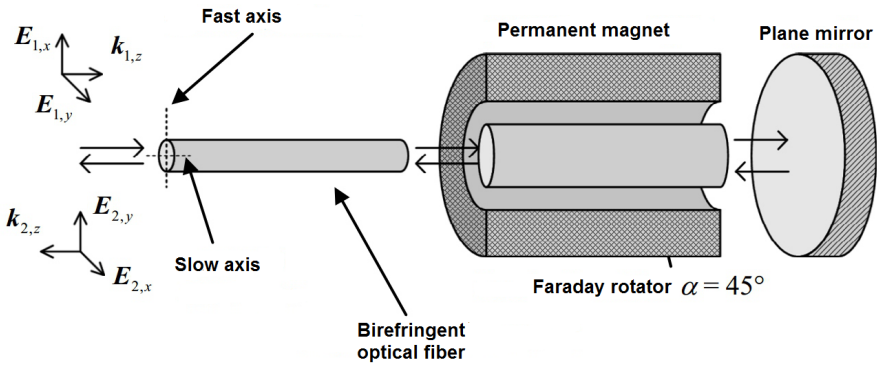


Figure 2: Principle of the Faraday mirror [4].

After the first pass of the wave through FM, the plane of polarization will be twisted by an angle $\alpha = 45°$ and a reflection from the mirror will occur. At the FM output, the plane of the

polarization is, due to the reciprocity, twisted by an angle $\alpha = 90^\circ$ in relation to the input [5]. The resulting effect is identical as in the previous solution.

3. Novel solution of linear birefringence compensation

Jones notation of FR 90°, which can be seen in the figure 1, can be understood as so-called rotation matricies of two oposit directions, +z and -z, subsequently:

$$\mathrm{FR}_{90^\circ+z} = \begin{bmatrix} \cos 90^\circ & \sin 90^\circ \\ -\sin 90^\circ & \cos 90^\circ \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \text{ and} \tag{3}$$

$$\mathrm{FR}_{-90^\circ-z} = \begin{bmatrix} \cos(-90^\circ) & \sin(-90^\circ) \\ -\sin(-90^\circ) & \cos(-90^\circ) \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \tag{4}$$

The idea is to reach the orthogonal conjugation using two half-wave plates (HWP), which are oriented under the angle of -45° to each other. Its principle is shown in the figure 3. The advantage of this method is the elimination of the temperature dependence of FR and its sensitivity to a strong external magnetic field.

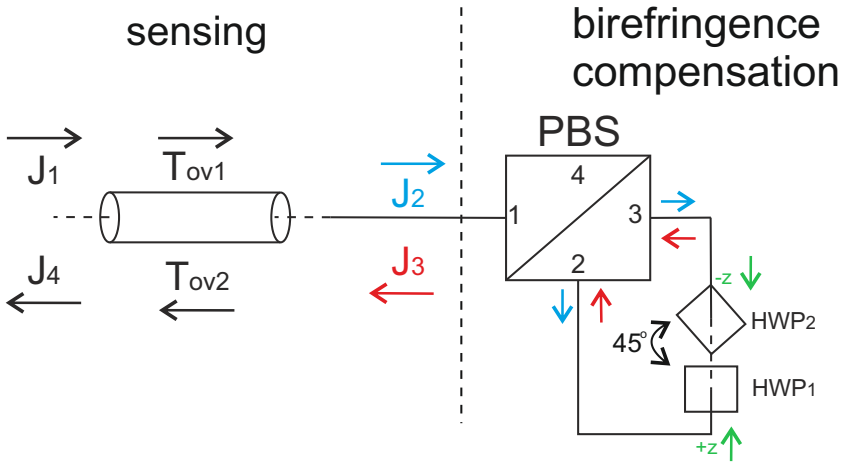


Figure 3: Linear birefringence compensation using half-wave plates.

For the direction +z is the Jones notation following:

$$[\mathrm{OUT}] = [45^\circ] \cdot [\mathrm{HWP}_2] \cdot [-45^\circ] \cdot [\mathrm{HWP}_1] \cdot [\mathrm{IN}], \tag{5}$$

where $[-45^\circ]$ denotes the angle between HWPs and the element $[45^\circ]$ is included due to the compensation of the rotation of the coordinate system to the HWP axes. [IN] and [OUT] notation represents vectors, respectively the states of polarization at the input of HWP_1 or at the output of HWP_2 and vice versa.

Full Jones notation:

$$[\mathrm{OUT}] = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot [\mathrm{IN}]. \tag{6}$$

Simplified notation:

$$[\mathrm{OUT}] = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \cdot 2 \cdot [\mathrm{IN}]\,. \tag{7}$$

Notation (7) equals to the notation (3) (direction +z) with the difference in term 2, representing so-called global phase, which can be for the purpose of the analysis neglected.

For the direction -z is HWP_2 turned under the angle of 45°:

$$[\mathrm{OUT}] = [-45^\circ] \cdot [-45^\circ] \cdot [\mathrm{HWP}_1] \cdot [45^\circ] \cdot [\mathrm{HWP}_2] \cdot [45^\circ] \cdot [\mathrm{IN}]\,. \tag{8}$$

Full Jones notation:

$$[\mathrm{OUT}] = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \cdot [\mathrm{IN}]\,. \tag{9}$$

Simplified notation:

$$[\mathrm{OUT}] = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \cdot 4 \cdot [\mathrm{IN}]\,. \tag{10}$$

Notation (10) equals to the notation (4) (direction -z). Thus it can be assumed that the functionality of FR 90° can be substituted by the concatenation of two half-wave plates. It has been proved that the orientation of particular half-wave plates is not important, but the preservation of the angle $\theta = 45^\circ$ between HWPs is shown to be critical.

4. Example of calculation

The dextrorotatory circular polarization described by the Jones vector $\vec{J}_2$ (figure 3) was chosen as an input polarization: $\begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix} \cdot \frac{1}{\sqrt{2}}$.

4.4. Direction +z

State of polarization after the port P_{12}:

$$\begin{bmatrix} 0 \\ -\mathrm{i} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix} \cdot \frac{1}{\sqrt{2}}. \tag{11}$$

State of polarization after the light wave passing HWP_1 and HWP_2 (7):

$$\begin{bmatrix} -\mathrm{i} \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ -\mathrm{i} \end{bmatrix}. \tag{12}$$

State of polarization after the back-passing PBS (port P_{31}):

$$\begin{bmatrix} -\mathrm{i} \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} -\mathrm{i} \\ 0 \end{bmatrix}. \tag{13}$$

From (13) can be noticed that back-passing wave through PBS does not affect previous SOP.

4.4. Direction -z

State of polarization after the port P_{13}:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix} \cdot \frac{1}{\sqrt{2}}. \tag{14}$$

State of polarization after the light wave passing HWP_2 and HWP_1 (10):

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{15}$$

State of polarization after the back-passing PBS (port P_{21}):

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{16}$$

5. Evaluation

PBS divides input SOP of the light wave into two, orthogonal and linear SOPs. In the loop, the orthogonal conjugation occurs and the resultant modes (SOPs) will merge into the final one. This output polarization $\vec{J_3}$ should be orthogonal to the original one - input $\vec{J_2}$.

The light waves from both directions, with their SOPs (13) and (16) will merge and the resultant light wave is described by the Jones vector:

$$\begin{bmatrix} -\mathrm{i} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -\mathrm{i} \\ 1 \end{bmatrix}. \tag{17}$$

Due to the generally established convention (according to [3]) the notation of the coordinate system is changed for the backward propagating wave; **z** goes to **-z** and **x** to **-x**. In this case, the resulting vector (17) will be modified as follows:

$$\begin{bmatrix} -\mathrm{i} \\ 1 \end{bmatrix} \Longrightarrow \begin{bmatrix} \mathrm{i} \\ 1 \end{bmatrix}. \tag{18}$$

Verifying that vectors $\vec{J}_2$ and $\vec{J}_3$ are orthogonal:

$$\begin{bmatrix} 1 \\ -\mathrm{i} \end{bmatrix} \cdot \begin{bmatrix} \mathrm{i} \\ 1 \end{bmatrix} = 1 \cdot \mathrm{i} + (-\mathrm{i}) \cdot 1 = 0 \tag{19}$$

Above stated theoretical analysis offers the capability of the practical usage of the proposed method in the field of linear birefringence compensation. In the new concept of sensor, it is possible to avoid using FR.

6. Conclusion

In this paper, the novel model of linear birefringence compensation in fiber optic current sensors, has been proposed. The theoretical analysis was based on the Jones calculus which shows the capability of this compensation method to be used in practice. The use of properly oriented half-wave plates gives an opportunity of avoidance the unadvantages showed by Faraday rotator. Further research will be oriented to practical measurements in order to validate proposed compensation technique.

7. Acknowledgement

The research was carried out under support of grant No. LO1401 covered by the National Sustainability Programme. The infrastructure of the SIX Center was utilized to facilitate the research and the assistance provided by the general student development project being executed at Brno University of Technology.

References

1. Day G. W., Rose A. H.: ”Faraday effect sensors” *The state of the Art. SPIE*, 1988, vol. 985

2. Drexler P., Fiala P.,: ”Suppression of polarimetric birefringence effect in optical fiber and its application for pulsed current sensing.”, *International Waveform Diversity and Design Conference*, 2009

3. Esman, R.D., Marrone M.J.: ”Passive elimination of polarization sensitivity of fiber-optic microwave modulators.” *IEEE Transactions on Microwave Theory and Techniques*, vol. 43, no. 9

4. Drexler P.: ”Techniky potlačování dvojlomných jevů v optických vláknech: Techniques for optical fiber birefringence suppression” *teze habilitační práce*, VUTIUM, Brno, 2010

5. Drexler P., Fiala P.: ”Utilization of Faraday mirror in fiber optic current sensors” *Radioengineering*, 2008, vol. 17, no. 4

Rolando INGLÉS[1]

A C++ Shared-Memory Inter-Process Communication Framework for High-Throughput Monitoring Systems

High-Throughput Monitoring Systems are required in scientific experiments since a large amount of data is produced, becoming thus a challenge for developing software applications capable of managing such bulk of data promptly and accordingly. The paper presents the experimental evaluation of a C++ framework for inter-process communication in order to demonstrate the effectiveness of shared-memory as a mechanism for data transferring and processes synchronization when independent processes are involved in high-throughput monitoring systems. The assessment of the effectiveness is based on the latency measurement when the data is placed into the shared-memory and measuring the longest time when such data is gathered. According to the results, it is confirmed that it is possible to develop a C++ framework for programming high-throughput monitoring system applications.

1. Introduction

This paper presents the effectiveness evaluation of a C++ Shared-Memory Inter-Process Communication Framework following the requirements established in the ITER CODAC framework [1] specifications. Such specifications establish the design of a C++ framework for supporting the development of monitoring applications and control systems for the TOKAMAK device [2]. These control systems involve the use of no less than 300 high-speed cameras [3] implying that a large-scale amount of data is produced and acquired in real-time through high-speed digital interfaces, e. g., Camera Link, PCIe or GbE [4]. Once the images are digitized for specialized grabber boards [5], the images must be transferred to the control systems for a proper processing. The transmission must be done in the fastest manner possible. In Figure 1 is schematized the High-Throughput Monitoring System workflow of the ITER project [6].

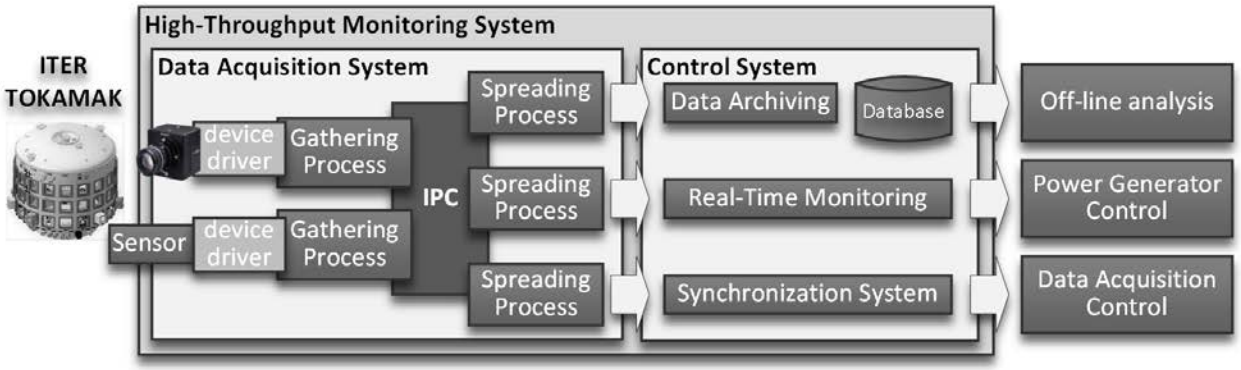


Fig. 1: ITER's High-Throughput Monitoring System.

[1] Ph.D. student, Lodz University of Technology, Department of Microelectronics and Computer Science, ul. Wólczanska 221/223 budynek B18, 90-924 Łódź, POLSKA,
tel. +48 575 182 638, e-mail: roling@dmcs.pl

During the operation, this kind of systems (see Figure 1) require an efficient mechanism of data communication among the involved processes. Additionally, the concurrent access to the data must be controlled by specialized mechanisms [7]. Having into account that each sub-system represents an independent process, the C++ framework evaluated in this paper uses shared-memory as the main mean of communication among these processes. By using shared-memory offers the advantage that the data are not copied; meaning that once the data are placed in the shared-memory region, the data are available for all those processes that have mapped such region properly [8]. Indeed, the shared-memory technology it was intended for providing high-speed inter-process communication among cooperating processes [9]. For managing the amount of data produced by the devices abovementioned, the processes must run concurrently, preferably one process running is its own CPU [10]. Regardless of this, the concurrent access to the shared-memory should be synchronized by using specialized mechanisms of control and thus elude the race-conditions incidents [11].

2. The High-Throughput Monitoring System

For monitoring the *TOKAMAK* [1], the ITER project makes use of the *EoSens® 3CL Full CL MC3010* high-speed cameras. Figure 2 illustrates the possible configurations of such cameras according to the technical documentation provided by the manufacturer [12].

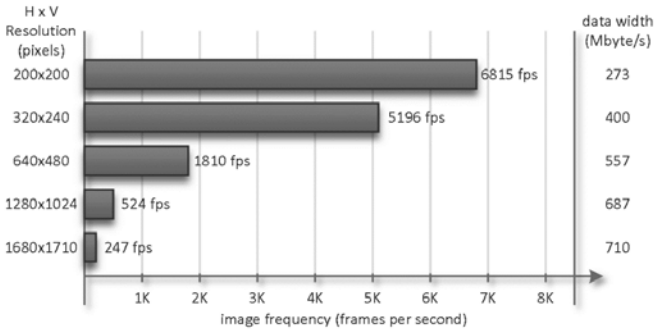


Fig. 2: Frameset factory profiles of EoSens® 3CL Full CL MC3010 high-speed camera [12].

Monitoring and controlling tasks are dependent on the physical time when the images are produced, thus the Control System runs in a real-time environment [13]. Such tasks require that the acquired images be sent prompt toward two sub-systems namely, *Data Archiving* and *Synchronization System*. The block diagram in Figure 3 outlines how the data are acquired and transferred to the two mentioned previously sub-system.

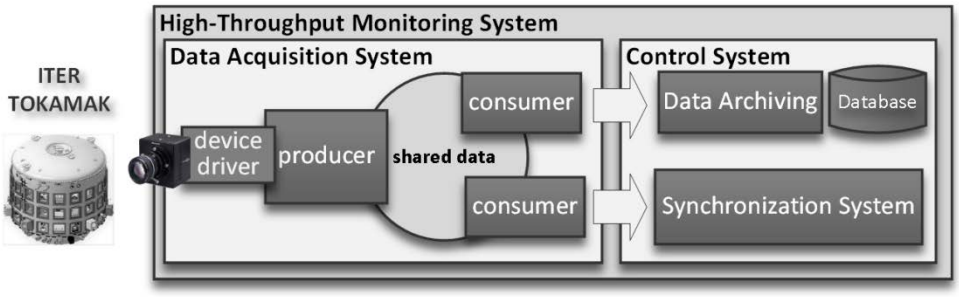


Fig. 3: High-Throughput Monitoring System Block Diagram.

The *Data Archiving* sub-system performs not critical storing data tasks; consequently, this system is considered a soft real-time system. Nonetheless, the *Synchronization System* hard real-time system since its operations are critical for monitoring and controlling of the TOKAMAK [13]; based on the foregoing, a 100% of accuracy of the data transfer it is required. Furthermore, a 100% of reliability is mandatory to evaluate the quality and consistency of the transferred data.

An additional aspect depicted in the schema of Figure 3 is the *producer/consumer* approach implemented with the C++ framework, where the *producer* process is in charge of getting the images from the cameras, and placing them into the shared-memory region. Thus, the *consumers* are responsible of taking such images and process them. During this transferring process, the producer conducts special signaling process in order to notify the *consumers* when new image have arrived. Additionally, each *consumer* notifies the *producer* when the images are gathered.

3. Related Works

An important component of modern computer operating systems is the inter-process communication (IPC) mechanisms. Nevertheless, these mechanisms are available as low-level operating system calls; precisely this fact becomes the primordial motivation for developing a framework that encloses the details of performing such calls and thus provide a user-friendly interface for developing applications software oriented to high-performance data transfer [15]. One of these frameworks is the *FastFlow*. This is a C++ framework for parallel processing advocating high-level and pattern-based parallel programming [16].

The *ADAPTIVE* Communication Environment (ACE) is a toolkit that implements patterns for communication software [17]. This framework is object-oriented and aimed to develop network applications involving concurrency and inter-process communication. Additionally, the *Dynamic Shared-Memory Allocator* (DSMA) provides IPC mechanisms to ensure communication, data transfer and synchronization among different processes [18]. Nevertheless, this framework uses virtual-memory paging mechanism, which is implemented in most operating systems and allows using a virtual-memory space larger than the available physical system memory.

4. The C++ Framework

The C++ framework under consideration uses the templates feature included in the C++ language. By using this C++ feature, the framework is capable of working with generic data types. Notwithstanding, the use-case for the experimental evaluation conducted in this papers is related with images produced by high-speed cameras above mentioned.

Figure 4 outlines how the framework is designed, and as the concept of templates included in C++ has been used. Likewise, as the operating systems calls for managing the shared-memory segments have been wrapped in the `shared resource interface class`. This particular class is in charge of creating/opening each shareable work region making it accessible for those processes that refer to it.

Developing application using the framework under evaluation requires create objects referring the `producer` and `consumer` classes for creating the producer and consumer programs respectively.

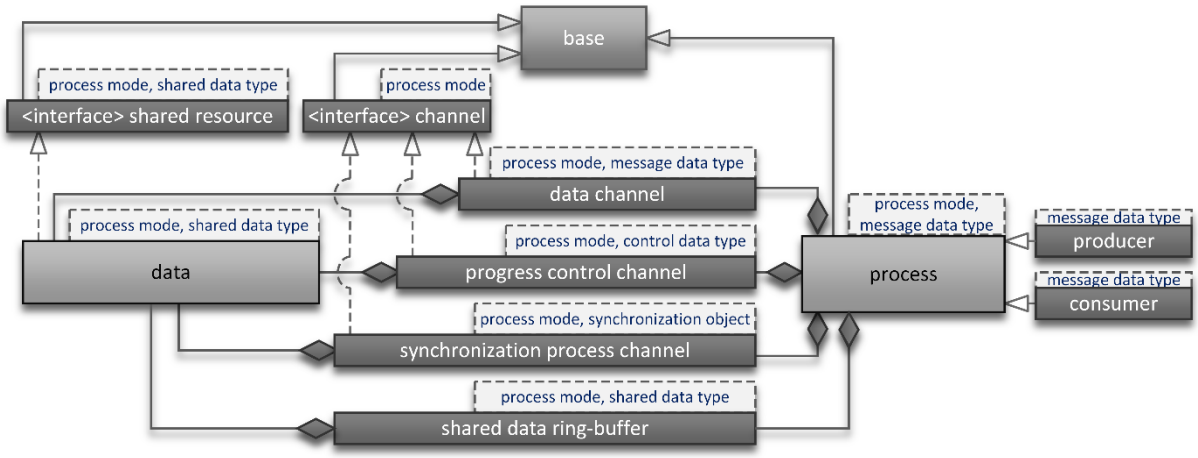


Fig. 4:C++ Framework Physical Design.

5. The Experimental Evaluation

The experimental evaluation was performed in a computer system with sixteen CPUs Intel® Xeon® CPU C5549 @ 2.53GHz., running the Red Hat Enterprise Linux® operating system, release 6.5 based on the Linux kernel version 2.6.32-431.20.3.el6.x86_64. Additionally, the computer system was equipped with 24GB of RAM memory.

The images production by the cameras was simulated by using a data file containing 7,348 frames with 200x200 pixels resolution each one. With the aim of making efficient the access of the input data, the producer process maps such input file into memory by using of *mmap* system call [19] before of entering into the main loop.

Figure 5 depicts how the timing process was conducted; the *producer* reads sequentially an input file containing testing frames and waits until a slot is available in the shared ring-buffer. On the other hand, the consumers wait until a new frame is present in the shared ring-buffer. It is important to point out that each frame is uniquely identified, first for checking that all frames were sent accurately, and second, as a method for helping the latency calculation process.

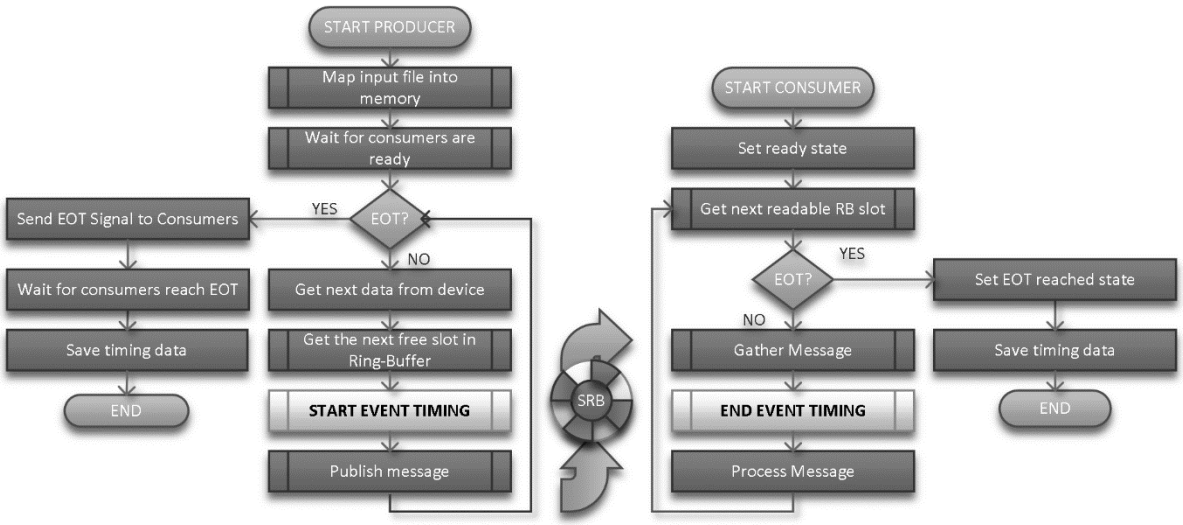


Fig. 5: Timing Algorithm

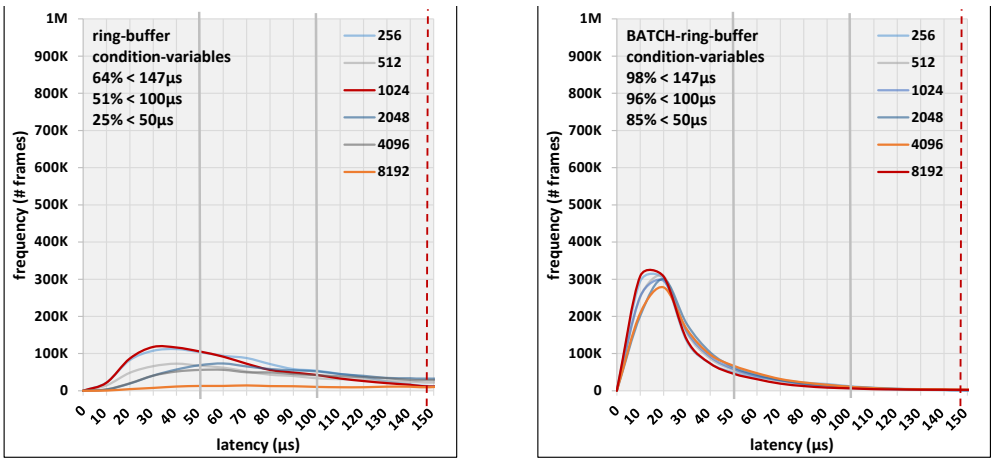
The timing for an frame starts when the *producer* calls the *START_EVENT_TIMING* sub-routine, within which the system function *clock_gettime(CLOCK_REALTIME,&timespec)* is called, this system function uses the *CLOCK_REALTIME* parameter in reference to the system-wide real-time clock and the *&timespec* parameter for storing the clock value. Subsequently, the *producer* publish the message by notifying to the *consumer* processes that a new frame has been placed into the shared-memory region and it is ready to be taken. The *consumer* processes stays in waiting mode until they receive the signal from the producer that a new frame has been published. Once the *consumer* takes the frame, the *END_EVENT_TIMING* is called, within which the *clock_getime* function is used for establishing when the message has been gathered.

The timing process is finished when the *producer* send the *end_of_transmission (EOT)* flag into *consumers* and wait until the *consumer* processes finishing. Likewise, when the *consumer* processes detect the *EOT* flag, the timing process is finished by dumping the timing data into a local file. The next step is to process the timing data files by checking each measurement using the identification of each frame. The starting time is subtracted from each stopping time and the greatest difference is taken as the latency time for the frame under consideration.

6. RESULTS

The effectiveness of the C++ framework was evaluated taking as success barrier the 147µs latency time. This specific latency time was obtained from the factory configuration of the cameras under consideration. It means that, when the high-speed camera is configured to produce frames with 200x200 pixels of resolution, the camera is capable of generating 6,815 frames per second. In other terms, the cameras produce a new frame each 1/6,815 part of a second. Based on this fact, a latency measurement under 147µs is considered success otherwise it is considered as an error in the transferring. It is worth mentioning that the 147µs barrier is represented in the following plots as a dotted red line. Furthermore, each scatter line represents the results obtained for different numbers of slots, starting from 256 slots and ending in 8,192 slots.

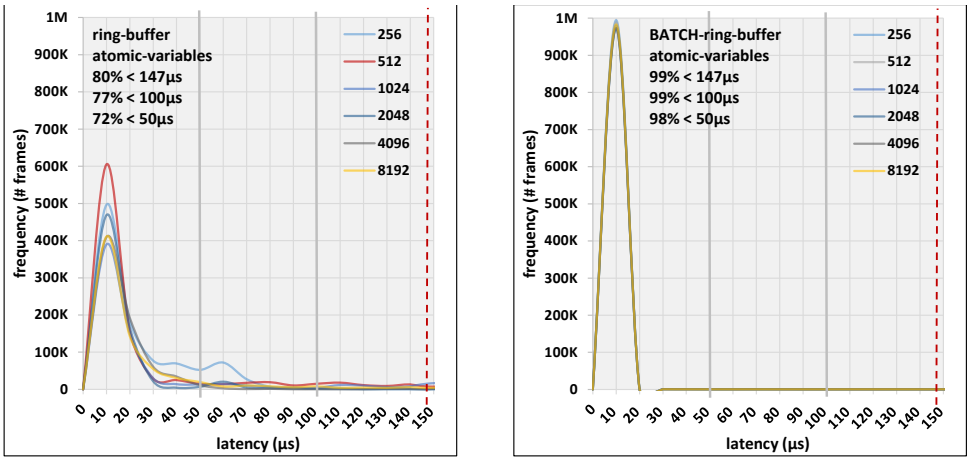
Figure 6 reports the experimental results when the C++ framework uses condition-variables as the mechanism for synchronizing the frame publishing and consuming processes. The latency results plotted in Figure 6(a) shows how only 64% of effectiveness was reached by using conditional-variables together with a typical ring-buffer algorithm of consuming. In order to improve the results in Figure 6(a), a new algorithm of batch-consuming it was introduced into the framework. By using the batch-consuming algorithm, the consumers takes a set of slots to be read instead of taking slot-by-slot. The plot presented in Figure 6(b) illustrates the results obtained by using batch-consuming approach, reporting 98% of effectiveness in the latency times. While improving the latency times was important, still a 2% of frames were sent above the 147µs established barrier.



(a) (b)

Fig. 6: Condition-Variables Latency Measurement Results

The results reported in Figure 7 were obtained changing the way as the processes of publishing and consuming were synchronized. Instead of using condition-variables, atomic-variables were used as synchronization mechanism. As well as the evaluation of condition-variables, the usage of atomic-variables was evaluated for slot-by-slot and batch-consuming algorithms. Figure 7(a) outlines the results when the combination slot-by-slot and atomic-variables is used obtaining a 80% of effectiveness. While these results represent an improvement in the slot-by-slot approach, a 20% of ineffectiveness was still obtained. The best results were obtained by using the batch-consuming algorithm synchronized with atomic-variables. These results reported in Figure 7(b) show that 99% of effectiveness was reached. Even though an 1% of error is still present, it is important to point out that with this combination is possible to obtain a 98% of effectiveness under 50µs.



(a) (b)

Fig. 7: Atomic-variables Latency Measurement Results

7. Conclusions

Based on the experimental evaluation effected, the use of atomic-variables combined with the batch-consuming algorithm is an effective and efficient method for handling the inter-process communication given the requirements established. With this approach, significant amount of waiting time is saved because the *consumers* do not waste time waiting for a signal from the *producer* only until when they reach the last published frame. It is important to highlight that by using condition-variables as inter-process communication method are not suitable with the requirements of the project. In the case of high-throughput monitoring systems where large-scale of data is produced, the use of atomic variables is recommended. Since by using atomic-variables has the side effect of increasing the CPU usage, additional IPC mechanisms should be evaluated in order to establish which of them could manage the latency levels similar to atomic-variables but with a moderate usage of CPU.

References

[1] F. Di Mario, M. Park and A. Wallander, "ITER - CODAC Core System Overview," 8 February 2007. [Online]. Available: http://static.iter.org/codac/cs/CODAC_Core_System_Overview_34SDZ5_v5_4.pdf. [Accessed 23 April 2017].

[2] ITER, "The ITER Tokamak," ITER, [Online]. Available: https://www.iter.org/mach/tokamak. [Accessed 23 04 2017].

[3] S. Simrock, P. Patil, V. Martin, L. Abadie, R. Barnsley, L. Bertalot, P. Makijarvi, D. Makowski, J.-Y. Journeaux, R. Reichle, D. Stepanov, G. Vayakis, I. Yonekawa, A. Wallander and M. Walsh, "Diagnostic use Case Examples for ITER Plant Instrumetnation and Control," in *14th International Conference on Accelerator & Large Experimental Physics Control Systems 2013*, 2013.

[4] J. Y. Journeaux, "Plant Control Design Handbook," 08 April 2013. [Online]. Available: http://static.iter.org/codac/pcdh7/Folder%201/1-Plant_Control_Design_Handbook_27LH2V_v7_0.pdf. [Accessed 09 June 2016].

[5] A. Mielczarek, P. P., D. Makowski, M. Orlikowski and G. N. A. Jabłoński, "AMC frame grabber module with PCIe interface," in *20th International Conference Mixed Design of Integrated Circuits and Systems*, Gdynia, Poland, 2013.

[6] V. Martin, V. Moncada and J.-M. Travere, "Challenges of Video Monitoring for Phenomenological Diagnostics in Present and Future Tokamaks," in *IEEE/NPSS 24th Symposium on Fusion Engineering*, 2011.

[7] A. Alexandrescu, A. Stan, N. Botezatu and S. Caraiman, "Real-Time inter-process communication in heterogeneous programming enviroments," in *20th International Conference on System Theory, Control and Computing (ICSTCC)*, Sinaia, Romania, 2016.

[8] R. Stevens, UNIX Network Programming Volume 2, Second Edition, Interprocess Communications, Prentice Hall, 1999.

[9] F. Hijaz, B. Kahne, P. Wilson and O. Khan, "Efficient parallel packet processing using a shared memory many-core processor with hardware support to accelerate communication," in *IEEE International Conference on Networking, Architecture and Storage (NAS)*, Boston, 2015.

[10] A. Williams, C++ Concurrency in Action, New York: Manning Publicaitons Co, 2012.

[11] K. Wall, M. Watson and M. Whitis, Linux Programming Unleashed, Sams Publishing, 1999.

[12] *EoSens® 3CL Camera Manual, Rev. 1.01, Copyright ©*, Mikrotron GmbH, 2010.

[13] H.Kopetz, Real-Time Systems, Design Principles for Distributed Embedded Applications. Second Edition, Springer, 2011.

[14] A. D. Association, *TIFF Revision 6.0*, 1992.

[15] A. Alexandrescu, A. Stan, N. Botezatu and C. S, "Real-time inter-process communication in heterogeneous programming environments," in *20th International Confrence on System Theory, Control an dComputing (ICSTCC), October 13-15*, Sinaia, Romania, 2016.

[16] M. Aldinucci, M. Danelutto, P. Kilpatrick and M. Torquati, "FastFlow: high-level and efficient streaming on multi-core," in *Programming Multi-core and Many-core Computing Systems*, Wiley, 2014.

[17] D. C. Schmidt, "The ADAPTIVE Communication Environment An Object-Oriented Network Programming Toolkit," in *12th Sun User Group conferences*, San Francisco, 1993.

[18] R. BENOSMAN, K. BARKAOUI and Y. ALBRIEUX, "A New Dynamic IPC-Memory Allocator Based on a Paging Approach," in *High Performance Computing and Simulation (HPCS)*, Paris, 2013.

[19] M. Mitchel, J. Oldham and A. Samuel, Advanced Linux Programming, New Riders Publishing, 2001.

Bartosz KANIA[1]

Nonlinear PID controller for ionisation gauge

The quality of an electron beam should be as good as possible, because it has an influence on the accuracy and repeatability of measurements or processes realised with vacuum instruments such as ionisation gauges or evaporators. There are many analogue and digital solutions of electron beam controllers working with hot cathode electron source, but none of them takes into account that the controlled system is nonlinear. This work presents a nonlinear, digital controller of thermionic emission current, the tuning parameters of which are suited to the system in a whole range of work. Such a controller is designed in order to improve the stabilisation quality of electron emission.

1. Introduction

There are many vacuum instruments that use an electron beam to realise specific purposes, e.g. evaporation in evaporators or measurement in ionisation gauges or isotope ratio mass spectrometers. The accuracy and repeatability of those processes or measurements depend on the properties of the electron beam. The most important property is the quality of stabilisation, because all the devices mentioned demand a constant value of the electron current. In the case of ionisation gauges and isotope ratio mass spectrometers the electron current is used to generate ion current according to (1)

$$I_i = I_e Q_T(U_s) nl \tag{1}$$

where I_i is the ion current, I_e is the electron emission current, Q_T is the cross section of ionisation, U_s is the accelerating voltage, n is the gas concentration and l is the ionisation chamber length. As one can see in equation (1), stabilisation quality depends on the electron beam. Therefore the emission current must be controlled and readjusted. There are some analogue solutions [1, 6, 7] and better digital designs [2, 4, 5], but none of them take into account the nonlinear character of the controlled system. This means that a controller's optimal tuning parameters for one value of the emission current do not have to be optimal for another value of the emission current, which results from changing the dynamic and static parameters of the controlled system within a range of studies.

This work presents a step in improving the quality of stabilisation of the electron emission current for a hot cathode electron source, which involves matching the controller's tuning parameters to the nonlinearity of the controlled system.

2. Design

A block diagram of the designed control system is presented in figure 1, where I_{e0} is the reference value, e is the error, u is the non-real-time controller (NI DAQ card) output voltage, U_c is the cathode voltage, I_e is the electron emission current and U_{fb} is the feedback voltage proportional to the system output value I_e. It consists of a National Instruments USB data

[1] Ph.D. student, Lublin University of Technology, Nadbystrzycka 38D, 20-618 Lublin, POLAND,
tel. +(48-81) 538 43 08, fax: +(48-81) 538 43 15, e-mail: b.kania@pollub.pl

acquisition card (NI USB-6251), a personal computer with Windows and LabVIEW software, the controlled system and intermediary circuits. The data acquisition card uses a 16-bit, 1MS/s analogue to digital converter, a 16-bit, 2.8MS/s digital to analogue converter and is responsible for signal conversion.

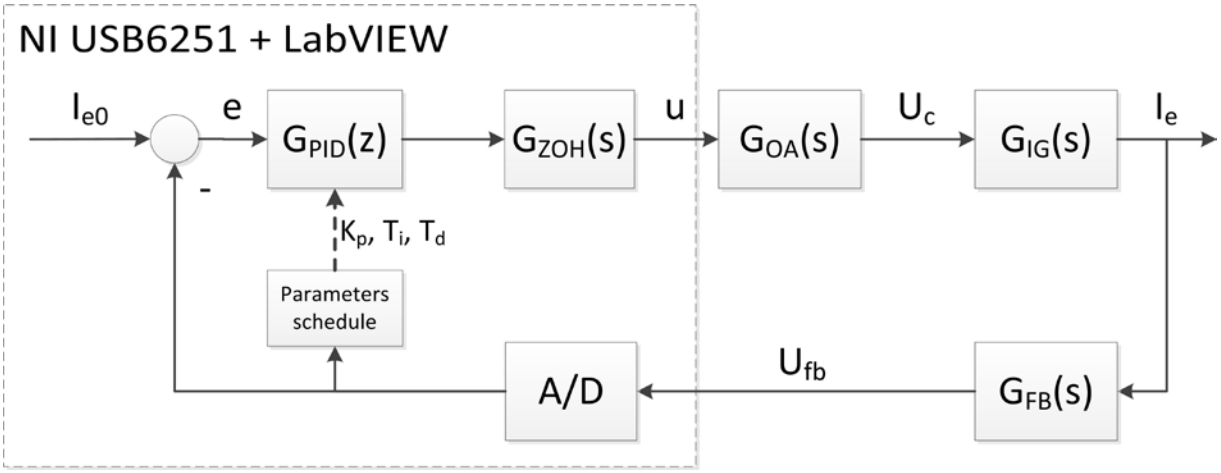


Figure. 1: A block diagram of the designed control system. $G_{OA}(s)$ is a power operational amplifier transfer function, $G_{FB}(s)$ is a feedback circuit transfer function, $G_{IG}(s)$ is the transfer function of the controlled system – an ionisation gauge (IG).

An algorithm is realised by personal computer and LabVIEW software. It realises the summation and nonlinear PI/PID algorithms described by $G_{PID}(z)$ (2).

$$G_{PID}(z) = k_p(I_e)\left(1 + \frac{T_s}{T_i(I_e)}\left[\frac{z+1}{z-1}\right] + \frac{T_d(I_e)}{T_s}\left[\frac{z-1}{z}\right]\right) \tag{2}$$

where $k_p(I_e)$ is the proportional gain, $T_i(I_e)$ is the integral time constant, $T_d(I_e)$ is the derivative time constant and T_s is the sampling period. The tuning parameters are dependent on the output value. The intermediary circuits are the same as in [4]. The assumption is made that transfer functions $G_{OA}(s)$ and $G_{FB}(s)$ are pure gain: $G_{OA}(s)$ = 2 V/V and $G_{FB}(s)$ = 387.5 mV/mA. It was noticed that such a system is strongly nonlinear and behaves like high order inertia, therefore a modified transfer function for ionisation gauge is proposed (3). It is described by gain $k_p(I_e)$, time constant $T(I_e)$ and time delay $T_d(I_e)$ dependent on output value I_e.

$$G_{IG}(s) = \frac{k(I_e)}{T(I_e)s+1} e^{-sT_d(I_e)} \tag{3}$$

The gain is growing, the time constant and time delay are decreasing with thermionic emission. Better modelling, like frequency domain identification, will be carried out in further investigations using the previously designed hot cathode electron source simulator [3].

3. Hardware

The hardware part of the controller is based on the controller designed in [4]. A simplified diagram of the emission current controller is shown in figure 2.

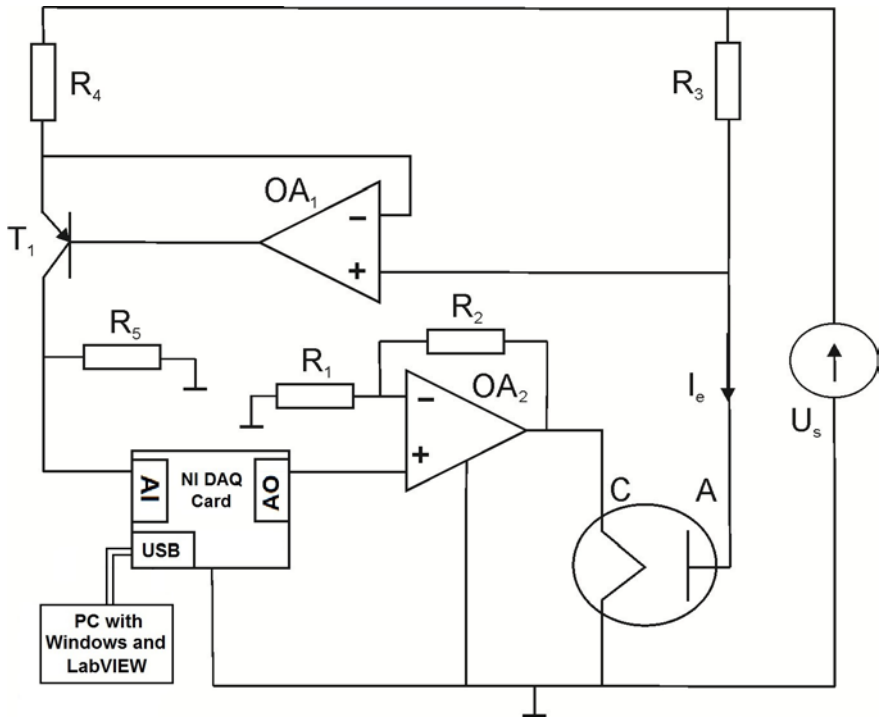


Figure. 2: A simplified diagram of the electron emission current controller. R_3, OA_1, R_4, T_1 create the current mirror for the emission current I_e. R_5 forms a feedback signal. The operational amplifier OA_2 (OPA549) drives the cathode heating. U_s is an electron accelerating voltage. C is the cathode and A the anode of the ionisation gauge.

Symmetrical power supply (+/- 15VDC) is delivered from an external power supply unit. The heating current is hardware limited to circa 3 A by the operational amplifier OA_2 to protect the cathode against overheating. The thermionic emission current is sensed by a current mirror circuit [6]. High voltage for electron acceleration U_s is delivered from an external high voltage power supply unit (0-120 V).

4. Software

A personal computer with LabVIEW software and the author's program (in the case of LabVIEW called virtual instrument) applies an algorithm which includes filtering, the main nonlinear PID algorithm and other tasks like computing relative standard deviation. The general algorithm is presented in figure 3.

The virtual instrument was created using the LabVIEW software. Stable work of this software was ensured by using a timer, which was a clock controller with changeable frequency. Investigation of long term relative standard deviation was carried out for T_s = 40 ms. The timer is a trigger for ADC data acquisition and the PID algorithm. Fast 1 MS/s ADC was used to average 2000 samples in order to obtain one process value of the thermionic emission current. The presented controller works in two modes: open loop, which means controlling the cathode voltage, and closed loop, which means controlling the thermionic current. Only the closed loop mode uses linear/nonlinear PI/PID controller.

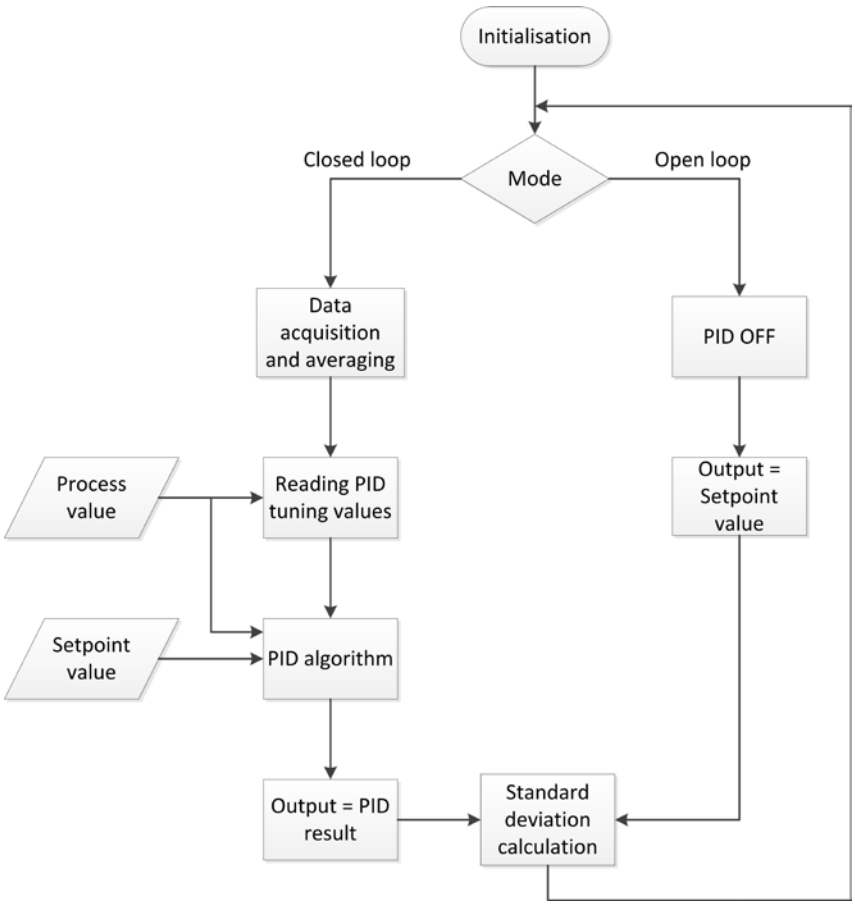


Figure. 3: Main algorithm of the nonlinear PID controller.

5. Results

The measurements have been performed for a long tungsten cathode (45 mm long, 0.13 mm diameter) installed in a Bayard-Alpert ionisation gauge. The results concerning 180 s relative standard deviation are showed in figure 6 and table I. The controller was tuned by the Ziegler-Nichols method using previously obtained approximation data, but the resulting tuning parameters made the system unstable, therefore the autotuning method was applied. In the case of the PI/PID controller the autotuning method for 1 mA of thermionic current was carried out. In the case of the nonlinear PI/PID controller autotuning was made for each value of the emission current.

The results show that applying a nonlinear PI controller gives the lowest standard deviation, its average value being 0.0187 %, while for a linear PI controller it is 0.0277 %. Furthermore, the results are better than those reported in [6] and comparable with [4], where the average standard deviation for the same ionisation bulb was 0.03 % and 0.015 %, respectively. However, the nonlinear PI controller gave somewhat worse results than the linear PID described in [4], but it should be noted that the controller used here is non-real-time and its sampling time equals 40 ms, while in [4] there was a real-time controller with 1 ms of sampling time. Therefore, it is certainly possible to improve the standard deviation of thermionic current by decreasing sampling time and using a real-time nonlinear PID controller. The nonlinear PID controller gave better results than the linear PI or PID controller, but worse than the nonlinear PI controller. One can conclude that the autotuning method is not the best way to obtain the lowest standard deviation of such a system.

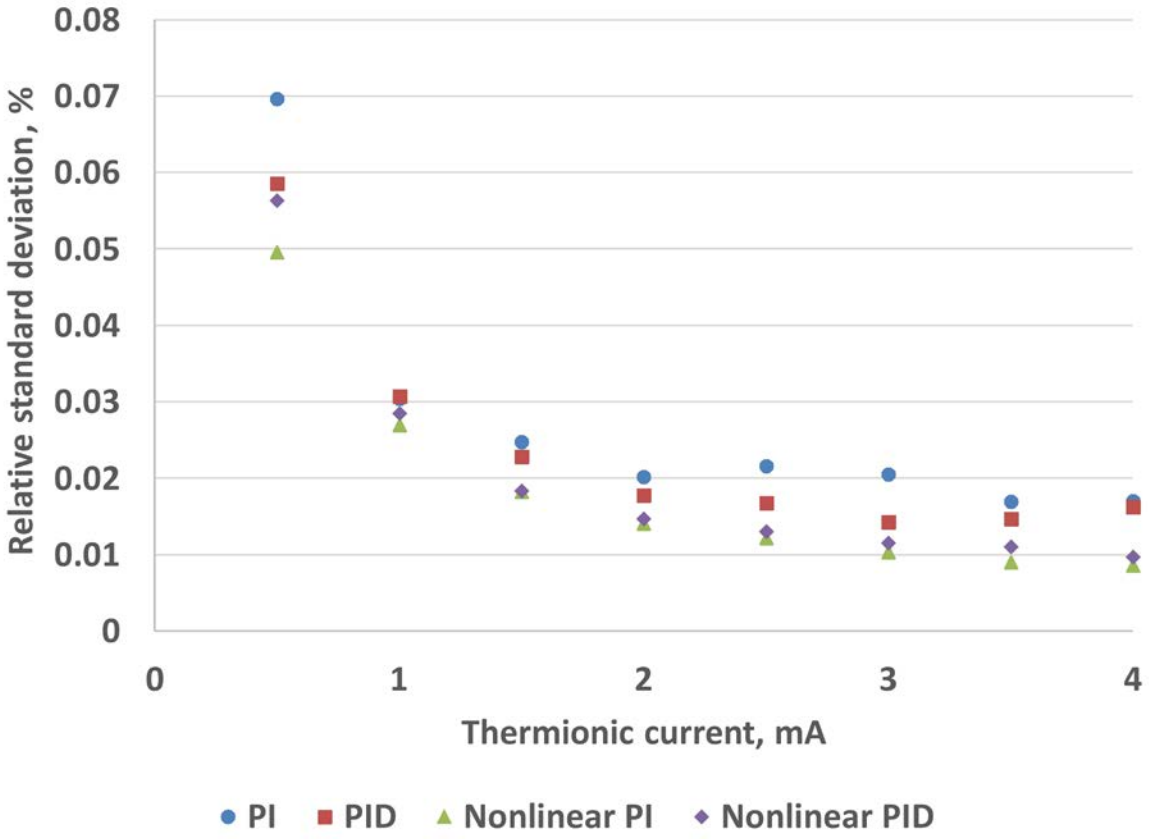


Figure. 4: Results of the electron emission current standard deviation.

Table I. Averaged relative standard deviation.

	PI	PID	Nonlinear PI	Nonlinear PID
Averaged relative standard deviation, %	0.0277	0.0240	0.0187	0.0204

6. Remarks and conclusion

To sum up, investigations of a single-input, single-output thermionic emission controller have been carried out. The results of emission current stabilisation quality are comparable or better than those reported in previous works. The investigations confirmed suppositions that a nonlinear PI or PID algorithm can give better results than the classic PI or PID algorithm in the case of non-real-time controller. However, it is strongly recommended to use a real-time controller with a short sampling time. Further investigations should answer the question if a real-time nonlinear controller can offer better results than those currently best ones. This is a key issue for the application of a nonlinear PID controller. If the results are better, a digital, nonlinear PID controller can improve the accuracy and measurements repeatability of vacuum instruments with a hot cathode, such as electron-impact gas ionisation mass spectrometers or ionisation gauges.

References

1. Donkov N., Knapp W.: *Control of hot-filament ionization gauge emission current: mathematical model and model-based controller*, Measurement Science and Technology, 1997, 8, 7

2. Flaxer E.: *Programmable smart electron emission controller for hot filament*, Review of Scientific Instruments, 2011, 82

3. Kania B.: *A LabView-based hot cathode electron source simulator with no-boundary conditions computation method*, International Journal of Numerical Modelling, 2016, 29, 6

4. Kania B., Sikora J.: *Thermionic emission controller with PID algorithm*, The 23rd International Conference "Mixed Design of Integrated Circuits and Systems", Łódź, Poland, June, 2016,

5. Meyer zu Heringdorf F.-J., Belton A. C.: *Flexible microprocessor-based evaporation controller*, Review of Scientific Instruments, 2004, 12, 75.

6. Sikora J.: *Dual application of a biasing system to an electron source with a hot cathode*, Measurement Science and Technology, 2004, 15, 1

7. Sikora J., Hałas S.: *A novel circuit for independent control of electron energy and emission current of a hot cathode electron source*, Rapid Communications in Mass Spectrometry, 2011, 25, 6

Zuzana MŰNSTEROVÁ[1], Jana PAŘÍLKOVÁ[2], Milan GOMBOŠ[3]

Applications EIS in water management

In this report are considered issues related to the application of electrical impedance spectrometry method in observation of selected hydrological processes and parameters in water management. These are, in particular, processes related to the safety of the operation of earthen dikes and the infiltration of water into the soil associated with solving the issues of flood and drought protection. The presented field survey results were identified by pilot projects in Hornice (Czech Republic) for documentation of BioSealing process used for remediation of the seepage and the process of water infiltration into the heavy soil was tested in Poľany (Slovak Republic).

1. Introduction

Water, contained in both organic and inorganic substances, is an essential condition of life on Earth and an important raw material for many industrial branches. It has a significant impact on engineering structures and, regarding potential construction risks, it is necessary to account for mutual interactions. Therefore, monitoring hydrogeological conditions is an important part of engineering-geological surveys. E.g. flood events, as well as drought, represent the greatest threats of natural disasters in the Czech Republic, including their impact on buildings and soil environments. Processes relating to the penetration of water into a very complex porous environment are relatively difficult to formulate mathematically. Therefore, monitoring these will always have an irreplaceable role. The method of electrical impedance spectrometry (EIS) has been utilized in the Czech Republic (CR) in water management since approximately 2001 [1], thanks to the outcomes of projects of the Grant Agency of the Czech Republic and the projects of the international program of applied research EUREKA.

EIS is one of indirect electrical measuring methods that enable the detection of electrical characteristics of a porous environment containing water. Due to the high complexity of the natural environment, its geological structure, geotechnical diversity, morphology, hydrogeology and wide typology of individual constructions [3, 7], the correct interpretation of attained results is a demanding issue requiring the expertise of surveyors and combined knowledge from several disciplines. The EIS method and the apparatus with the Z-meter device [13] were mainly used in water management to monitor anomalies in the soil, the spread of pollutants through soil and water, the penetration of water into the soil, to evaluate seepage conditions of sealing elements of dikes or penetration of water and changes in conductivity of the dike core soil, and to monitor the morphology of the bottom of watercourses and water reservoirs [10], etc.

On the basis of the field tests carried out so far, it is safe to assert the convenience of handling the measuring apparatus and its resistance to weather conditions, sensitivity [14], both accuracy and repeatability of measurements and reproducibility of data [4]. If the output is required in the form of a non-electric parameters the measuring range and the need to calibrate the measured characteristics against the required must be respected. It is therefore

[1] Ph.D. student, Brno University of Technology, Veveří 331/95, 602 00 Brno, CZECH REPUBLIC,
tel. +420 541 147 283, fax: +420 541 147 288, e-mail: munsterova.z@fce.vutbr.cz
[2] doc., CSc., Brno University of Technology, Veveří 331/95, 602 00 Brno, CZECH REPUBLIC,
tel. +420 541 147 284, fax: +420 541 147 288, e-mail: parilkova.j@fce.vutbr.cz
[3] CSc., Slovak Academy of Sciences, Dúbravská cesta 9, 841 04 Bratislava, Slovak REPUBLIC,
tel. 056/6425147, fax: 0946/6425147, e-mail: gombos@uh.savba.sk

often more appropriate to carry out relative monitoring, where the changes occurring in porous environment are monitored against initial findings.

Measurements, selected results of which are the subject of the submitted paper, were carried out in observation of relevant hydrodynamic processes and parameters in field conditions at locations of Hornice, CR [11, 12], and Poľany, SR [5, 8, 9].

2. Electrical impedance spectrometry, Z-meter device

The EIS method represents a sensitive instrument and an experimental technique designed to determine the parameters of observed hydrodynamic effect. It is based on a periodic harmonic signal of a small amplitude providing minimal concentration changes at the surface of the electrode associated with the measured environment [14]. The basic principle of the electrical impedance spectrometry method is the measuring of the frequency characteristic of the electrical impedance Z of the environment.

Measuring the electrical impedance of soils is based on the consideration that two electrodes are installed apart at a distance *l*, which always make up one EIS probe sensor. The monitored area, defined by the electrodes, always has the character of a resistor, it is the real part of electrical impedance, and the imaginary part is represented by apparent resistivity, which in soils has the character of capacitor.

By the frequency characteristic of electrical impedance of soil $\mathbf{Z} = R + \mathrm{j} \cdot X$ [Ω] can be used to characterize properties of soil. E.g. water content, porosity, humidity, ion content, temperature, etc. by measuring the *R* (resistance), while the *X*, apparent resistivity (reactance, changes with frequency), characterizes properties such as the texture, grain size, soil deposition, etc. Generally, the lower the *R* value, the higher the water content in the measured profile [14].

The apparatus used to monitor selected hydrological processes and parameters in the field conditions (Hornice and Poľany) included Z-meter III or Z-meter IV devices and divided probes (Figure. 1, Table 1), where sensors formed two opposing, parallel, tubular, stainless electrodes in different layer in soil. The length and inner diameter of the electrodes were based on the site requirements, the observed phenomenon, and the monitoring concept.

Table 1: Basic technical parameters of Z-meter devices [13]

Parameter	 **Z-meter III**	 **Z-meter IV**
Impedance range	100 Ω – 1 MΩ	10 Ω – 1 MΩ
Frequency range	1 kHz – 100kHz	100 Hz – 200kHz
Measuring Voltage	1.0 V	0.2 V, 1.0 V
Accuracy of module Z measurement	± 2% from range	± 2% from range
Accuracy of phase measurement	± 2°	± 2°
Communication interface	USB, SD card	USB, SD card, bluetooth
Number of measurement points	1, 8, 16, 32, 64, 128	1, 8, 16, 32, 64, 128, 256
Switcher	internal, external	internal, external
Power	battery	battery

Figure. 1: Used measuring probes – location of Hornice (left), location of Pol'any (right)

3. Measurements at Hornice and Poľany

The hydraulic structure (HS) Hornice (Figure. 2), located in the District of Třebíč, Czech Republic, was approved on 31.05.1985. Originally was served as an irrigation reservoir, currently it is used for retention and accumulation of surface water and utilized for breeding fish and ducks. Agricultural Cooperative Dešov is the owner of the HS since 2012.

Figure. 2: Situation of the HS at the location Hornice (www.mapy.cz)

The height of the earthen dike above the original ground is 8.22 m [6]. On the dike crown lies a gravel road runs with the average width of 4.25 m. According to the project [6], the dike at its base is 46.80 m wide. The upstream face slope has a gradient of 1:2.5 and is fortified with stone packing; the downstream face slope of the dike has a designed gradient of 1:2.5 with grass cover.

Due to the deformation of the dike body caused by a progressing seepage, the dike underwent remediation and monitoring. The non-traditional BioSealing method was used, for the first time in the Czech Republic, for the remediation of the dike of the HS [12]. The impacts of the application of BioSealing technology [2, 15], the fluctuation of water level in reservoir and weather conditions were monitored. Two types of sensors were used to measure the water level in the reservoir; the first one was based on measuring the electrical impedance and the other (ESMWLPS Water level sensor) on measuring the relative hydrostatic pressure. The main parameter documenting the application of BioSealing nutrient aqueous solution was the water penetration (seepage) measured by the PF500 shuttle flowmeter. The system was further supplemented with an ESP11 electrical conductivity, pH and temperature sensor. The probes of the EIS method were realized in the total lengths of 3 m (on the downstream face of the dike in the area of the observed seepage) and 10 m (in the crown of the dike 0.15 m from the upper edge of the downstream slope). The 3 m long probe is vertically situated 3.03 m below the crown of the dike, the total number of sensors is 16, and the first sensor is 0.03 m below the surface, the 10 m long probe carries a total of 22 sensors with the first sensor at a depth of 1.0 m under the crown of the dike.

The application of BioSealing nutrient aqueous solution on the ground dike of HS Hornice took place from 09.05.2016 to 03.06.2016, when a total volume of 205 litres of aqueous nutrient solution (Figure. 3 and Figure. 4) was injected into 5 boreholes daily, except for Saturdays and Sundays. At the same time, an increase of the water level in the reservoir was

secured by partitioning the auxiliary safety overflow until 06.06.2016. Until 28.06.2016 the level in the reservoir was kept at approximately constant level, on 29.06.2016 the partitioning was reduced by 0.10 m, then the water level was changing in relation to the weather conditions and from 20.10.2016 the reservoir was being discharged.

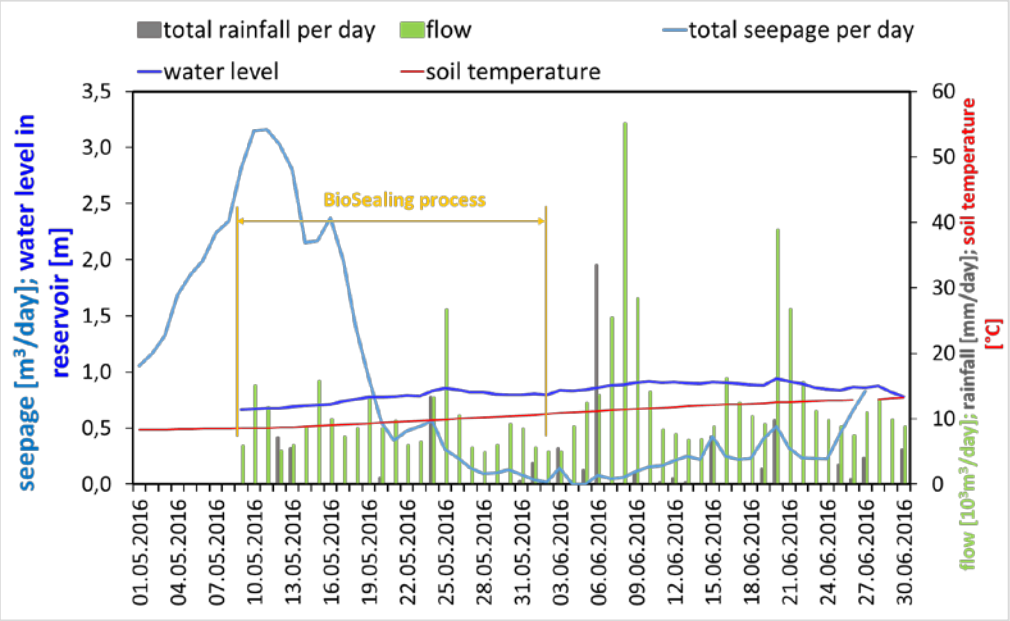


Figure. 3: Documentation of the BioSealing process and measured data

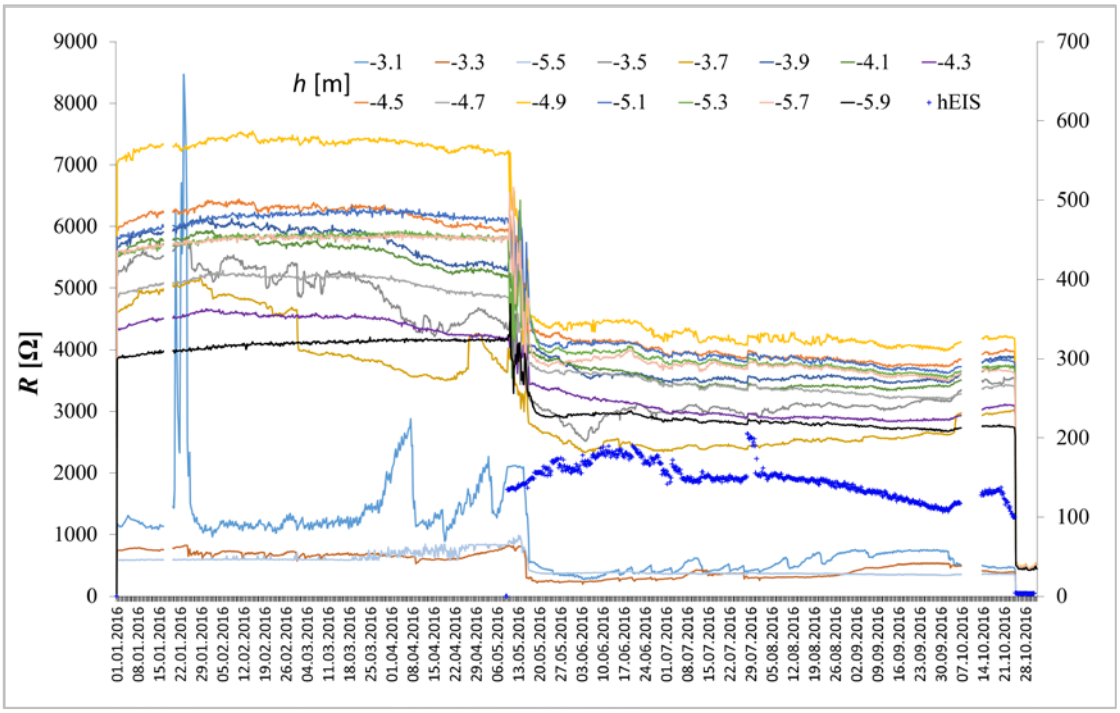


Figure. 4: Documentation of EIS probe measurement (3 m) of BioSealing process

Because the elements of the BioSealing method are newly applied to an earth environment quite different in terms of character, it is difficult even to estimate within what time span the seepage can be remediated by the “swelling” bacteria. Therefore, it is hard to evaluate what action has resulted in changes of the pattern of resistance values on the monitored profiles. The measured data were subjected to another analysis and a correlation of the results of

measurement and dosing of elements of the BioSealing method was searched. The long-term patterns of resistance, where it holds true that the higher the value, the drier the soil at the monitored level, show that the environment at greater depths is the same over the entire duration of monitoring. From the view of the short-term monitoring is in some layers to higher values of resistance were measured. It correspond with the measured approximately zero or very low value of the seepage at the end of period of BioSealing process to the June, 24. The conclusion of the monitoring is possibility of using the BioSealing process for short-term remediation of earthen dike and the ability of EIS method to document of this process.

Changes during which new seepage pathways are sealed or, conversely, formed in the dike soil can be attributed to the weather effects, to the movement of the water level in the reservoir and to the elements of the BioSealing method provided that its individual elements do not affect the measured components of electrical impedance.

Measurement in field conditions in the Slovak Republic took place in the village of Poľany, District of Trebišov, on a windblown sand dune originating in the quaternary. For the clay soil detection experiment in the ground profile, a sample of clay was taken by soil sampling auger at the location of Senné, District of Michalovce (Figure. 5). For the hydrological description of soils, grain size was determined based on the USDA classification triangle, the retention curve expressing the relation between the moisture potential and the soil moisture, and the EIS method was used for frequency analysis.

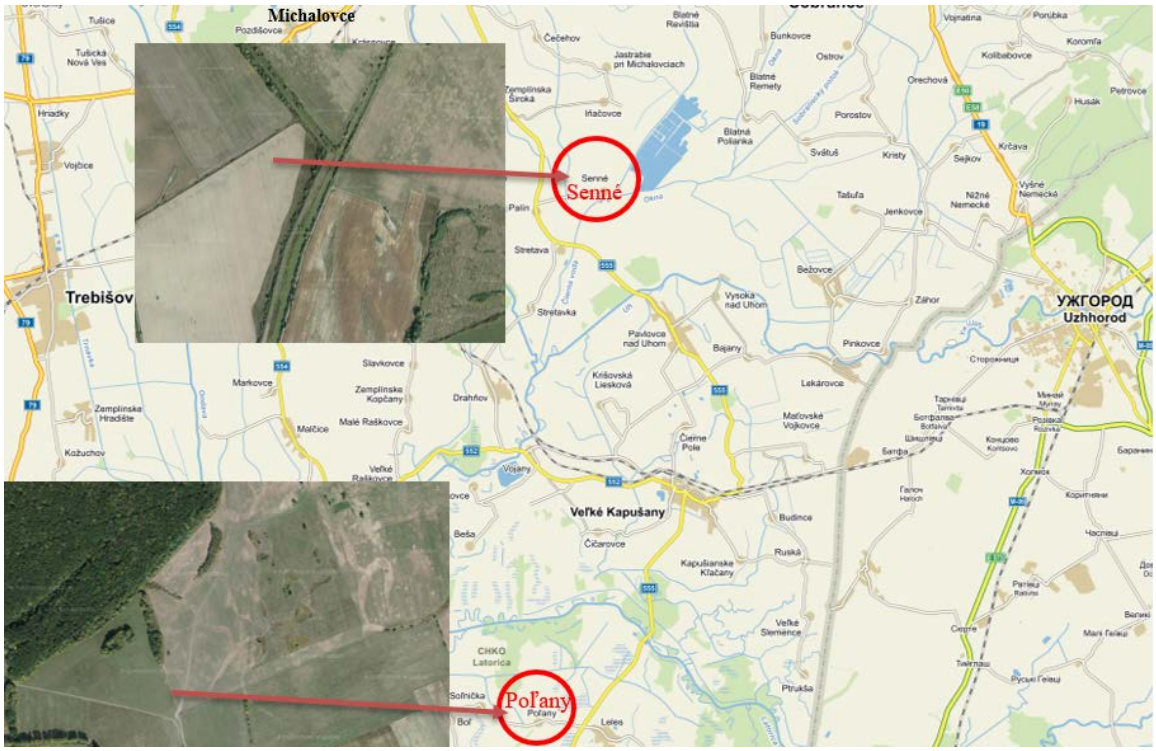


Figure. 5: Situation at the locations of Poľany and Senné (www.mapy.cz)

Artificial soil inhomogeneity at the location of Poľany was made by a hole with the floor plan dimensions of 0.15 m × 0.30 m and a depth of 0.20 m, into which clay of 0.10 m thickness, transported from Senné (Figure. 6), was installed. The clay was compacted by a rubber mallet so as to evenly fill the entire space. The rest of the hole was filled with the original soil (Figure. 6). For the application of the EIS method, four holes were drilled into the ground at a distance of 0.15 m from the edge of the hole, into which two paired divided probes (Figure. 1) were placed. The average value of soil moisture at the beginning of measuring determined by the TDR device was 6 %.

Figure. 6: Experiment of detection soil inhomogeneity at the location of Pol'any

The paired divided probe contained 7 sensors and the measurement frequency was 8 kHz. The measurements took place in two directions, on the shorter (SS) and the longer (LS) side. The soil environment was measured before digging the hole (natural ground) and after filling it. Based on the results of the measurement (Figure. 7) it can be stated that in both measured directions the soil anomaly, artificially realized with a clay lens, was detected at the experimentally set level of 0.15 m below natural ground level. Furthermore, it is obvious that, in terms of the measured resistivity, the natural ground was not homogeneous in any of the measured directions, and the vertical decrease of resistivity indicates an increase of soil moisture with depth. The basic conclusion of this type of measurements – it can be recommended the electrode distance in the range of 0.40 m up to a maximum of 2 m. For the shorter distance is hard to describe soil inhomogeneity resp. structure and for higher distance can be higher influence of geo-, electromagnetic field of the Earth to the measured values.

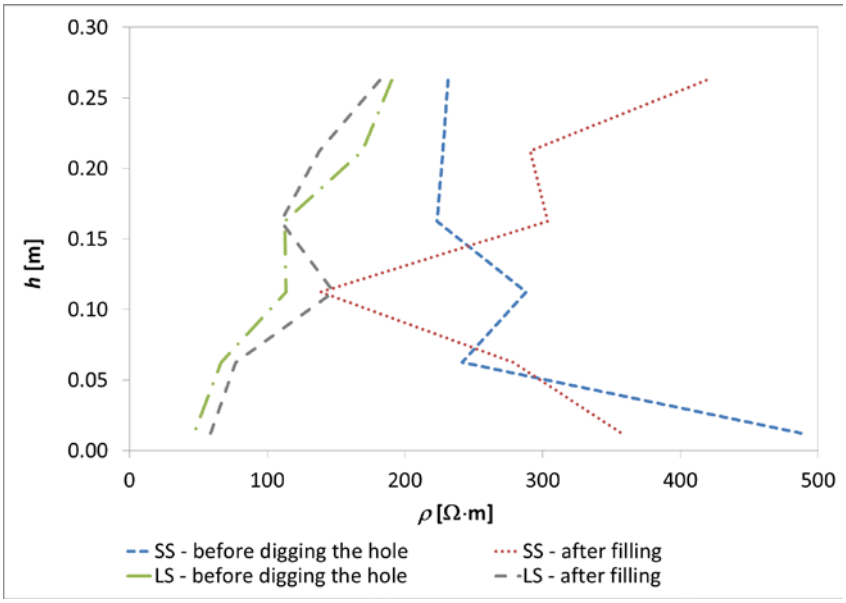


Figure. 7: Experiment results – description of soil inhomogeneity through resistivity

3. Conclusion

The method of BioSealing was first applied in the Czech Republic to the medium of the earth-fill embankment dam of the Hornice water reservoir in 2015 and 2016, i.e. to a completely different medium in terms of character. Even despite intensive consultations with the authors and other users of the given method, it was very difficult on the basis of their experience just to estimate what concentration of the aqueous solution of Nutrolase should be dosed into the soil and in which time horizon the remediation of seepage by "swelling" bacteria should be evaluated. The experiment in 2015 was significantly influenced by the weather because the summer months were characteristic with a long-lasting drought that was also manifested by

a strong drop of the water level in the Hornice reservoir. The experiment in 2016 was characterized of higher water level in reservoir in all time of monitoring. But also was influenced by the weather, in this case higher rainfall after the application of Nutrolase into the dike body. It is possible that it could draining through preference ways from the dike body, so that only small volume could serve as a nutrient for the bacteria living in the soil. However, in the course of the process of dam remediation using the method of BioSealing, it was possible to observe reduced discharge in the pipes. Before earthen dike remediation, a total seepage of 0.24 l/min. was measured on 21 March 2016; after earthen dike remediation the seepage in the left pipe was immeasurable and in the right pipe was 0.13 l/min on 22 June 2016, which represents a 45% change of its value. The seepage through earthen dike is very complicated 3D problem with a lot of unknown parameters like moisture of the soil, grain size, structures, tension and so on. Also it is possible that to the seepage area the water may flow under the catchment system, where the bacteria wasn't in contact with Nutrolase. It is possible to believe, and the results showed it, that from the view short-term the BioSealing process is possible to use for seepage remediation of dike and EIS monitoring method and apparatus is possible to use for short- and long-term monitoring. Other application of both methods goes at the location Opatovicky channel in this year.

Detection of the soil inhomogeneity using EIS method and apparatus was made like pilot measurement. The authors would like on the known samples of the soil environment to search the configuration of the measurement, measuring frequency for clay detection and which the effect of inhomogeneity in the measured values. The work can be describe like basic research with this new apparatus.

The paper was created within the scope of junior specific research FAST-J-17-4689 Possibilities of Determining the Relation between Electrical and Hydraulic Conductivity and the Sustainability of EUREKA project E!7614. It illustrates, on the selected examples, the possibilities of using the EIS method in water management practice. The method was used to monitor the efficacy of remediation of the earthen dike in the Czech Republic and in a pilot study in Slovakia to detect inhomogeneity of the soil environment (changes in the soil structure and moisture) using the EIS method and an apparatus with the Z-meter device.

References

1. Barbiero G., Miracapillo C.: *International application of EIS method for soil humidity measuring*, 2008
2. Blauw, M., Lambert, J.,Latil, M.: Biosealing: A Method for in situ Sealing of Leakages. 125-130. 10.3850/GI132m 2009
3. Cooper, D.: Soil water measurement: a practical handbook. Hoboken, NJ: John Wiley, 2016. ISBN 9781405176767.
4. Fejfarová M.: *Možnosti detekce obsahu vody v pórovitém prostředí metodou elektrické impedanční spektrometrie*, doktorská práce, 2014
5. Gomboš M., Kadrna B., Vasiľová I., Pavelková D., Tall A.: *Variability of electrical impendance in the soil environment*, 1st Conference and Working Session Within the frame of the Unternational Program EUREKA, Brno, 2013
6. Hejtman Z.: *Provozní a manipulační řád pro vodní nádrž „Hornice“ na p. č. 124/33 v k. ú. Hornice. Podrobná situace*. Vypracoval zodpovědný projektant Ing Zdeněk Hejtman zak č 31/2012 č výkresu 3 část H 3, 2012
7. Kearey P., Brooks M., HILL I.: *An introduction to geophysical exploration*. 3rd ed./. Malden, MA: Blackwell Science, 2002. ISBN 0632049294.
8. Pařílková J., Gardavská Z., Fejfarová M., Zachoval Z.: *Dlouhodobé monitorování zemní hráze malé vodní nádrže metodou EIS*, SAV, Bratislava, 2012

9. Pařílková J.: *Monitoring the Morphology of the Bottom of Streams and Reservoirs Using Electrical Impedance Spectrometry Method*, 2010
10. Pařílková J., Pařílek L., Fránková H., Nehudek A., Pazdírek O.: *The monitoring of the earth-fill dam of the Hornice water reservoir using the EIS method*, 4th conference and working session, Brno, 2016
11. Rupp D., Pazdírek O.: *BioSealing – the final report of a pilot project EUREKA 2015* Jaroměřice nad Rokytnou pp 197-205 ISBN 978-80-214-5338-8 ISSN 2464-4595
12. Pařílková J., Radkovský K.: *Z-meter IV – User's manual*, 2016
13. Pařílková J., Ševčík R.: *The physical substance of the method of electrical impedance spectrometry*, Brno, 2010
14. van der Zaan, B.: *BioSealing Column-experiment 1: Durability of Biosealing in field material*, 2015

Jan DUSEK[1], Jan MIKULKA[2]

Using a GPU to Accelerate Electrical Impedance Tomography Algorithms

The presented paper discusses parallelizing an algorithm for image reconstruction in electrical impedance tomography (EIT). An Nvidia GPU platform was used for the data processing. The introductory section comprises a general description of EIT, relevant practical applications, and current status of the program. The following chapter outlines the objective function with Tikhonov regularization and the possibility of a mathematical extension to increase the image resolution. In addition, we also characterize the modification of the reconstruction algorithm with respect to the computational speed. The results achieved via the novel GPU implementation are discussed, with a survey of the achieved algorithm acceleration and related problems. The CPU and GPU methods are compared by means of a time-consuming solution.

1. Introduction

Electrical Impedance Tomography (EIT) is a non-destructive and non-invasive diagnostic method suitable for reconstructing the impedance distribution in an examined space. Principally, the method exploits an alternating current progressively applied to all the electrodes. The electrodes are deployed at regular intervals along the border of the examined object. The internal impedance is evaluated according to the voltage values measured on the electrodes, and the spatial image resolution can be changed in accordance with the number of the electrodes. [2][4][6]

EIT as a diagnostic method can be found in a variety of technical and research branches, including biomedicine, materials engineering, and geophysical mapping. In the first of these domains, the technique is used to detect blood clots or tumors and observe internal organs, such as the brain or lungs. In small children, brain monitoring is often performed instead of X-ray examination. As regards materials engineering, the described tomographic method remains of major importance due to its non-destructive character, finding application in the monitoring of inhomogeneities, cracks, or defects inside a material. The importance of EIT has also recently increased within geophysical mapping, where this type of imaging is instrumental for detecting the formation of bedrock and subsoil or observing groundwater and seepage in artificial lakes. [1][2][6]

The numerical EIT algorithm comprises the forward solution and ill-posed problem, which need to be calculated in every iteration of an image reconstruction procedure. When computing the forward solution, the impedance inside the examined object is evaluated with the finite element method (FEM). Based on the results obtained in this manner, the objective function determines the voltage values on the electrodes. In our case, the computation of the program was first performed via a CPU. Then, the Jacobi matrix solution was modified, with

[1] Ph.D. student, Brno University of Technology, Technická 12, 616 00 Brno, CZECH REPUBLIC,
tel. +420 541 146 277, fax: +420 541 146 276, e-mail: xdusek19@stud.feec.vutbr.cz
[2] Assoc. Prof., Brno University of Technology, Technická 12, 616 00 Brno, CZECH REPUBLIC,
tel. +420 541 146 292, fax: +420 541 146 276, e-mail: mikulka@feec.vutbr.cz

the aim to parallelize and increase the speed of the implemented CUDA platform. Exploiting the relevant results, GPU computing was applied to the entire algorithm. [1][2]

2. Theoretical Framework

Image reconstruction in EIT is a non-linear problem that combines the forward solution and ill-posed problem with regularization. The objective function described herein is based on the least squares method with Tikhonov regularization. The objective function of the inverse problem assumes the form:

$$\Psi(\boldsymbol{\sigma}) = \frac{1}{2}\sum_{\mathrm{ne}} \|\mathbf{U}_{\mathrm{M}} - \mathbf{U}_{\mathrm{FEM}}(\boldsymbol{\sigma})\|^2 + \alpha\|\mathbf{R}\boldsymbol{\sigma}\|^2, \tag{1}$$

where $\Psi(\boldsymbol{\sigma})$ is the regularized objective function, $\boldsymbol{\sigma}$ denotes the vector of the unknown conductivity distribution in the monitored object, $\mathbf{U}_{\mathrm{M}}$ represents the vector of the voltages measured on the electrodes, $\mathbf{U}_{\mathrm{FEM}}(\boldsymbol{\sigma})$ is the vector of the voltage obtained via solving the forward problem, α stands for the regularization parameter, and $\mathbf{R}$ is the regularization matrix connecting neighboring elements having different conductivities. [2][4][5]

In the inverse task, regularization facilitates the computation of an ill-posed problem. Tikhonov regularization then ensures comparatively good convergence, stability, and noise sensitivity reduction. The effect of the regularization matrix on the objective function is reduced via the regularization parameter α, according to the increasing number of iterations.

In the proposed non-linear problem, we seek such a vector σ that will minimize the objective function Ψ(σ). For this reason, iterative techniques are often applied; in this case, we use the Gauss-Newton method, defined by the expression

$$\boldsymbol{\sigma}_{i+1} = \boldsymbol{\sigma}_i + \left(\mathbf{J}_i^{\mathrm{T}}\mathbf{J}_i + \alpha\mathbf{R}^{\mathrm{T}}\mathbf{R}\right)^{-1}\left(\mathbf{J}_i^{\mathrm{T}}\left(\mathbf{U}_{\mathrm{M}} - \mathbf{U}_{\mathrm{FEM}}(\boldsymbol{\sigma})\right) - \alpha\mathbf{R}^{\mathrm{T}}\mathbf{R}\boldsymbol{\sigma}_i\right), \tag{2}$$

where $\boldsymbol{\sigma}_{\mathrm{i+1}}$ is the novel conductivity approximation, $\boldsymbol{\sigma}_{\mathrm{i}}$ denotes the conductivity vector from the previous step, and $\mathbf{J}_{\mathrm{i}}$ represents the Jacobian expressing the sensitivity of the electrode potentials to a change of the conductivity in the given element. [1][2]

The Jacobian is realized from the linear part of the Taylor polynomial. The Jacobi matrix is numerically simpler; however, the yield information volume is smaller. To obtain a more detailed image, we need to extend the Taylor polynomial to higher orders of derivatives. The expression of the quadratic polynomial is defined as follows:

$$T_2(\mathbf{x}) = f(\mathbf{x}_0) + \frac{1}{1!}\sum_{j=0}^{m}\frac{\partial f(\mathbf{x}_0)}{\partial \mathbf{x}_{0j}}(\mathbf{x} - \mathbf{x}_0) + \frac{1}{2!}\sum_{j=0}^{m}\frac{\partial^2 f(\mathbf{x}_0)}{\partial \mathbf{x}_{0j}{}^2}(\mathbf{x} - \mathbf{x}_0)^2\,, \tag{3}$$

where $\mathbf{x}$ is the vector of unknown points, and $\mathbf{x}_0$ denotes the vector of the function roots.

The second order of the Taylor polynomial does not lead to the solution of the Jacobian but facilitates the computation of the Hessian matrix, which is more time-consuming and mathematically more complex. [3][4]

3. Proposed solution

This section briefly introduces the inverse problem algorithm with Tikhonov regularization, including the modifications in the GPU-based implementation. The algorithm was designed in Matlab R2016a. The flowcharts of the original program and the novel GPU-based implementation are presented in Fig. 1.

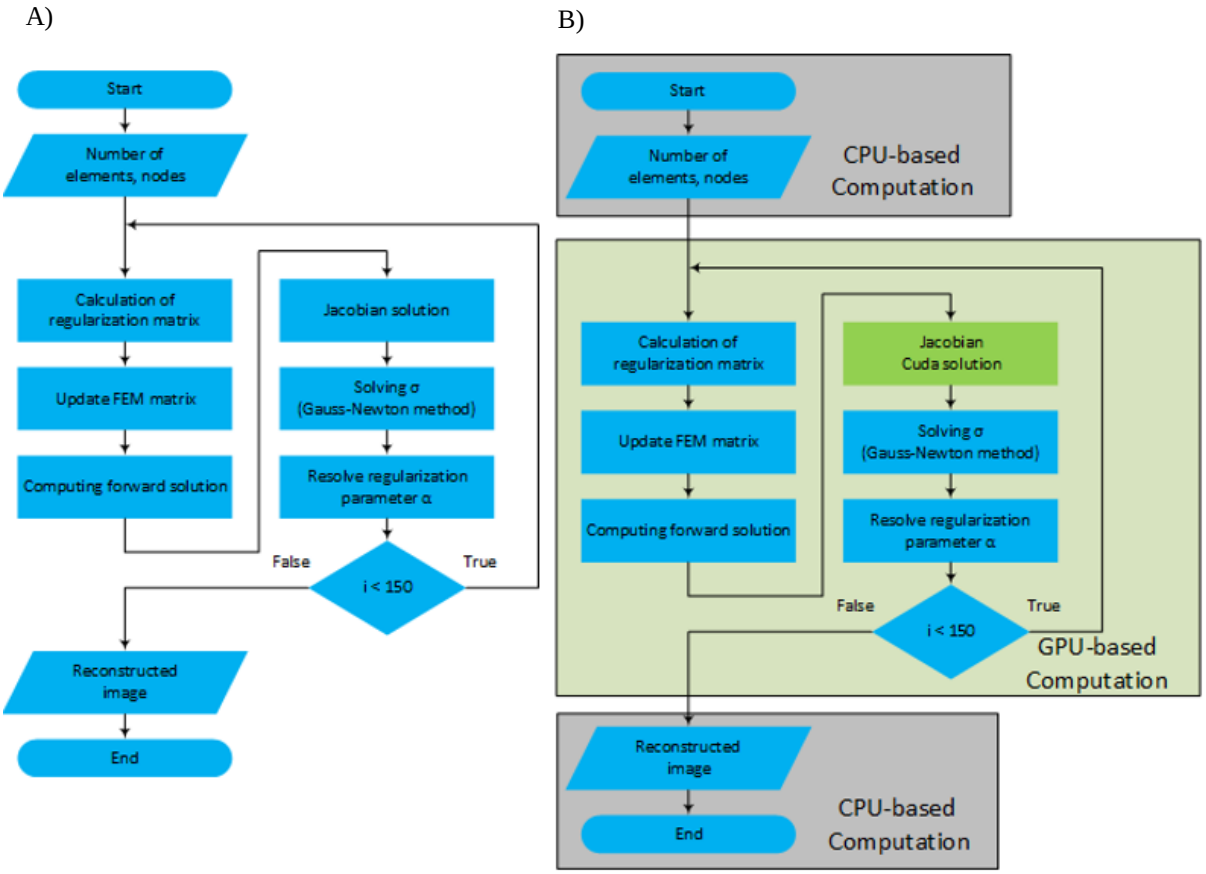


Figure 1: The flowcharts of a) the program for image reconstruction via EIT and b) the proposed GPU implementation (the CPU-based computation is shown in grey, while the GPU-based calculation is indicated in light green; the Jacobian resolved via the CUDA function is then presented as the highlighted green box).

The major part of the program was originally solved via the employed CPU; more concretely, the operations thus solved included the input data loading, computation of the regularization matrix, forward solution by means of the FEM, evaluation of the impedance inside the examined object, computing the conductivity vector σ via the Gauss-Newton method, and resolving the regularization parameter α. The Jacobian matrix was implemented in the CUDA function [2]. As regards speeding up the Jacobi matrix calculation, the present work is focused on implementing the rest of the algorithms into the GPU. The proposed GPU-based implementation required corrections in the initialization phase, where we had to relocate the input data from the RAM to VRAM.

The computation utilized the GPU but one part of the algorithm could not be implemented in Matlab. This problem with the *symamd* function for the symmetric definite matrix could not be resolved because Matlab does not have an equivalent feature with the GPU support. The

necessary data transfer between the RAM and VRAM caused a higher time consumption, as verified through relevant testing and evaluated for each part of the algorithm.

Table 1: The times required to compute individual steps in the CPU and GPU-based reconstruction processes.

Process	**CPU – time [ms]**	**GPU – time [ms]**
Calculating the regularization matrix	0.6	0.8
Updating the FEM matrix	0.3	0.3
Forward solution	1.6	4.3
Jacobian computation	10.9	0.5
Solving σ with the Gauss-Newton method	2.3	14.6
Total time consumption per iteration	**15.7**	**20.5**

The results shown in Table 1 indicate that the GPU-based implementation proceeded at a slower pace in most of the steps. The *Calculating regularization matrix* procedure increased the overall time intensity due to the multiplication of a one-dimensional vector with a two-dimensional regularization matrix. The multiplication of the vector with the matrix was slower than expected, whereas the *Update FEM matrix* process proved to be slightly faster. We assume a larger difference for higher numbers of elements. The proposed *Forward solution* implementation was hampered by the data transfer between the RAM and VRAM. Within this phase, the function for the symmetric definite matrix was resolved. The CUDA-based implementation of the *Jacobian computation* exhibited a very high speed, being more than twenty times faster than the original CPU process. The last of the procedures, *solving σ with the Gauss-Newton method*, increased the total time consumption. In this type of computation, a GPU is used to facilitate the matrix division. The matrix division function comprises a process resolved via systems of linear equations, Ax = B. During the testing, we established that each evaluation cycle strongly depends on the results previously achieved by the function. The sequential behavior of the systems of linear equations was found to embody the factor that most prominently decelerates the discussed process.

4. Hardware and software

The applied hardware and software tools are as follows:

- CPU: Intel Core i5 – 4460; number of cores: 4; clock rate: 3.2 GHz; main memory: 8 GB RAM; operating system: Windows 10, version x64; Matlab R2016a;
- GPU: NVIDIA GTX 970; max. number of threads in a block: 1024; clock rate: 1.215 GHz; graphics memory: 4 GB, GDDR5; Windows 10 x64; Matlab R2016a.

5. Conclusion

The paper mathematically describes the reconstruction of an image in EIT, with a special view to the possibility of increasing the yield information volume and implementing the GPU-based computational procedure. Such a process is advantageously employed to accelerate the solution of the Jacobian. During the realization and testing of the proposed solution, a problem was detected with the *symamd* function to compute the symmetric definite matrix. The function could not be resolved via Matlab due to a lack of support for the GPU platform and had to be solved via the applied CPU. This approach nevertheless required data transfer between the RAM and VRAM. Another issue consisted in the sequential behavior of the

Gauss-Newton solving process, which slowed down the actual computation. The discussed procedure resolved systems of linear equations, depending on the previous results of the actual function. The time consumption of the proposed solution was higher (20.5 ms) compared to the original approach (15.7 ms). The knowledge acquired through the experiments and outlined herein shows that Matlab does not constitute a suitable option for either fast VRAM administration or the parallelization of computational algorithms in EIT. The results of the GPU-based parallel processing can be improved via a direct implementation of CUDA C; such an operation is generally supposed to comprise manual setting of graphical memory [7], using libraries to enable fast computation with sparse matrices and vectors [8].

Acknowledgement

The paper was supported from the general student development project guaranteed by Brno University of Technology, within National Sustainability Programme under grant LO1401. The infrastructure of the SIX Center was utilized to facilitate the research.

References

1. Dedkova, Jarmila. *Image reconstruction methods based on Electrical Impedance Tomography*. Brno, 2007. Brno University of Technology.
2. Dusek, Jan, David Hladky, and Jan Mikulka. *Electrical Impedance Tomography Methods and Algorithms Processed with a GPU*, In Progress in Electromagnetics Symposium Proceedings 2017. Accepted for publication.
3. EI Jing, Shi Liu, Li Zhihong, Sun Meng, *An image reconstruction algorithm based on the extended Tikhonov regularization method for electrical capacitance tomography*, Measurement, Volume 42, Issue 3, 2009, Pages 368-376, ISSN 0263-2241.
4. Lionheart, WRB. *Developments in EIT Reconstruction Algorithms*. Manchester, 2003, 17. UMIST, Department of Mathematics
5. Lukaschewitsch, Michael, Peter Maass, and Michael Pidcock. *Tikhonov regularization for electrical impedance tomography on unbounded domains*. Inverse Problems 19.3 (2003): 585.
6. Vauhkonen, Marko. *Electrical impedance tomography and prior information*. Kuopio: Dept. of Applied Physics, University of Kuopio, 1997. ISBN 9517817002.
7. Wajman, R., R. Banasiak, L. Mazurkiewicz, T. Dyakowski and D. Sankowski. *Applying Parallel and Distributed Computing for Image reconstruction in 3D Electrical Capacitance tomography*. Automatyka. Krakow: AGH, 2010, vol. 14, no. 3/2, p. 711-722.
8. Kapusta, Pawel, Michal Majchrowicz, Dominik Sankowski and Lidia Jackowska-Strumillo. *Acceleration of image reconstruction in 3D Electrical Capacitance Tomography in heterogeneous, multi-GPU system using sparse matrix computations and Finite Element Method*. Federated Conference on Computer Science and Information Systems. Gdansk, 2016, 679 - 683.

Jiří SLIŽ[1]

Using Perona-Malik Filtering to Detect the Transition between an Embryogenic Tissue and a Culture Medium in MRI Images

In this paper, the Perona-Malik (PM) filtering method is applied to MRI images of an embryogenic tissue. The filtering enables us to reduce the noise and to preserve the shape of the narrow transition between the given tissue and the related culture medium. For comparison, Gaussian filtering is used as well; this technique then shows an SNR nearly twice higher than that obtained with PM. The following stage consists in segmenting the transition region and employing selectivity and specificity to compare the results. The segmentation via PM filtering yielded the selectivity value of 93.8 %.

1. Introduction

Real images invariably contain certain amounts of noise, which is usually spurious and makes the actual image processing more difficult. In cases where only a small level of noise is present, simple denoising methods are applicable; these include, for example, mean, median, or Gaussian filtering. Such methods, although very fast and impressive in noise reduction, nevertheless suppress small details and sharp edges too: generally, the higher the noise level, the more visible the adverse effect of denoising. It is then advisable to utilize an approach that preserves the edges and details; in this context, the procedure outlined by Perona and Malik (PM) [1] embodies an advantageous instrument. The concept exploits the heat equation (HE), whose solution leads to Gaussian filtering. Importantly, PM supplies a nonlinear part into the HE, and when the modified equation is employed as the image filter, the diffusion process is slower in regions with large brightness changes than in homogeneous ones. This scenario has become known as anisotropic diffusion.

The practical application of the PM method is demonstrated through filtering an image obtained through magnetic resonance imaging (MRI). Reference [2] presents a characterization of the embryogenic tissue of the Norway spruce. In our paper, the purpose of the image processing consists in localizing the narrow transition between the embryogenic tissue and the culture medium discussed in [2]. As the MRI image contains a high level of noise, we assume that PM filtering is suitable for the corresponding reduction because it preserves the shape of the objects. The outcomes of the filtering and segmentation are compared, using the signal-to noise ratio (SNR) to assess the noise reduction quality and employing selectivity (SE) and specificity (SP) as the segmentation measure.

[1] Ph.D., Brno University of Technology, Technická 12, 616 00 Brno, CZECH REPUBLIC,
tel. +420 54114 6277, e-mail: jiri.sliz@vutbr.cz

2. Anisotropic diffusion

The commonly used Gaussian noise filter is virtually equivalent to isotropic diffusion; moreover, it embodies the solution of the HE, which can be edited to create an anisotropic diffusion filter. We have the HE formula

$$\frac{\partial u}{\partial t} = \Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \nabla \cdot (\nabla u). \ , \tag{1}$$

where u is the image function depending on the coordinates x and y; t denotes the time; and $s=|\nabla u|$. To maintain the edges during the filtering, a nonlinear part c(s) can be added to the HE; the resulting nonlinear model is represented by the equation

$$\begin{cases} \dfrac{\partial u}{\partial t} = \nabla (c(\mathrm{s}) \nabla u) \\ u(x, y, t = 0) = u_0(x, y). \end{cases} \tag{2}$$

If c(s) = 1, then eq. (2) is the same as eq. (1). In order to preserve the edges, the function c(s) should be chosen to yield small values for a large gradient value and vice versa. This approach is used in PM.

3. Perona-Malik method

The relevant anisotropic diffusion model for image filtering was originally presented in [1]. As mentioned in the previous section, PM basically exploits the modified HE, which, in the discussed case, comprises the nonlinear function c(s). Two options for the c(s) function were presented, namely, c_1(s) and c_2(s); we then have

$$c_1(\mathrm{s}) = \mathrm{e}^{(-\frac{s}{k})^2}, \ c_2(\mathrm{s}) = \frac{1}{1 + \left(\dfrac{s}{k} \right)^2}, \ . \tag{3}$$

The coefficient k represents the diffusivity, and its setting influences the final result; more details are available from source [3]. There is a minor difference between the c_1(s) and the c_2(s): while the former option prefers high-contrast, very sharp edges to weak ones, the latter preserves larger regions rather than small ones. In the basic sense, the applied PM is formulated in eq. (4) below, where the variant c_2(s) is used, and k = 1.

$$\begin{cases} \dfrac{\partial u}{\partial t} = \nabla \left(\dfrac{\nabla u}{1 + |\nabla u|^2} \right), \\ u(x, y, t = 0) = u_0(x, y). \end{cases} \tag{4}$$

The above formulation, however, has a disadvantage in that false objects can be generated due to a high noise level. A solution is proposed in reference [4], whose authors discuss the regularized PM.

4. NMR image filtering

As already pointed out in Section 1, source [2] employs MRI to characterize an embryogenic tissue. The measured images indicate two items: the tissue proper and a small transition between the tissue and the culture medium. In this context, we focus on localizing the transition. Because of the high noise level in the reconstructed MRI images, the filtering is carried out via PM; Gaussian filtering is then applied for comparison, with the SNR utilized to correlate the results. Figure 1 shows a reconstructed MRI image of the tissue and the culture medium.

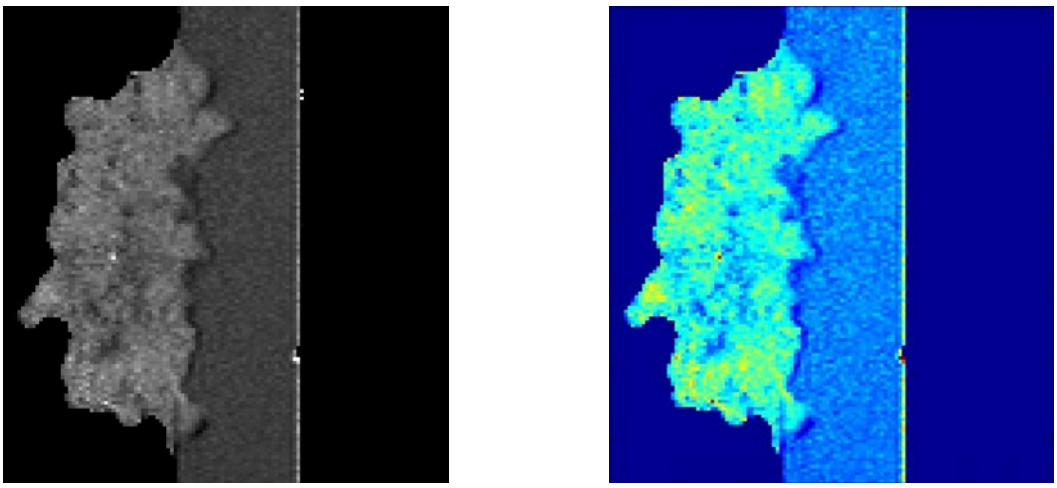


Figure 1 A reconstructed MRI image of the tissue and the culture medium. The grayscale (left) and color (right) versions are identical; the colors are added only to improve the visualization.

Dataset was provided by Institute of Scientific Instruments of the CAS and contains twenty reconstructed images for three samples of axial tissue slices gathered during growing process for T1 and T2 relaxation times. More information in [2]. Images were scaled for better view in figures but original resolution is 132x123 pixels and colorspace is 8-bit grayscale. In addition, color visualization was added to images below so the transition is more visible to human.

The transition detection occurs via simple thresholding preceded by noise reduction to prevent the segmentation result from being influenced by the noise. Firstly, the Gaussian filter is utilized (the filtering and segmentation results are shown in Figure 2).

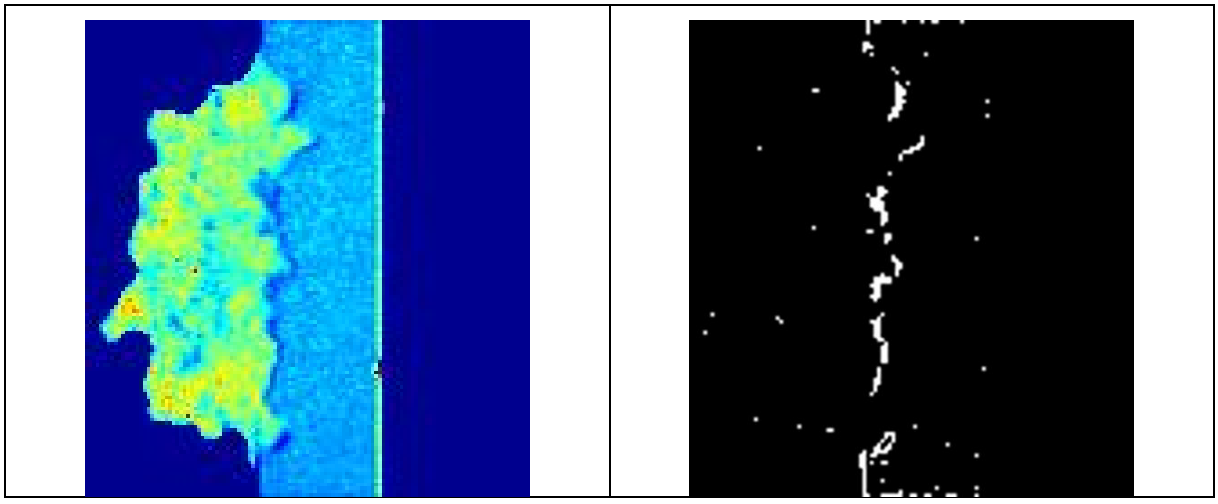


Figure 2 The Gaussian filtering (left) and segmentation of the transition (right).

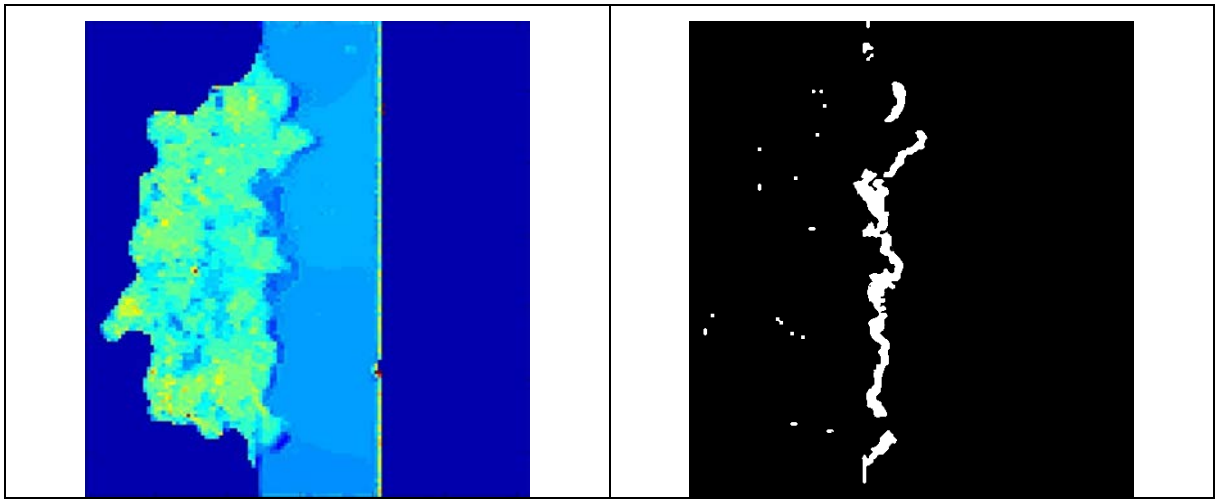


Figure 3 The PM-based result (left) and segmentation of the transition (right).

Secondly, the PM filter is used (the resulting images appear in Figure 3). An overall look at the results yielded through the two methods reveals the advantages of the PM technique. The quantitative comparison of the outcomes is performed by using the SNR, and also the quality segmentation is investigated. The SNR parameter compares the signal level with the noise level. Within this article, the SNR is computed using the formula

$$SNR = \frac{\mu}{s}, \tag{5}$$

where μ is the mean value of brightness in the area of the object (the green region in Figure 1, 2, and 3); and s denotes the standard deviation expressing the noise level of the background (the light blue area). The SNR values for the original image, Gaussian filter, and PM filter are summarized in Table 1; a higher SNR value indicates better filtering, and it is obvious that the PM-based option is more effective than the procedure using the Gaussian filter.

Filter	SNR [-]	Time [s]
None	17.2	0
Gaussian	22.1	0.036
PM	40.6	1.54

Table 1 The SNR parameter values related to time. Ten images were randomly selected from dataset for testing purposes because of manual standard transition creation need. Table shows one representative image results while other testing images had similar values. Image resolution: 132x123 pixels, colorspace: 8-bit grayscale.

The segmentation results are not as clear-cut as those of the filtering process; thus, the parameters SE and SP were chosen for the given purpose. At the initial stage, let us define four types of pixels: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). Using these parameters, the SE and SP can be computed as follows:

$$SE = \frac{TP}{TP + FN}, \quad SP = \frac{TN}{TN + FP}. \tag{6}$$

In the SE, a higher value expresses a better ability to select the pixels of an object; the SP operates similarly, only for the pixels of a background [5]. The SE and SP values in relation to the segmentation results are in Table 2, where the SP parameter is not very descriptive, because the major portion of the segmented image consists in the background; a small pixel selection error thus has a merely negligible impact on the resulting value. Conversely, the SE parameter shows a higher segmentation efficiency after the actual PM filtering.

Filter	**SE** [%]	**SP** [%]
Gaussian	28.4	99.3
PM	93.8	99.7

Table 2 The selectivity and specificity values for the segmentation results. Same images used as in Table 1.

5. Remarks and conclusion

The initial section of the present paper discusses the image filtering procedure proposed by Perona and Malik, namely, a prominent approach using an anisotropic diffusion filter. The main advantage of the method lies in the ability to reduce noise while preserving the sharp edges of the investigated object. In its original form, PM is derived from the heat equation, whose solution leads to a Gaussian filter behaving as isotropic diffusion.

The following parts then describe the filtering and segmentation of the narrow transition in an MRI image containing the embryogenic tissue of the Norway spruce and the related culture medium. Two different filters find use in the experiment: a Gaussian filter behaving as isotropic diffusion, and a PM filter exhibiting traits of anisotropic diffusion. The results are compared using the SNR, with the PM method having a value nearly twice as large as that

yielded by the Gaussian filter. After noise reduction, segmentation via threshold is performed. Consequently, the selectivity and specificity parameters are used to correlate the segmentation results of both the methods. In the Gaussian filter, the selectivity corresponds to 28.4 % - a value which makes the approach unusable for this application. The PM method achieved the specificity of 93.8 %, confirming the assumption that PM is suitable for the described purpose. The specificity parameter is not descriptive in the case of the given image.

References

[1] Perona P, Malik J. Scale-space and edge detection using anisotropic diffusion. IEEE Transactions on pattern analysis and machine intelligence. 1990;12(7):629–639.

[2] Korínek R, Mikulka J, Hrib J, Hudec J, Havel L, Bartušek K. Characterization of the Embryogenic Tissue of the Norway Spruce Including a Transition Layer between the Tissue and the Culture Medium by Magnetic Resonance Imaging. Measurement Science Review. 2017;17(1):19–26.

[3] Samsonov AA, Johnson CR. Noise-adaptive nonlinear diffusion filtering of MR images with spatially varying noise levels. Magnetic Resonance in Medicine. 2004;52(4):798–806.

[4] Catté F, Lions PL, Morel JM, Coll T. Image selective smoothing and edge detection by nonlinear diffusion. SIAM Journal on Numerical analysis. 1992;29(1):182–193.

[5] PROVAZNÍK I. Úvod do biomedicínske informatiky; Pravdepodobnost v rozhodování prednáška. ÚBMI VUT Brno. 2009;.

Dawid BAK[1],Przemyslaw MAZUREK[2]

Antenna Control using Assignment Algorithm for Dynamic Distributed Sensory Network

Monitoring animals' behaviors and live quality using distributed sensory network is possible. Antenna control algorithm for the monitoring station is proposed with the use k–means clustering and compared for 2D clustering and azimuth scenarios. The proposed 2D clustering solution is better two times for small number of assumed clusters up to eight times for more then tens of clusters. Monte Carlo tests are provided for random and sequential initialization of clustering.

1. Introduction

Sensory networks are applied in numerous applications [4, 9]. The availability of low power devices, biometric sensors and GPS system allows the application of sensory network in advanced monitoring of animals [13]. Geofencing is the solution available nowadays for the monitoring of animals in herds, what is important for cost reduction and safety enhancement of animals [1]. Geofencing allows the transmission of animal position data to management computer as well as to local portable devices weared by farmers. The progress in development of biometric sensors allows the measurements of temperature, humidity and estimation of behavior (motion) of animals [17]. Such data could be applied for the monitoring of behaviors and detection of diseases or injuries. Improvement of animal live is important for modern farming and animals live quality.

Such sensory network is dynamic, because animals' changes own positions and attached transmitting devices. One of the most important problems is the power requirement for data transmission [7] and the solar powered device, supported with small rechargeable battery, is the most interesting solution. The lack of direct connection between node (animal) and monitoring station is typical due to land profile as well as signal losses due to trees and bushes. The node antenna should be omnidirectional, because orientation of particular animal is unknown. The problem of transmitting power of node could be reduced using directional antenna at monitoring station. Mechanical scanning (rotating) antenna or electronic scanning antenna (phase array) [16] could be applied for improving directional sensitivity. Such system is depicted in Fig. 1.

The rotation of antenna in full range of azimuth is not necessary in most cases. Only some azimuth ranges should be used for the data reception from nodes. The availability of GPS data allows the estimation of direction to nodes. The rotating antenna can be oriented in specific azimuth at one time. The electronic scanning antenna could be oriented in one specific azimuth or multiple azimuths at one time. The multiple azimuth mode requires more sophisticated digital signal processing device.

[1]M.Sc. Eng., Department of Signal Processing and Multimedia Engineering, West Pomeranian University of Technology Szczecin, Gen. Sikorskiego 37 St., 70–313 Szczecin POLAND, e-mail: dawid.bak@zut.edu.pl

[2]Ph.D. Eng. Sc., Department of Signal Processing and Multimedia Engineering, West Pomeranian University of Technology Szczecin, Gen. Sikorskiego 37 St., 70–313 Szczecin, POLAND, tel: (+48) 91 44 95 307, e-mail: przemyslaw.mazurek@zut.edu.pl

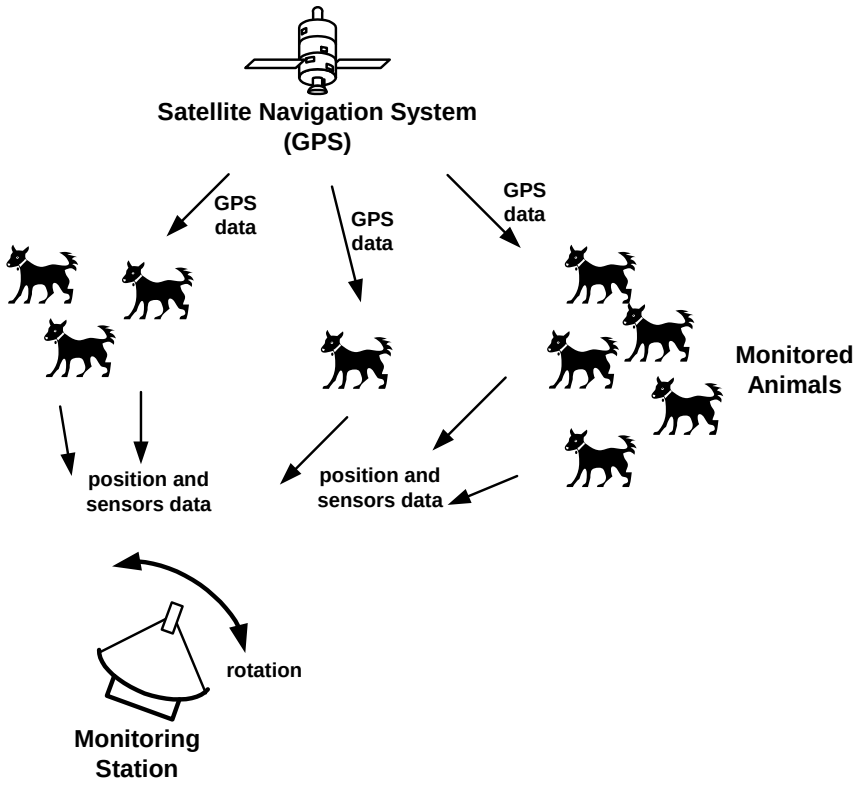


Figure 1: The model of dynamic sensory network.

1.1. Related Works

The problem of tracking animal is similar to the radar system with TWS (Track–While–Scan) mode [2]. TWS mode assumes radar time assignments for specific azimuths related to the detected and tracked object (airplanes or rockets). This radar time assignment is necessary for the improving estimation of object position and velocity (especially for Doppler radar). Second time assignment is related to scanning area, because new objects should be detected if they are in range. TWS is more effective then simple radar that uses rotating antenna with fixed angular velocity.

Dynamic sensory networks are widely used in geofencing [12]. The GPS system connected to the interconnected nodes of the sensory network [15] is ideally suited for creating virtual fences around a geographical area. Such solutions can be successfully used to monitor animals moving around a particular area [5], and by using sensory networks, there is no need for a direct connection to each of the monitored node. In addition, wireless sensory networks are ideal for collecting information about animal behavior and health [18].

1.2. Content and Contribution of the Paper

Assignment algorithm for nodes (animals) using k–means algorithm is proposed and analyzed in this paper. The Monte Carlo approach is applied for the evaluation of properties using synthetic model of herd. A specific initialization is proposed also for dynamic system.

The proposed solution using k–means algorithm is considered in Section 2. Monte Carlo test results are shown in Section 3. Discussion of results is provided in Section 4. Conclusions are

considered in Section 5.

2. Proposed Solution using K–Means Algorithm

The segmentation of 2D distributed observations using unsupervised approach is possible using clustering algorithms, like k–means [14, 10].

This algorithm, depicted in Fig. 2, provides clustering data into predefined number N of groups of animals in particular case. The position of animal obtained from GPS is transmitted to monitoring station and used in next time moment for the prediction of animal position. It is not necessary to use predictors based on the motion model with small movements of animals and frequent data transmissions.

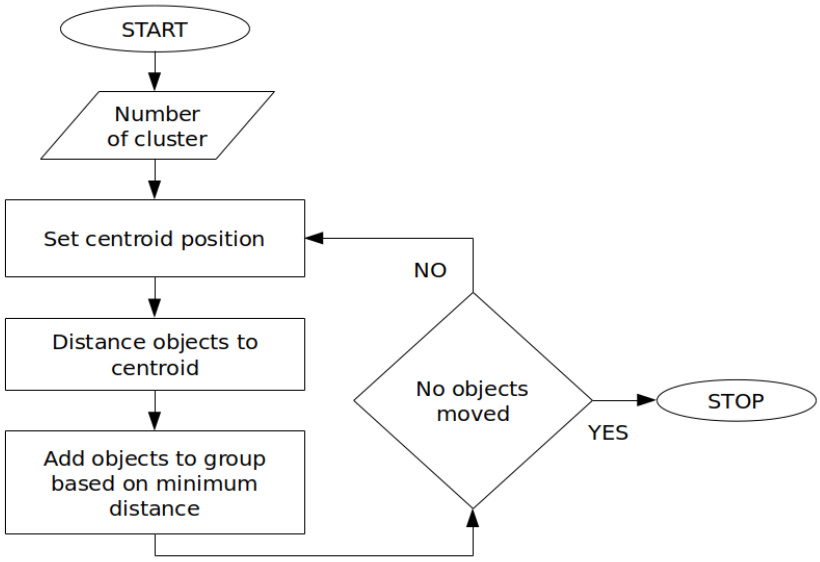


Figure 2: K–means algorithm

There are two possible approaches for clustering. The first variant assumes 2D position and clustering in 2D space. The second variant assumes azimuth based clustering. Both of them are compared in test. The initialization of K–means algorithm assumes random position in 2D space or azimuth 1D space. The obtained position is propagated to next iteration as an initialization.

3. Monte Carlo Tests and Results

Dedicated software tool was developed for the simulation of network. Simulation software is written in C++ language using the SFML (Simple and Fast Multimedia Library) library for rendering graphics. OpenCV library is used for optimal implementation of the K–means algorithm. Developed software uses a swarm algorithm to simulate the behavior of the monitored animals. The use of such approach gives the possibility of analyzing dynamic sensory networks that may arise between partially random moving groups of animals.

The simulation applied in our work uses Monte Carlo approach. A few parameters, like initial position, velocity and direction, are randomly generated using predefined constraints. The behavior of all objects is random in early stages of simulation. Velocities and directions are grouped during starting iterations. The boid algorithm is applied together with random distur-

bances of direction and velocity that are applied individually to every object. The applications of random parameters for initialization reduce the risk of erroneous results due to bias of test. Monte Carlo approach is applied for the initialization of centroids in k-means algorithm and assignment of objects to centers of clusters.

Monte Carlo tests give possibilities of complex system analysis by numerical evaluation of large case sets and random starting conditions. The number of test for particular number of groups N should be large for stable results. Mean value is calculated for particular set. Increasing numbers of test cases stabilize results. In this work there are 2000 cases for each N.

The evaluation criteria could be defined using numerous approaches. The antenna characteristic also influences communication quality as well as modulation, because reduced bitrate with proper selection modulation improves SNR (Signal–to–Noise Ratio). Distributed sensory networks give the possibility of communication using node as a proxy, that is important if poor communication conditions occurs for direct connection between node and monitoring station. The criterion of analysis use the following formula after clustering for every test case:

$$E_k(N) = \max_{i=1}^{N} \max_{j=1}^{N_j} \left| \overrightarrow{A_{i,j}} - \overrightarrow{C_i} \right|, \tag{1}$$

where: $A_{i,j}$ is the 2D position of animal (node) N_j is the number of nodes inside particular cluster j, C_i is the central node (that is used as proxy) inside cluster. For 2D clustering C_i is nearest existing cluster to cluster computed center. For 1D clustering the cluster center is computed using CoG (Center of Gravity) algorithm. This criterion assumes communication through proxy node from most distant node in cluster to proxy. The mean value of $E(N)$ is computed from $E_k(N)$ and the results are shown in Fig. 3 and Fig. 4.

3.1. Test for random initialization

In this case clustering is initialized randomly and previous clustering results are not applied.

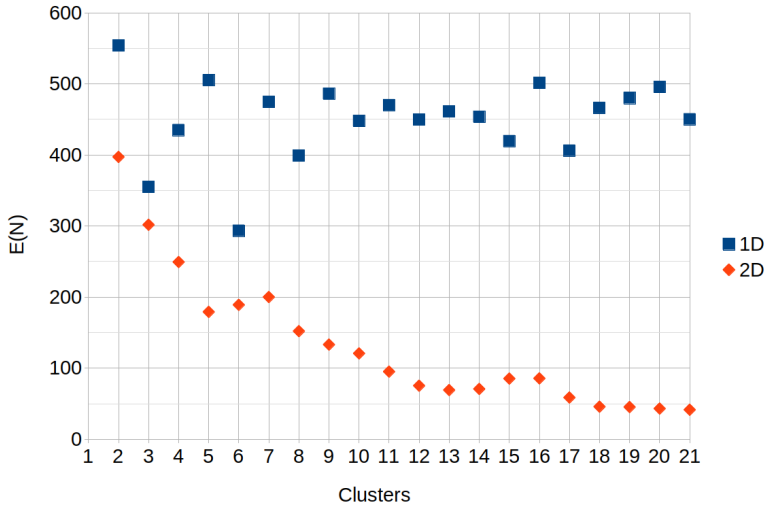


Figure 3: Performance of network for random initialization

3.2. Test for recursive system

In this case clustering is initialized randomly for first iteration and results of clustering are reused in next step.

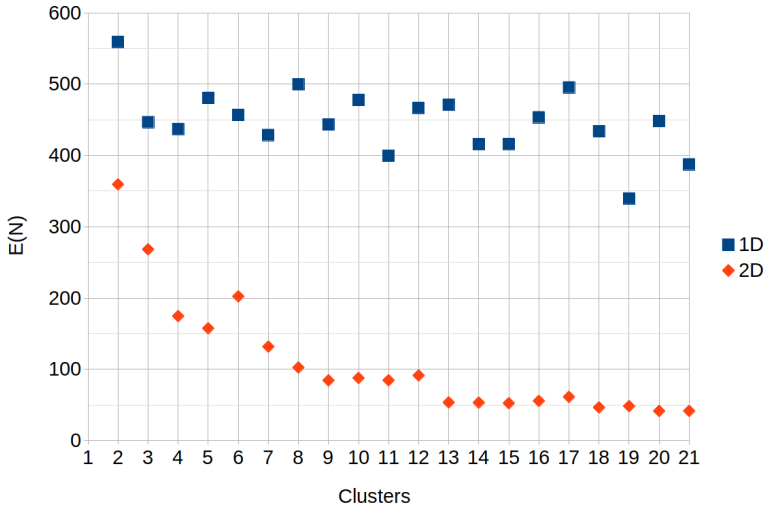


Figure 4: Performance of network for recursive initialization

4. Discussion

Monte Carlo tests shows stable results, after 2000 test for every configuration of cluster numbers. It is expected that errors are higher for small number of assumed clusters. Increasing of clusters number reduces error for all scenarios.

Proposed solution based on 2D clustering gives better result for assumed criteria, comparing to the azimuth only clustering scheme. It shows the advantage of the application of geolocalization data. The communication with proxy is more reliable in such case. The improvement is about two times for small number of clusters Fig. 3 and Fig. 4. The improvement is about eight times for large number of clusters. The advantage of proposed solution is that every proxy is dynamically selected node.

5. Conclusions and Further Work

Proposed solution, based on clustering 2D space, is better in comparison to the azimuth clustering. Considered scenario does not assume predictors for the estimation of possible position of proxy. Numerous factors could be incorporated to the analysis also, like directional sensitivity of antenna, that will be considered in next work.

Additional problem is related to the time of rotation for mechanically rotated antenna. The rotation could be based on steps or could be smoother. Both cases require different control algorithm, that influences assignment algorithm.

Sensory networks with the use of clustering algorithms for the antenna control could be useful in various applications. There are numerous cases where the results presented in this paper could be applied. One of the exemplary applications is wireless communication systems for

military use [9, 6, 11]. The application of sensory network allows the acquisition of soldiers' positions as well as biometric data. Such systems are used nowadays, and are important for the improvement of soldiers' effectiveness and safety. Another possible application is the communication with the group of autonomous mobile robots [19]. The transmission power is the important factor [8] for the design of power limited robots. The application of proposed algorithm reduces power requirements necessary for the communication purposes. It provides to the extension of working time of entire system. Sensory network devices could be spread from the air on ground [3] of Earth or surface of other celestial objects. The communication with such localization of static or dynamic network is difficult and proposed solution is important for the improvement of communication.

Acknowledgments

This work is supported by the UE EFRR ZPORR project Z/2.32/I/1.3.1/ 267/05 "Szczecin University of Technology – Research and Education Center of Modern Multimedia Technologies" (Poland).

References

1. I. Andonovic, Kae Hsiang Kwong, C. Michie, and M. Gilroy. Monitoring of farm animals with wireless sensor networks. In *2010 International Conference on Advanced Intelligence and Awarenss Internet (AIAI 2010)*, pages 5–5, Oct 2010.

2. G. L. Bair and E. D. Zink. Radar track-while-scan methodologies. In *IEEE Region 5 Conference, 1988: 'Spanning the Peaks of Electrotechnology'*, pages 32–37, Mar 1988.

3. S. Birtane, S. Kazdal, and O. K. Sahingoz. 2d coverage analysis of sensor networks with random node deployment. In *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*, pages 1–5, Sept 2017.

4. D. Chaudhari and V. Musale. Novel approach towards securing wireless sensor network systems for health care. In *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*, pages 164–168, Dec 2016.

5. S. Ehsan, K. Bradford, M. Brugger, B. Hamdaoui, Y. Kovchegov, D. Johnson, and M. Louhaichi. Design and analysis of delay-tolerant sensor networks for monitoring and tracking free-roaming animals. *IEEE Transactions on Wireless Communications*, 11(3):1220–1227, March 2012.

6. J. Hörst, F. Govaers, B. Rehbein, C. Barz, and P. Sevenich. Joint protection of a military formation using heterogeneous sensors in a mobile ad hoc network: Concept and field tests. In *2016 International Conference on Military Communications and Information Systems (ICMCIS)*, pages 1–8, May 2016.

7. J. Jayarajan and S. Prabhu. Comparison of energy minimization techniques in wireless sensor networks. In *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*, pages 593–598, Dec 2016.

8. J. K. Deepak Keynes and D. S. Punithavathani. Clustering methodology to prolong lifetime in wireless sensor networks. In *2017 International Conference on Information Communication and Embedded Systems (ICICES)*, pages 1–8, Feb 2017.

9. C. V. Mahamuni. A military surveillance system based on wireless sensor networks with extended coverage life. In *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*, pages 375–381, Dec 2016.

10. P. Milder. Memocode 2016 design contest: K-means clustering. In *2016 ACM/IEEE International Conference on Formal Methods and Models for System Design (MEMOCODE)*, pages 125–127, Nov 2016.

11. W. Pawgasame. A survey in adaptive hybrid wireless sensor network for military operations. In *2016 Second Asian Conference on Defence Technology (ACDT)*, pages 78–83, Jan 2016.

12. M. Rao, N. K. Kamila, and K. V. Kumar. Underwater wireless sensor network for tracking ships approaching harbor. In *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, pages 1872–1876, Oct 2016.

13. S. Sarangi, A. Bisht, V. Rao, S. Kar, T. K. Mohanty, and A. P. Ruhil. Development of a wireless sensor network for animal management: Experiences with moosense. In *2014 IEEE International Conference on Advanced Networks and Telecommuncations Systems (ANTS)*, pages 1–6, Dec 2014.

14. Zahra Zandieh Shirazi and Seid Javad Mirabedini. Dynamic kmeans algorithm for optimized routing in mobile ad hoc networks. In *International Journal of Computer Science and Engineering Survey (IJCSES) Vol.7, No.2*, April 2016.

15. R. Singh and G. M. Asutkar. Survey on various wireless sensor network techniques for monitoring activities of wild animals. In *2015 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, pages 1–5, March 2015.

16. R. Sturdivantt and E. K. P. Chong. System latency performance of mechanical and electronic scanned antennas for leo ground stations for iot and internet access. In *2017 Topical Workshop on Internet of Space (TWIOS)*, pages 1–4, Jan 2017.

17. T. Volk and D. Jansen. Implantable rfid sensor platform to monitor vital functions of small animals controlled by network based software. In *Smart SysTech 2012; European Conference on Smart Objects, Systems and Technologies*, pages 1–6, June 2012.

18. H. Wang, A. O. Fapojuwo, and R. J. Davies. A wireless sensor network for feedlot animal health monitoring. *IEEE Sensors Journal*, 16(16):6433–6446, Aug 2016.

19. A. Wichmann, T. Korkmaz, and A. S. Tosun. Robot control strategies for task allocation with connectivity constraints in wireless sensor and robot networks. *IEEE Transactions on Mobile Computing*, PP(99):1–1, 2017.

Marcin PĘKAL[1], Janusz FRĄCZEK[2]

Nullspace Method for the Reaction Forces Uniqueness Analysis in Rigid Multibody Systems with Redundant Nonholonomic Constraints

A nullspace method for Jacobian-based reaction uniqueness analysis of redundant rigid multibody systems (MBS) is presented. So far, the method was used for holonomic systems only. The aim of this paper is to show that this method may be extended for the case of linear nonholonomic systems. The approach is applicable for planar or spatial MBS described by absolute or natural coordinates. In order to illustrate the method, planar MBS with knife-edge Pfaffian constraints is considered.

1. Introduction

Algorithms for rigid multibody systems (MBS) are in general much less computationally expensive than those for flexible MBS. However, if the considered system is redundantly constrained (which is very common), rigid body approach is subjected to some limitations. One of them is a problem with numerical simulations. The simulations of redundantly constrained MBS must be carried out with the use of appropriate algorithms [9, 10, 12, 13] . It should be emphasized that some papers are devoted to algorithms which may be used to analyze systems without redundant constraints (see, e.g. [1]) or to study serial systems only, e.g. [2]. The other problem which is considered in this paper is a reaction indeterminacy (see, e.g. [4, 5, 7, 18, 21]). This feature is very important because it may lead to unrealistic results of simulations or faults in the design process [4, 17, 18, 21, 22]. Non-uniquelly calculated reactions may lead to false conclusions (e.g. when bearings for the mechanism are selected) or may affect results of analysis of motion, e.g. when the friction in joints is considered (it is worth noting that friction itself may cause problems with solution uniqueness which in some situations are exactly equivalent to problems caused by redundant constraints – see, e.g. [19, 20]). However, despite the general indeterminacy of reactions, some of them (or their components) may be always uniquely determined [17, 18, 21, 22].

There are methods which allow to detect unique reactions. They are based on the analysis of Jacobian matrix of constraints (see, e.g. [4, 15, 18]) or on kinetostatic equations (described, e.g. in [14]). These methods use one of two approaches based on linear algebra, i.e. direct sum or nullspace approach.

In this paper, a nullspace-based method of Jacobian matrix analysis is studied. The method may be used for planar or spatial mechanisms described by means of absolute or natural coordinates. Herein, absolute coordinates are used. In some works (see, e.g. [4, 6]), the nullspace method for reaction uniqueness analysis was formulated for single reaction components (Lagrange multipliers). It was stated that every Lagrange multiplier laying in the nullspace of

[1]Ph.D. student, Institute of Aeronautics and Applied Mechanics, Warsaw University of Technology, Nowowiejska 24, 00-665 Warsaw, POLAND, tel. +48 22 234 79 25, e-mail: mpekal@meil.pw.edu.pl

[2]Prof., Institute of Aeronautics and Applied Mechanics, Warsaw University of Technology, Nowowiejska 24, 00-665 Warsaw, POLAND, tel. +48 22 234 71 36, e-mail: jfraczek@meil.pw.edu.pl

transposed Jacobian matrix of constraints is non-unique. Unfortunately, in many cases it is not enough to study the Lagrange multipliers. For example, the total joint reaction in a compound joint with non-unique reaction components may be unique. Therefore, in [15], the nullspace method was extended to encompass such cases. The extension was made for holonomic systems only.

So far, it was shown in [22] that the direct sum method is applicable for nonholonomic systems with linear constraints (only Pfaffian constraints which are linear and scleronomic were considered). The aim of this paper is to show that the nullspace method can be used for linear nonholonomic systems as well. In order to illustrate the approach, a planar nonholonomic mechanism with knife-edge Pfaffian constraints is modelled and considered.

The structure of the publication is as follows. The next section introduces basic equations for the constraints described by absolute coordinates, Section 3. presents the nullspace approach. In Sect. 4. the exemplary system is considered. Finally, Sect. 5. contains conclusions.

2. Constraints in absolute coordinates

A vector of absolute coordinates used to describe the considered system may be written as [8]

$$\mathbf{q}_{n\times 1} = \left[\mathbf{q}_1^T \ \mathbf{q}_2^T \ \cdots \ \mathbf{q}_N^T\right]^T = \left[q_1 \ q_2 \ \cdots \ q_n\right]^T, \tag{1}$$

where N and n are numbers of bodies and coordinates, respectively.

In this formulation, the considered system is treated as a set of bodies subjected to constraints that describe kinematic pairs. A vector of holonomic position constraints has the following form [3, 8]

$$\mathbf{\Phi} = \mathbf{\Phi}(\mathbf{q}, t) = \mathbf{0}, \tag{2}$$

which, after differentiation, yields [8]

$$\dot{\mathbf{\Phi}} = \mathbf{\Phi}_\mathbf{q}\dot{\mathbf{q}} + \mathbf{\Phi}_t = \mathbf{0}, \tag{3}$$

where $\mathbf{\Phi}_\mathbf{q}$ is a Jacobian matrix of constraints later on used in the joint reactions uniqueness test.

When the considered system is additionally subjected to linear nonholonomic (Pfaffian) constraints, the following equations at velocity level must be fulfilled [22]

$$\mathbf{\Psi}\dot{\mathbf{q}} = \mathbf{0}, \tag{4}$$

where $\mathbf{\Psi}$ is a constant matrix.

Eqs. (3) and (4) may be combined to obtain

$$\begin{bmatrix}\mathbf{\Phi}_\mathbf{q} \\ \mathbf{\Psi}\end{bmatrix}\dot{\mathbf{q}} = \begin{bmatrix}-\mathbf{\Phi}_t \\ \mathbf{0}\end{bmatrix}, \tag{5}$$

where $\left[(\mathbf{\Phi}_\mathbf{q})^T \quad (\mathbf{\Psi})^T\right]^T$ is so-called constraint matrix [22]. Transposition of this matrix is used for uniqueness test. Hence, the following notation may be introduced

$$\mathbf{J} = \left[(\mathbf{\Phi}_\mathbf{q})^T \quad (\mathbf{\Psi})^T\right]. \tag{6}$$

Finally, Lagrange equations of the first kind of the rigid MBS have the form [11]

$$\mathbf{M}\ddot{\mathbf{q}} - \mathbf{J}\boldsymbol{\lambda} = \mathbf{Q}, \tag{7}$$

where $\mathbf{M}$ is a matrix of inertia terms, $\boldsymbol{\lambda}$ is a vector of Lagrange multipliers, $\mathbf{Q}$ is a vector of remaining generalized forces and velocity dependent inertial components. Since $\mathbf{J}\boldsymbol{\lambda}$ represents generalized reactions, the uniqueness of joint reactions may be examined by analyzing matrix $\mathbf{J}$.

3. Nullspace method

If the considered system is redundantly constrained, Jacobian matrix is rank deficient (see, e.g. [9, 10]). This is also true for constraint matrix. In that case, a nullspace matrix $\mathbf{N}$ of matrix $\mathbf{J}$ (which contains basis vectors of the nullspace $\mathscr{N}(\mathbf{J})$) may be calculated. This matrix fulfills the following condition [16]

$$\mathbf{J}\mathbf{N} = \mathbf{0}. \tag{8}$$

Subsequently, a studied element U is selected. It may be, e.g. total joint reaction in a compound joint. For this element, submatrices $\mathbf{J}_U$ and $\mathbf{N}_U$ are selected. Matrix $\mathbf{J}_U$ is built from the columns of $\mathbf{J}$ describing the studied element U, while matrix $\mathbf{N}_U$ is created analogously from the rows of nullspace matrix which correspond to the element U. Eventually, a uniqueness criterion is checked. This criterion has the following form [15]

$$\mathbf{J}_U\mathbf{N}_U = \mathbf{0}. \tag{9}$$

If it is fulfilled, then the studied element is unique. The justification of the criterion for the case of nullspace method based on Jacobian matrix may be found in [15] and is not repeated here. For the constraint matrix, the justification is analogous.

4. Example

To illustrate the nullspace method for the uniqueness analysis based on matrix of constraints, planar nonholonomic mechanism is considered. This system is shown in Fig. 1. It has a mobile platform and a trailer. Such mechanism may be used, e.g. in mobile robotics. It has 7 bodies connected by: 1 translational joint denoted Gt, 6 revolute joints: Ar, Cr, Er, Fr, Hr and Ir respectively, and 1 doubled revolute joint Jr built of 2 revolute joints $Jr1$ and $Jr2$ (and placed in the same point J). Such redundancy may be introduced in real system, e.g. because of the strength reasons. Moreover, nonholonomic constraints are introduced by 5 wheels denoted Aw, Bw, Cw, Dw and Kw, which (for the sake of simplicity) are represented by knife-edge Pfaffian constraints. The main dimension of the considered system is $h = 0.2\ m$ (see Fig. 1). The uniqueness test is performed when the system is in the position shown in Fig. 1. Moreover, vector of absolute coordinates for body i has the form $\mathbf{q}_i = [x_i \quad y_i \quad \varphi_i]^T$ (see Eq. (1)).

For this system, an appropriate vector of constraints $\boldsymbol{\Phi}$ (see Eq. (2)) has to be specified. This vector is not written here due to the limited space, but constraints describing exemplary joints are shown. Hence, the constraints specifying revolute joint Ar have the form

$$\boldsymbol{\Phi}^{Ar} \equiv \left[\mathbf{p}_1 + \mathbf{R}_1\mathbf{r}_A^1 - \left(\mathbf{p}_2 + \mathbf{R}_2\mathbf{r}_A^2\right)\right] = \mathbf{0}, \tag{10}$$

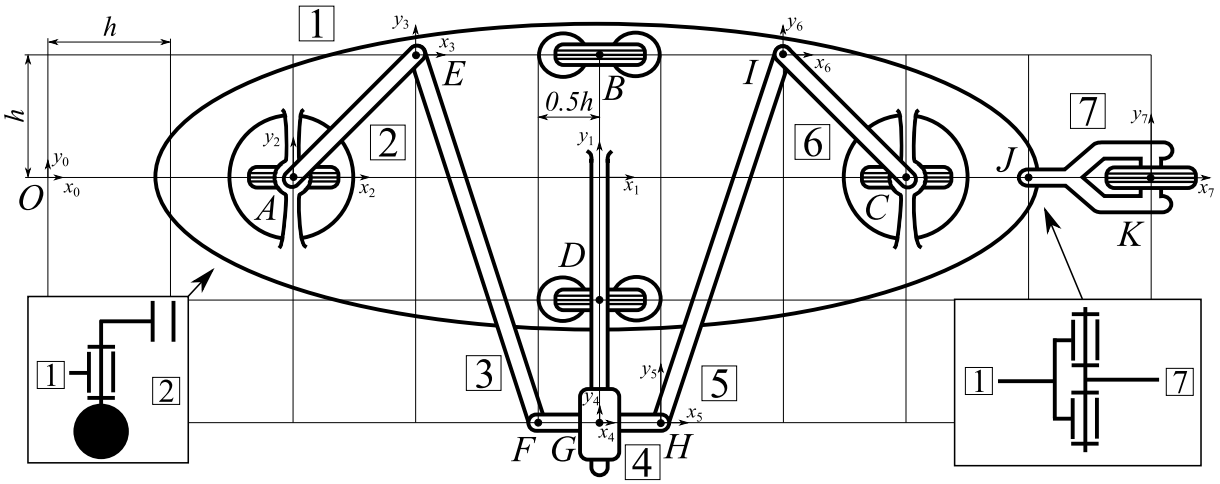


Figure 1: Planar nonholonomic mechanism.

where $\mathbf{p}_i = [x_i \quad y_i]^T$, $\mathbf{r}_K^i$ is a distance vector from the center of the body i reference frame to point K (resolved in local reference frame i) and $\mathbf{R}_i = \begin{bmatrix} cos(\varphi_i) & -sin(\varphi_i) \\ sin(\varphi_i) & cos(\varphi_i) \end{bmatrix}$ is a rotation matrix.

The constraints describing translational joint Gt may be written as

$$\mathbf{\Phi}^{Gt} \equiv \begin{bmatrix} \varphi_1 - \varphi_4 \\ (\mathbf{R}_4\mathbf{v}^4)^T(\mathbf{p}_4 - \mathbf{p}_1 - \mathbf{R}_1\mathbf{r}_{Gt}^1) + (\mathbf{v}^4)^T\mathbf{r}_{Gt}^4 \end{bmatrix} = \mathbf{0}, \tag{11}$$

where $\mathbf{v}^4$ is a vector perpendicular to the axis of joint Gt (resolved in local coordinate frame 4), $\mathbf{r}_{Gt}^1 = [0 \quad 5h]^T$ is a position vector specifying a point laying in the axis of translational joint Gt described in local coordinate frame 1 and $\mathbf{r}_{Gt}^4$ is an analogous position vector. The last vector will vanish in Jacobian matrix, hence its exact value is not introduced here.

Assume that the kinematic constraints are written in vector $\mathbf{\Phi}$ in the following order: Gt, Ar, Cr, Er, Fr, Hr, Ir, $Jr1$ and $Jr2$. Differentiation of constraints vector $\mathbf{\Phi}$ with respect to coordinates $\mathbf{q}$ yields Jacobian matrix of kinematic constraints which may be written as

$$\mathbf{\Phi}_\mathbf{q} = \begin{bmatrix}
\mathbf{0}_{1\times2} & 1 & \mathbf{0}_{1\times2} & 0 & \mathbf{0}_{1\times2} & 0 & \mathbf{0}_{1\times2} & -1 & \mathbf{0}_{1\times2} & 0 & \mathbf{0}_{1\times2} & 0 & \mathbf{0}_{1\times2} & 0 \\
-(\mathbf{R}_4\mathbf{v}^4)^T & \mathbf{a} & \mathbf{0}_{1\times2} & 0 & \mathbf{0}_{1\times2} & 0 & (\mathbf{R}_4\mathbf{v}^4)^T & \mathbf{b} & \mathbf{0}_{1\times2} & 0 & \mathbf{0}_{1\times2} & 0 & \mathbf{0}_{1\times2} & 0 \\
\mathbf{I}_{2\times2} & \mathbf{\Omega}_1\mathbf{r}_A^1 & -\mathbf{I}_{2\times2} & -\mathbf{\Omega}_2\mathbf{r}_A^2 & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} \\
\mathbf{I}_{2\times2} & \mathbf{\Omega}_1\mathbf{r}_C^1 & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & -\mathbf{I}_{2\times2} & -\mathbf{\Omega}_6\mathbf{r}_C^6 & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} \\
\mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{I}_{2\times2} & \mathbf{\Omega}_2\mathbf{r}_E^2 & -\mathbf{I}_{2\times2} & -\mathbf{\Omega}_3\mathbf{r}_E^3 & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} \\
\mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{I}_{2\times2} & \mathbf{\Omega}_3\mathbf{r}_F^3 & -\mathbf{I}_{2\times2} & -\mathbf{\Omega}_4\mathbf{r}_F^4 & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} \\
\mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{I}_{2\times2} & \mathbf{\Omega}_4\mathbf{r}_H^4 & -\mathbf{I}_{2\times2} & -\mathbf{\Omega}_5\mathbf{r}_H^5 & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} \\
\mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{I}_{2\times2} & \mathbf{\Omega}_5\mathbf{r}_I^5 & -\mathbf{I}_{2\times2} & -\mathbf{\Omega}_6\mathbf{r}_I^6 & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} \\
\mathbf{I}_{2\times2} & \mathbf{\Omega}_1\mathbf{r}_J^1 & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & -\mathbf{I}_{2\times2} & -\mathbf{\Omega}_7\mathbf{r}_J^7 \\
\mathbf{I}_{2\times2} & \mathbf{\Omega}_1\mathbf{r}_J^1 & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & \mathbf{0}_{2\times2} & \mathbf{0}_{2\times1} & -\mathbf{I}_{2\times2} & -\mathbf{\Omega}_7\mathbf{r}_J^7
\end{bmatrix} \begin{matrix} \}Gt \\ \\ \}Ar \\ \}Cr \\ \}Er \\ \}Fr \\ \}Hr \\ \}Ir \\ \}Jr1 \\ \}Jr2 \end{matrix}, \tag{12}$$

where $\mathbf{a} = -(\mathbf{R}_4\mathbf{v}^4)^T\mathbf{\Omega}_1\mathbf{r}_G^1$, $\mathbf{b} = -(\mathbf{R}_4\mathbf{v}^4)^T\mathbf{\Omega}(\mathbf{p}_4 - \mathbf{p}_1 - \mathbf{R}_1\mathbf{r}_{Gt}^1)$, $\mathbf{\Omega}_i = \mathbf{\Omega}\mathbf{R}_i$, $\mathbf{\Omega} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, $\mathbf{I}_{2\times2}$ is an identity matrix of dimensions 2×2 and $\mathbf{0}_{i\times j}$ is a zero matrix of dimensions $i\times j$.

Knife-edge Pfaffian nonholonomic constraints are introduced as in article [22]. Hence, the

matrix of nonholonomic constraints for the whole system may be written as

$$\boldsymbol{\Psi}=\begin{bmatrix} \mathbf{0}_{1\times 2} & 0 & (\mathbf{R}_2\mathbf{n}_A^2)^T & (\mathbf{n}_A^2)^T\boldsymbol{\Omega}\mathbf{r}_A^2 & \mathbf{0}_{1\times 9} & \mathbf{0}_{1\times 2} & 0 & \mathbf{0}_{1\times 2} & 0 \\ (\mathbf{R}_1\mathbf{n}_B^1)^T & (\mathbf{n}_B^1)^T\boldsymbol{\Omega}\mathbf{r}_B^1 & \mathbf{0}_{1\times 2} & 0 & \mathbf{0}_{1\times 9} & \mathbf{0}_{1\times 2} & 0 & \mathbf{0}_{1\times 2} & 0 \\ \mathbf{0}_{1\times 2} & 0 & \mathbf{0}_{1\times 2} & 0 & \mathbf{0}_{1\times 9} & (\mathbf{R}_6\mathbf{n}_C^6)^T & (\mathbf{n}_C^6)^T\boldsymbol{\Omega}\mathbf{r}_C^6 & \mathbf{0}_{1\times 2} & 0 \\ (\mathbf{R}_1\mathbf{n}_D^1)^T & (\mathbf{n}_D^1)^T\boldsymbol{\Omega}\mathbf{r}_D^1 & \mathbf{0}_{1\times 2} & 0 & \mathbf{0}_{1\times 9} & \mathbf{0}_{1\times 2} & 0 & \mathbf{0}_{1\times 2} & 0 \\ \mathbf{0}_{1\times 2} & 0 & \mathbf{0}_{1\times 2} & 0 & \mathbf{0}_{1\times 9} & \mathbf{0}_{1\times 2} & 0 & (\mathbf{R}_7\mathbf{n}_K^7)^T & (\mathbf{n}_K^7)^T\boldsymbol{\Omega}\mathbf{r}_K^7 \end{bmatrix}\begin{matrix} \}Aw \\ \}Bw \\ \}Cw, \\ \}Dw \\ \}Kw \end{matrix} \tag{13}$$

where $\mathbf{n}_k^i$ is a vector perpendicular to the allowed direction of motion of the nonholonomic knife-edge Pfaffian constraint k described in local reference frame i.

Hence, constraint matrix (see Eq. (5)) has dimensions 23×21, and rank $r = 19$.

The uniqueness analysis is performed using transposed constraint matrix $\mathbf{J}$. For this matrix, nullspace matrix $\mathbf{N}$ is calculated. It has the following form

$$\mathbf{N}_{23\times 4} = \begin{bmatrix} \mathbf{0}_{4\times 3} & (\mathbf{N}_{1_{1\times 4}})^T & \mathbf{0}_{4\times 1} & (\mathbf{N}_{2_{1\times 4}})^T & \mathbf{0}_{4\times 8} & (\mathbf{N}_{3_{8\times 4}})^T & \mathbf{0}_{4\times 1} \end{bmatrix}^T, \tag{14}$$

where $\mathbf{N}_i$ is a submatrix containing nonzero elements. The exact elements are not shown, because they depend on the algorithm used for the computation of $\mathbf{N}$. From this nullspace matrix, it may be easily seen (see, e.g. [4,6]) that single Lagrange multipliers: λ_4, λ_6, $\lambda_{15} - \lambda_{22}$ are non-unique. However, in many cases, e.g. when the uniqueness of total joint reactions is requested, it is not enough to know the uniqueness of a single Lagrange multiplier. This is because a unique joint reaction may have non-unique components. Hence, the nullspace method should be applied. In Tab. 1 the data for uniqueness test using this approach are shown. Column ***Studied Element*** U specifies the examined element, column ***Constraint Localization*** specifies the row localization of the constraints in Jacobian matrix (which is equivalent to the column localization in matrix $\mathbf{J}$). Column ***Value of*** (9) presents the results of uniqueness test. If the condition is fulfilled (i.e. it equals zero), then the studied element is unique.

Studied Elem. U	***Constr. Loc.***	***Value of*** (9)	***Studied Elem.*** U	***Constr. Loc.***	***Value of*** (9)
Reaction in Gt	1 – 2	$= \mathbf{0}$	Reaction in $Jr2$	17 – 18	$\neq \mathbf{0}$
Reaction in Ar	3 – 4	$\neq \mathbf{0}$	Reaction in Jr	15 – 18	$= \mathbf{0}$
Reaction in Cr	5 – 6	$\neq \mathbf{0}$	Reaction in Aw	19	$\neq \mathbf{0}$
Reaction in Er	7 – 8	$= \mathbf{0}$	Reaction in Bw	20	$\neq \mathbf{0}$
Reaction in Fr	9 – 10	$= \mathbf{0}$	Reaction in Cw	21	$\neq \mathbf{0}$
Reaction in Hr	11 – 12	$= \mathbf{0}$	Reaction in Dw	22	$\neq \mathbf{0}$
Reaction in Ir	13 – 14	$= \mathbf{0}$	Reaction in Kw	23	$= \mathbf{0}$
Reaction in $Jr1$	15 – 16	$\neq \mathbf{0}$			

Table 1: Reaction uniqueness analysis of nonholonomic planar mechanism

The uniqueness analysis of the nonholonomic constraints gave the results consistent with intuition, i.e. reactions in the wheels of the mobile platform (Aw, Bw, Cw and Dw) are non-unique, while the reaction in wheel Kw of the trailer is unique. As may be expected also, reactions in joints $Jr1$ and $Jr2$ are non-unique. However reaction in doubled joint Jr (composed of $Jr1$ and $Jr2$) is unique. It confirms the fact that a unique reaction may have non-unique components. A quite unexpected result is also obtained. Non-unique nonholonomic constraints Aw and Cw cause non-uniqueness of reactions in two revolute joints Ar and Cr.

5. Conclusions

The paper presents an extension of the nullspace method used for the reaction uniqueness analysis. The approach is based on the analysis of the constraint matrix and it is applicable for planar and spatial rigid MBS described by absolute or natural coordinates. Using the method, the range of applicability, and thus fidelity of the rigid body model, may be assessed.

It was shown that redundant rigid MBS with linear nonholonomic constraints may be studied using the presented method. In this paper, nonholonomic constraints in the form of knife-edge Pfaffian are considered. It is useful to point out that the extension presented in this paper significantly improved the applicability of the nullspace method, because the usage of nonholonomic constraints allows us to study broader classes of MBS, e.g. some kinds of vehicles or mobile robots.

6. Acknowledgements

This work has been supported by the Faculty of Power and Aeronautical Engineering Dean's Grant 2017.

References

1. Chadaj K., Malczyk P., Frączek J.: "A parallel Hamiltonian formulation for forward dynamics of closed-loop multibody systems", *Multibody Syst. Dyn.*, 2017, Vol. 39, nr 1–2

2. Chadaj K., Malczyk P., Frączek J.: "A Parallel Recursive Hamiltonian Algorithm for Forward Dynamics of Serial Kinematic Chains", *IEEE T. Robot.*, 2017, Vol. 33, nr 3

3. de Jalón J.G., Bayo E.: "Kinematic and Dynamic Simulation of Multibody Systems, The Real-Time Challenge", Springer-Verlag, *New-York*, 1994

4. de Jalón J.G., Gutiérrez-López M.D.: "Multibody dynamics with redundant constraints and singular mass matrix: existence, uniqueness, and determination of solutions for accelerations and constraint forces", *Multibody Syst. Dyn.*, 2013, Vol. 30, nr 3

5. Featherstone R.: "Rigid Body Dynamics Algorithms", Springer US, 2008

6. Featherstone R., Orin D.E.: "Dynamics". Siciliano B., Khatib O. (Eds.): *Springer Handbook of Robotics*, Springer-Verlag Berlin Heidelberg, 2008

7. González F., Kövecses J.: "Use of penalty formulations in dynamic simulation and analysis of redundantly constrained multibody systems", *Multibody Syst. Dyn.*, 2013, Vol. 29, nr 1

8. Haug E.J.: "Computer aided kinematics and dynamics of mechanical systems, Volume I: Basic methods", Allyn and Bacon, 1989

9. Mariti L., Belfiore N.P., Pennestrì E., Valentini P.P.: "Comparison of solution strategies for multibody dynamics equations", *Int. J. Numer. Meth. Eng.*, 2011, Vol. 88, nr 7

10. Mariti L., Pennestrì E., Valentini P.P., Belfiore N.P.: "Review and Comparison of Solution Strategies for Multibody Dynamics Equations", *The 1st Joint International Conference on Multibody System Dynamics*, Lappeenranta, Finland, May 25–27, 2010

11. Nikravesh P.E.: ”Computer-Aided Analysis of Mechanical Systems”, Prentice Hall, 1988

12. Pękal M., Frączek J.: ”Comparison of natural complement formulations for multibody dynamics”, *Journal of Theoretical and Applied Mechanics*, 2016, Vol. 54, nr 4

13. Pękal M., Frączek J.: ”Comparison of selected formulations for multibody system dynamics with redundant constraints”, *Archive of Mechanical Engineering*, 2016, Vol. 63, nr 1

14. Pękal M., Frączek J., Tomulik P.: ”Solvability of reactions and inverse dynamics problem for complex kinematic chains”, *The 21st International Conference on Methods and Models in Automation and Robotics*, Międzyzdroje, August 29–September 1, 2016

15. Pękal M., Frączek J., Wojtyra M.: ”Nullspace Method for the Analysis of Uniqueness of Reactions and Driving Forces in Redundantly Constrained Multibody Systems”, *8th ECCOMAS Thematic Conference on MULTIBODY DYNAMICS*, Prague, June 19–22, 2017

16. Strang G., Borre K.: ”Linear Algebra, Geodesy, and GPS”, Wellesley-Cambridge Press, 1997

17. Wojtyra M.: ”Joint Reaction Forces in Multibody Systems with Redundant Constraints”, *Multibody Syst. Dyn.*, 2005, Vol. 14, nr 1

18. Wojtyra M.: ”Joint reactions in rigid body mechanisms with dependent constraints”, *Mech. Mach. Theory*, 2009, Vol. 44, nr 12

19. Wojtyra M.: ”On Some Problems With Modeling of Coulomb Friction in Self-Locking Mechanisms”, *J Comput. Nonlin. Dyn.*, 2016, Vol. 11, nr 1

20. Wojtyra M.: ”Modeling of static friction in closed-loop kinematic chains—Uniqueness and parametric sensitivity problems”, *Multibody Syst. Dyn.*, 2017, Vol. 39, nr 4

21. Wojtyra M., Frączek J.: ”Joint reactions in rigid or flexible body mechanisms with redundant constraints”, *Bull. Pol. Ac.: Tech.*, 2012, Vol. 60, nr 3

22. Wojtyra M., Frączek J.: ”Solvability of reactions in rigid multibody systems with redundant nonholonomic constraints”, *Multibody Syst. Dyn.*, 2013, Vol. 30, nr 2

Edyta ROLA[1]

The effect of child position in child restraint system on safety

The objective of this study was to investigate the effect of selected positions of a child in a CRS on the injury potential during the frontal crash. The Q3 dummy model travelling in a forward-facing child restraint systems in the rear seat of a passenger car was considered. The dummy model kinematic response and the serious injury risk in each case were compared. It was concluded that child's position prior to the crash influences the injury response of Q3 dummy model.

1. Introduction

Child restraint systems (CRSs) are designed to provide protection for child occupants in the event of a crash. The effectiveness of CRSs is influenced by many variables [4][5][9]. Several studies found out that the position of a child which is different from a standard dummy position may have influence on the kinematics and passive safety of a child [1][7][6][14]. The objective of this study was to investigate the effect of selected positions of a child in a CRS on the injury potential.

The research presented in this paper focuses on the frontal impact, which is the most frequent type of crash. In this study, particular attention was paid to the neck injury criteria. Cervical spine injuries are not common but because of the tendency for occurring in the upper cervical spine in young children, the effects can be severe [2][10]. In the ECE Regulation No. 129 (Date of entry into force: 9 February 2017) upper neck tension force and upper neck flexion moment for frontal impact are for monitoring purpose only [3]. In this study, neck injury criteria based on additional components of neck load were taken into account.

2. Methods

The Q3 dummy model travelling in a forward-facing CRSs in the rear seat of a passenger car was considered. The frontal impact was configured according to the European regulation ECE R129 (with a closing speed of about 50 km/h) [3] (Figure 6 in Appendix). Before the impact pre-crash braking with acceleration of 0.7g was simulated. The following cases were investigated:

a) Reference case - standard dummy position - Case A (Figure 4 in Appendix),
b) Head lean right - Case B (Figure 4 in Appendix),
c) Head lean right and hand near head - Case C (Figure 5 in Appendix),
d) Limbs out of positions - Case D (Figure 5 in Appendix).

[1] Phd student, Warsaw University of Technology, Institute of Aeronautics and Applied Mechanics, ViSEB Lab, ul. Nowowiejska 24, 00-665 Warszawa, POLAND, tel. +48 22 234 7925, e-mail: erola@meil.pw.edu.pl

A series of numerical simulations were carried out with the use of MADYMO v7.5.1 software [8]. A facet CRS model was taken from previous research [13]. During the simulations, injury criteria connected with the head, neck, chest and pelvis were controlled.

The defined injury criteria connected with the head were: the highest center of gravity of head acceleration that is exceeded during at least 3ms - Cumulative 3ms Head Injury Criterion (Head_3ms) and the Head Injury Criterion with the time interval of 15ms (HIC_15). Injury criterion related to the chest was Cumulative 3ms Thorax Injury Criterion (Chest_3ms) – chest resultant acceleration that is exceeded during at least 3ms. Injury criterion connected to the pelvis was Cumulative 3ms Pelvis Injury Criterion (Pelvis_3ms) - pelvis resultant acceleration that is exceeded during at least 3ms. Injury criteria for the neck were: Neck Injury Predictor - Tension-Extension (N_TE) and Tension-Flexion (N_TF), Compression-Extension (N_CE) and Compression-Flexion (N_CF), Neck Injury Criterion - Tension Negative (NIC_TN) and Positive (NIC_TP), Shear Negative (NIC_SN) and Shear Positive (NIC_SP), Bending Negative (NIC_BN) and Positive (NIC_BP). The defined injury criteria were based on the signals that may interact with the Q3dummy measurements. Reference values at which injury may occur with probability of about 10-20 percent of AIS3+ injury are called critical values. These critical values for a 3-year-old child are presented in Table I [11][12].

TABLE I
PROPOSED INJURY CRITERIA AND THEIR CRITICAL VALUES

	Injury Criterion	Units	Critical Value
1	*Head_3ms*	g	80
2	*Chest_3ms*	*g*	55
3	*Pelvis_3ms*	*g*	80
4	*HIC_15*	-	800
5	*N_TE*	-	1
6	*N_TF*	-	1
7	*N_CE*	-	1
8	*N_CF*	-	1
9	*NIC_TN*	*N*	1130
10	*NIC_TP*	*N*	1130
11	*NIC_SN*	*N*	500
12	*NIC_SP*	*N*	500
13	*NIC_BN*	*Nm*	16
14	*NIC_BP*	*Nm*	16

In order to compare injury risk in different cases the resulting value of each injury was divided by its critical value which was based on Table I. The effectiveness of configurations was estimated by comparing the normalized values of all injury criteria. Q3 dummy model responses were monitored for 200ms from the initiation of the frontal impact pulse. All signals were filtered (SAE J211-1, CFC filters).

3. Results

For all cases the head resultant acceleration is presented in Figure 1. Most cases have a similar peak value of head and chest acceleration. Only the Case D has different acceleration history of the head: the first peak is lower and the second occurred earlier. Figure 2 shows the Q3 model chest acceleration in each case. The highest value of chest acceleration can be observed for the Case D (limbs out of positions) and maxiumum value occured the earliest. In other cases both timing and maximum value of the chest resultant acceleration are similar.

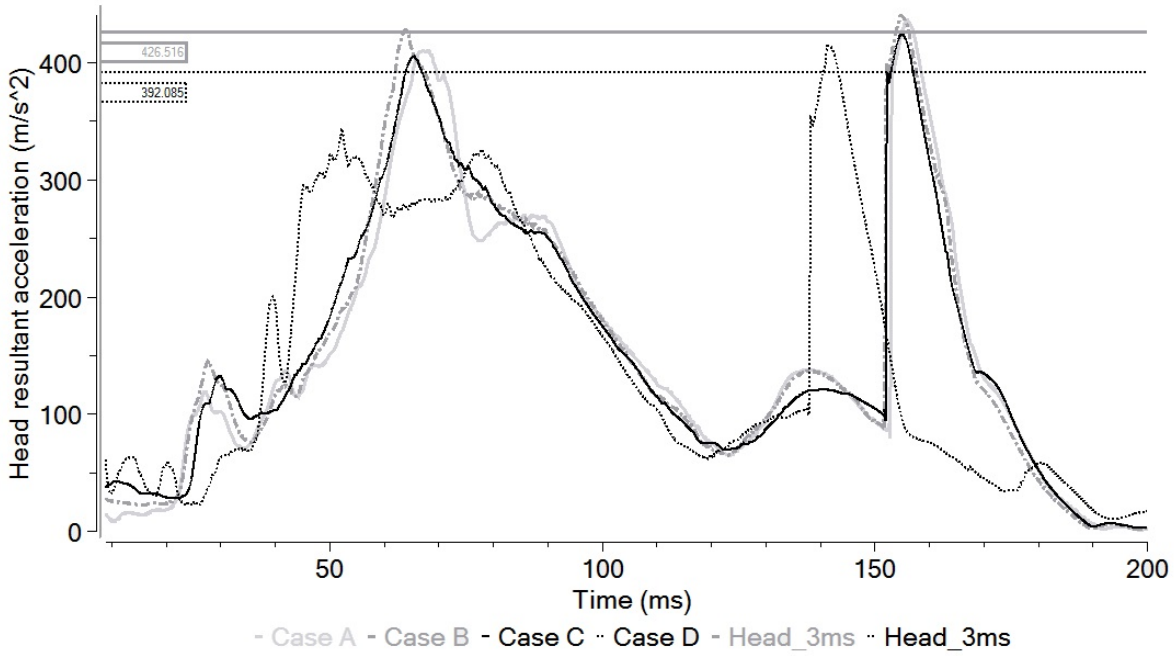


Figure. 1: Time histories of the center of gravity of the head resultant acceleration for Cases A-D.

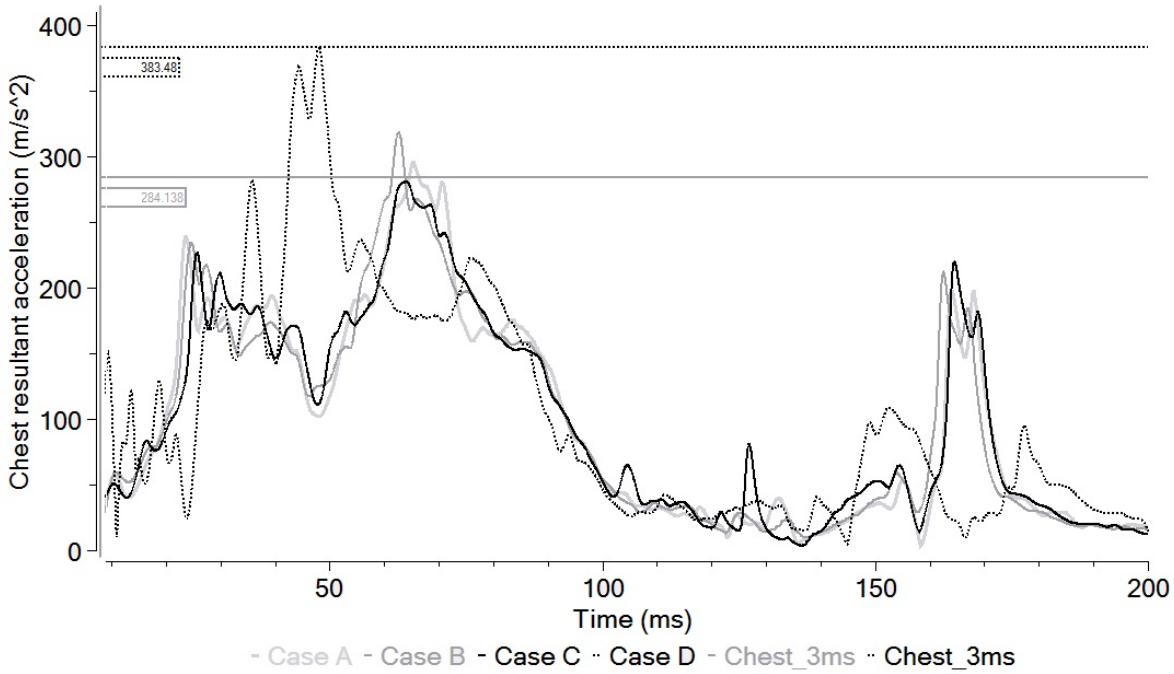


Figure. 2: Time histories of the center of gravity of the chest resultant acceleration for Cases A-D.

The injury criteria of each case are shown in Figure 3. In Cases A, C and D four of the injury criteria referring to the neck exceeded their limits. In Case B three injury criteria reached critical values. It should be emphasized that not only the criteria of injury based on

tensile and bending forces have reached critical values but also those based on the shear force. Table II in Appendix shows all normalized values of injury criteria.

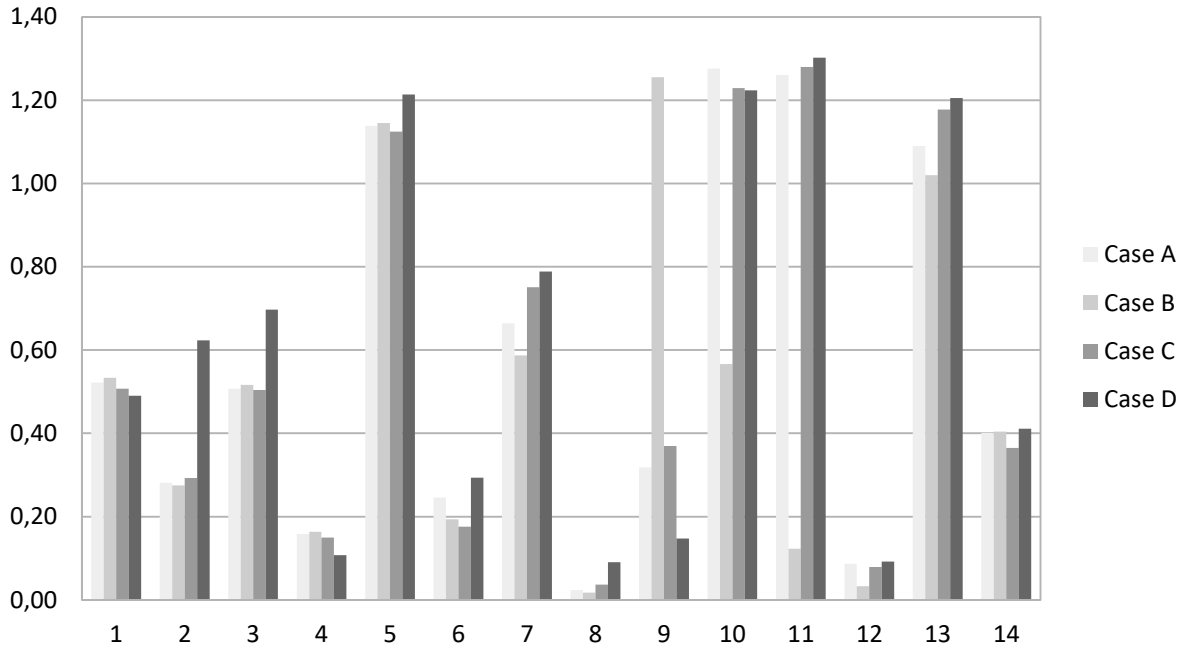


Figure. 3: Normalized values of injury criteria.

4. Remarks and conclusion

To provide a real life safety effect, the injury mitigation systems should provide appropriate protection in the range of pre-crash sitting position and natural variation between individuals in population [11]. The obtained results suggest that child's position prior to the crash influences the injury response of Q3 dummy model. Moreover, some of child occupant positions significantly increase the likelihood of occurring a serious injury.

It was noticed that in all cases the critical value of some neck injury criteria were exceeded. Despite of considering frontal impact not only neck tension force and upper neck flexion moment reached high values. This was due to the asymmetry of the position of Q3 dummy model. The current regulation requirements do not take into account various positions of children, so they are significantly different from the real conditions in which children travel and need to be improved. However, the research deals with uncertainties which are common in car accident analyses. Further research is essential both for numerical simulations and experiments.

References

1. Charlton J., Koppel S., Kopinathan C., Taranto D.: *How Do Children Really Behave in Restraint Systems While Travelling in Cars?*, 54th AAAM Annual Conference Annals of Advances in Automotive Medicine, 2010
2. Ching R., Nuckley D., et al.: *Tensile Mechanics of the Developing Cervical Spine*, 45th Stapp Car Crash Journal, Vol. 45 (P–375), San Antonio, 2001

3. ECE Regulation No. 129, *Uniform provisions concerning the approval of enhanced Child Restraint Systems used on board of motor vehicles (ECRS)*, 7 August 2013
4. Elliott M., Kallan M. et al.: *Effectiveness of child safety seats vs seat belts in reducing risk for death in children in passenger vehicle crashes*, Archives of Pediatrics and Adolescent Medicine, 160:617–621, 2006
5. Hummel T., Langwieder K. et al.: *Injury risks, misuse rates and the effect of misuse depending on the kind of child restraint system*, Proceedings of the 41st Stapp Car Crash conference, Lake Buena Vista, 1997
6. Lesire P., Cassan F. et al.: *TUB Report on relevant children injury in road accidents and specification of children models*, CASPER-30 MARCH 2010-UDs-WP2-DEL2.1.1_v2, 2010
7. Lesire P., Grant R., Hummel T.: *The CREST project accident data base*, 17th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Amsterdam, The Netherlands, 2001
8. MADYMO Manuals, ver. 7.5, TASS International 2013
9. Morris S., Arbogast K. et al.: *Misuse of booster seats*, Injury Prevention, 6:281–4, 2000
10. Pintar F., Mayer R., Yoganandan N.: *Child Neck Strength Characteristics Using an Animal Model*, 44th Stapp Car Crash Conference, Atlanta, 2000
11. Rola E., Rzymkowski C.: *Effectiveness of the Child Restraint System with a Special Airbag and Smart Seatbelt Pretensioner in Frontal Collisions*, w: IRCOBI Conference Proceedings, International Research Council on Biomechanics of Injury, vol. on-line, ss. 101-113, 2015
12. Rola E.: *Study of 3-Year-Olds in a Child Restraint System with Harness Pretensioner and Load Limiter*, w: Proceedings of 2016 IRCOBI Conference, IRCOBI, ss. 1-2, 2016
13. Świetlik M., Rzymkowski C., Kędzior K.: *Efficiency of moveable type of Child restraint Systems In the event of Head-on Collision*, 13th World Congress in Mechanism and Machine Science, Guanajuato, México, 2011
14. Van Rooij L., Harkema C., et al.:.: *Child poses in child restraint systems related to injury potential: investigations by virtual testing*, 19th International technical conference on the enhanced safety of vehicles (ESV), Washington, 2005

Appendix

Figure. 4: Reference position – Case A (left) and Head lean right position (right)) – Case B.

Figure. 5: Head lean right and hand Case C (left) and Limbs out of positions (right) – Case D.

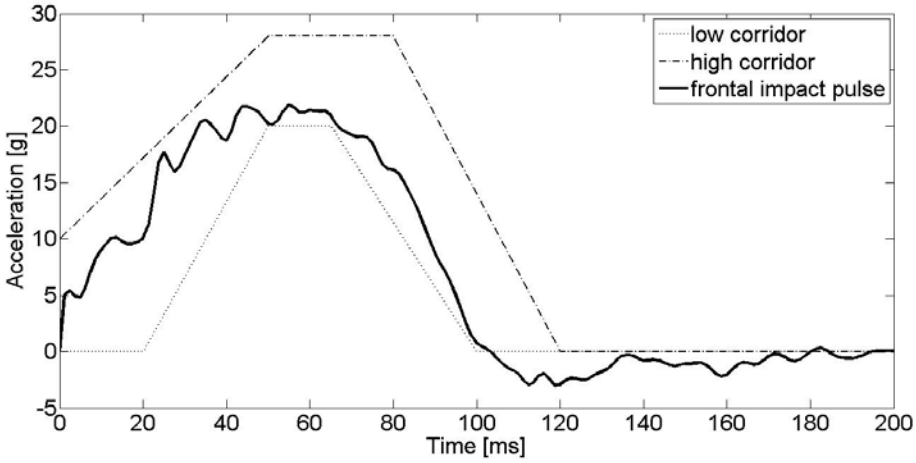


Figure. 6: The impact pulse with corridors.

TABLE II
NORMALIZED VALUES OF INJURY CRITERIA

Injury Criterion	Injury Criterion Value			
	Case A	Case B	Case C	Case D
Head_3ms	0,52	0,53	0,51	0,49
Chest_3ms	0,28	0,27	0,29	0,62
Pelvis_3ms	0,51	0,52	0,50	0,70
HIC_15	0,16	0,16	0,15	0,11
N_TE	**1,14**	**1,14**	**1,12**	**1,21**
N_TF	0,25	0,19	0,18	0,29
N_CE	0,66	0,59	0,75	0,79
N_CF	0,02	0,02	0,04	0,09
NIC_TN	0,32	**1,25**	0,37	0,15
NIC_TP	**1,28**	0,57	**1,23**	**1,22**
NIC_SN	**1,26**	0,12	**1,28**	**1,30**
NIC_SP	0,09	0,03	0,08	0,09
NIC_BN	**1,09**	**1,02**	**1,18**	**1,21**
NIC_BP	0,40	0,40	0,36	0,41

Grzegorz DRAŁUS[1], Zbigniew GOMÓŁKA[2], Damian MAZUR[3], Andrzej SMOLEŃ[4]

Seasonal models of energy forecasting in photovoltaic systems using artificial neural networks

The paper presents methods of energy forecasting in photovoltaic systems. A short-term forecast e.g. a day-ahead forecasting is presented. A forecasting methods that uses MLP and RBF neural networks are studied. A comprehensive set of explanatory variables for forecasting models are presented. The parameters of the forecasting model were determined and forecasts were made for several time periods (e.g. –year-round forecast, seasonal forecasts) and final conclusions were formulated.

1. Introduction

The production of energy from renewable sources, especially in the case of photovoltaic panels, is strongly dependent on weather conditions. Changes in the size of this production must be compensated on a regular basis through conventional sources using regulatory reserves or in the future through the use of energy storage. Particularly dangerous for the power system can be short-term changes in the amount of energy produced from renewable sources due to sudden changes in weather. Therefore, with the increase in the share of renewable energy in the total energy balance, it is increasingly important to forecast its production in the short term and to optimize production in real time.

Forecasting the production of electricity from renewable sources can be considered in several different horizons: ultra-short-term, short-term (up to several dozen hours), medium-term (from several to dozens of days), long-term (from several months). The terminology used for forecasting, especially for renewable energy sources, may be different depending on the application of the forecast or their complementarity. For example, short-term forecasts are used to plan the work of photovoltaic sources and the power system. Therefore, forecasting a one day-ahead is a category of short-term forecasts.

In addition to the time horizons, the quantities that are subject to forecasting are also important. Of course, the forecasted volume depends on the prediction objective. For short-term forecasts, average power values in portions of 15 minutes to 1 hour are usually used. For forecasts with longer horizons peak powers and cumulative energy produced in the longer term can be used. In case of the one day-ahead forecasts are usually used total solar irradiation in a day, total precipitation in a day, total energy generated on a given day, and average values (e.g. temperature, visibility, pressure, wind speed, cloudiness) or maximum values (e.g. temperature).

[1] Ph.D, Eng., Rzeszow University of Technology, ul. Pola 2, 35 595 Rzeszów, POLAND,
tel. +48 501 332 657, fax: +(48) 17 854 20 88, e-mail: gregor@prz.edu.pl
[2] Ph.D, University of Rzeszów, ul. Prof. Stanisława Pigonia 1, 35-959 Rzeszów POLAND,
tel. +(48), 48 17 872 10 00 fax: +(48) 17 872 12 65, e-mail: zgomolka@ur.edu.pl
[3] DSc, Ph.D, Eng., Rzeszow University of Technology, ul. Pola 2, 35 595 Rzeszów, POLAND,
tel. +(48) 17 743 2469, fax: +(48) 17 854 20 88, e-mail: mazur@prz.edu.pl
[4] Ph.D. student, Rzeszow University of Technology, 35-595 Rzeszów, POLAND
tel. +(48) 17 743 2469, fax: +(48) 17 854 20 88, e-mail: a.smolen@prz.edu.pl

Neural networks are the most popular tool among artificial intelligence methods for forecasting electrical energy generation. Their main advantage is the ability to generalize knowledge based on learning data, typically derived from the measurement data of objects or systems. In addition, they enable us to designate model parameters through a learning process. Their disadvantage is the lack of knowledge about the object under investigation [3].

The use of artificial neural networks for energy forecasting in photovoltaic systems is very popular. Different types of neural networks are used, and neural network combinations with fuzzy logic or genetic algorithms are also used. The choice of the network type depends on the horizon of forecasts, the availability of various historical data and forecasts from Numerical Weather Prediction (NWP).

The most commonly used neural network for predicting the amount of energy produced in photovoltaic systems is the MLP neural network (Multilayer Perceptron) [1, 2, 4, 7]. Other types of neural networks, such as RBF neural networks, are also used in forecasting [2, 7, 8]. In addition, support vector machines (SVM) are used [2, 5], recurrent neural networks [8]. Very promising results are provided by methods using several neural network models (several different learning algorithms) to obtain a final prediction. This solution was confirmed by using a set of four neural networks [7]. Physical models [4] and time series analysis [6] are also apply to energy forecasting. The physical model uses the conversion of the irradiation forecast to the energy forecast in the solar PV system.

2. The solar photovoltaic system and dataset

The data used in forecast models were collected from a solar PV installation. Data measurement covers the period from 1 November 2013 to 31 October 2014. The results of measurements of the intensity of solar irradiation, the power and energy of the photovoltaic system before and after the inverter in the period of 1 year were used to build prognostic models. The photovoltaic system from which the data were derived was located in Rzeszow (50°02' N 22°17' E). It consists of three monocrystalline modules with a total peak power of 330 [Wp], connected to a voltage inverter. Photovoltaic modules are mounted on a frame at a fixed angle of 30° to the horizon. Annual irradiation on inclined surface was 1284 [kWh/m^2/year]. The following measured data from the solar PV system were used: daily total irradiation, and daily total value of generated energy behind the inverter, converted into 1 [m^2] of the active surface of the photovoltaic panels.

The input data for the forecasting model were eight explanatory factors. Seven of them were weather factors taken from the meteorological station in the locality where the solar PV system is located. The eighth factor was the irradiation measured in the system. The following seven weather factors were chosen as the input data of the forecast models: average temperature, maximum temperature, visibility, average wind speed, precipitation, cloudiness, wind direction. The output of the model was the forecast of energy production.

3. One day-ahead energy forecasting models

Renewable energy in photovoltaic systems strongly depends on the seasons and weather conditions throughout the year [4, 7]. At different times of the year, the length of the day changes, such as the effect of the sun, the height of the sun on the horizon, cloudiness and temperature. Various strategies for constructing the one day-ahead forecasting models can be found in the literature. Mostly, the year-round models, which forecast the amount of energy production the one day-ahead for any day of the year on the basis of yearly learning data were

developed. Another way is to build seasonal or monthly the one day-ahead forecasting models.

Based on the analysis of the weather data and time periods of the year, we have decided to select several types of forecast models based on the length of period. Division of periods was due to the conditions of the irradiation and weather conditions in particular periods of the year.

The following the one day-ahead forecasting models were adopted:
- Year-round model,
- Seasonal models ('spring', 'summer', 'autumn', 'winter' models).

The purpose of this division was to check how the weather conditions and seasons affect the accuracy of energy production forecasts in the system.

The year-round model was based on all year data. Seasonal models were based on data from particular seasons (spring data: from 1 March to 31 May, summer data: from 1 June to 31 August, autumn data: from 1 September to 30 November, winter data: from 1 December to 28 February). Seasonal models were used to forecast production for a given season only. The differentiated approach will allow you to build more accurate models by taking appropriate weather conditions into account for each season.

The explanatory data were divided into two groups: teaching data (estimation) and test data (verification), respectively, for the periods. Teaching and testing data were scaled to a range of 0-1 by dividing them by the maximum values of the appropriate factors. In the estimation of the parameters of the year-round model used data from 280 days of the year. Verification of this model was conducted using selected data from 60 days of the year. In seasonal models, teaching data from 78 (or less) days of a given season were used to determine their parameters. To verification of these models were used test data from 14 days of the season. Test data did not participate in the estimation of model parameters.

4. Forecasting models

To build forecasting models, which will predict the amount of energy generated by the photovoltaic system, MLP neural networks and RBF radial bias neural networks were used.

A simple feed-forward two-layers MLP neural network is described in Figure 1.

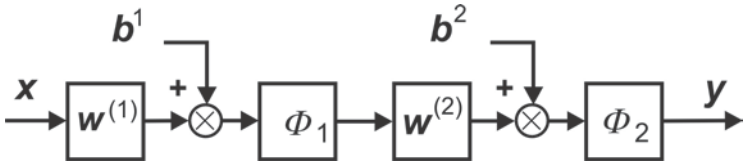


Figure. 1: Feed-forward MLP neural network.

The output of *j*-th neuron of neural network is given by equation (1):

$$y_j(t) = f_2\left[\sum_k^K w_{jk}^{(2)} \cdot f_1\left[\sum_{i=0}^I \left(w_{ki}^{(1)} x_i\,(t) + b_k^{(1)}\right)\right] + b_j^{(2)}\right] \tag{1}$$

where $\boldsymbol{w}^{(1)} \in R^{KxI}$, $\boldsymbol{w}^{(2)} \in R^{JxK}$, $\boldsymbol{b}^{(1)} \in R^K$, $\boldsymbol{b}^{(2)} \in R^J$, $\boldsymbol{x} \in R^I$, $\boldsymbol{y} \in R^J$, x_i is the *i*-th input of the neural network, y_j is the *j*-th output of a the neural network, $w_{ki}^{(1)}$ and $w_{jk}^{(2)}$ are the weights between neuros, $b_k^{(1)}$ and $b_j^{(2)}$ are the biases of neuros, *t* is a sample of input data.

The output layer uses the linear transfer function, which is denoted $f_2() \in \Phi_2$ in equation (1). The hidden layer typically processes using a nonlinear function, for example the sigmoidal transfer function, denoted as $f_1() \in \Phi_1$ in equation (1) described by the following equation:

$$f(x) = \frac{1}{1+e^{-x}} \tag{2}$$

For solar PV energy forecasting, the input should be the weather information, i.e. NWP data, and the output is the output energy of the solar PV system.

In order to determine the optimal parameters of the forecasting model, a quadratic quality index should be minimized:

$$\min Q(\boldsymbol{w}^{(1)}, \boldsymbol{w}^{(2)}, \boldsymbol{b}^{(1)}, \boldsymbol{b}^{(2)}) = \frac{1}{2}\sum_{t=1}^{T}\left(y^p(t) - y(t)\right)^2 \tag{3}$$

where: $y(t)$ is the generated energy in the PV system in t-th day, and $y^p(t)$ is predicted energy in t-th day by neural model, T is number of days.

The Levenberg-Marquardt algorithm [3] was used to minimize the quality index (3) and determine the parameters of the MLP neural networks as the forecasting models.

The models of MLP neural networks were made in two variants differing in the number of neurons in the hidden layer. The structure of the neural model were based on the analysis of the number of weather factors at the input of the neural network and based on preliminary simulations with different number of neurons in the hidden layer.

Forecast models as MLP neural networks have the following structure:

- Option 1: 8-7-1 (the structure: 8 input neurons, 7 neurons in the hidden layer, 1 neuron in the output layer);
- Option 2: 8-9-1 (the structure: 8 input neurons, 9 neurons in the hidden layer, 1 neuron in the output layer).

In MLP neural networks all neurons in hidden layers use the sigmoid activation function, and the neurons in the output layer use the linear activation function. In RBF networks, the number of neurons in the hidden layer was dependent on the set of learning data and accepted parameters of the network. The neurons in hidden layer use the Gaussian radial basis function.

Figure 2 shows the structure of MLP neural network as the one day-ahead forecasting model.

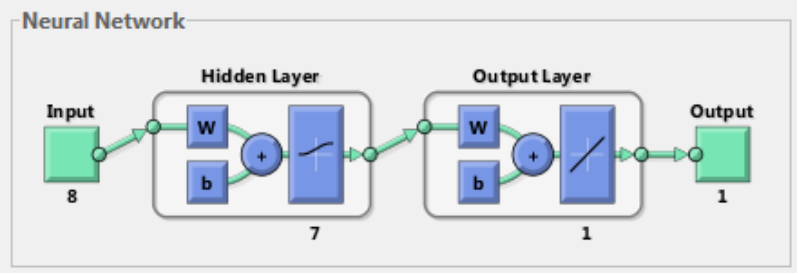


Figure. 2: The structure of neural network as a forecasting model (option 1, the net 8-7-1).

As a measure of the quality of the forecast of energy generation in the system, the *MAPE* (Mean Absolute Percentage Error) measure was used:

$$MAPE = \frac{1}{T}\sum_{i=1}^{T}\frac{\left|y_i^p - y_i\right|}{y_i} \cdot 100 \tag{4}$$

where: y_i - real measured energy in PV system in the *i*-th day, y_i^p - energy production forecast in the *i*-th day, *T* - number of days.

5. Forecasts of energy production in the photovoltaic system

For all types of models and their options the one day-ahead forecasts of energy production in the solar PV system were determined. In Table 1 are shown the errors forecasting the volume of energy production by the year-round model and seasonal models using MLP neural networks. The errors forecasting the volume of energy production by models using RBF neural networks are shown in Table 2. Forecast errors were estimated by the *MAPE* measure of accuracy, both for the estimation date (learning phase) and for the verification data (testing phase).

Table 1. Errors of the one day-ahead forecasts of energy production using the MLP neural networks as forecasting models.

Option	Model structure	Year-round model		Seasonal model 'spring'		Seasonal model 'summer'		Seasonal model 'autumn'		Seasonal model 'winter'	
		Estimation	Verification	Estimation	Verification	Estimation	Estimation	Estimation	Verification	Estimation	Verification
		MAPE [%]	*MAPE* [%]	*MAPE* [%]	*MAPE* [%]	*MAPE* [%]	*MAPE* [%]	*MAPE* [%]	*MAPE* [%]	*MAPE* [%]	*MAPE* [%]
1	8-7-1	3.195	3.283	2.187	2.092	1.544	2.077	3.282	3.219	3.103	3.041
2	8-9-1	3.051	3.334	2.405	2.505	2.184	2.145	2.809	2.902	3.536	3.565

Table 2. Errors of the one day-ahead forecasts of energy production using the RBF neural networks as forecasting models.

	Year-round model		Season model 'spring'		Season model 'summer'		Season model 'autumn'		Season model 'winter'	
	Estimation	Verification	Estimation	Verification	Estimation	Verification	Estimation	Verification	Estimation	Verification
Model structure	8-136-1		8-48-1		8-28-1		8-39-1		8-31-1	
MAPE [%]	4.326	3.674	3.132	3.198	2.190	3.186	3.228	3.241	3.132	3.198

Figure 3 shows the forecast of energy production and the real photovoltaic energy production for the test data (verification phase) by the seasonal 'summer' (option 1) model. The forecast covers 14 subsequent days throughout that season.

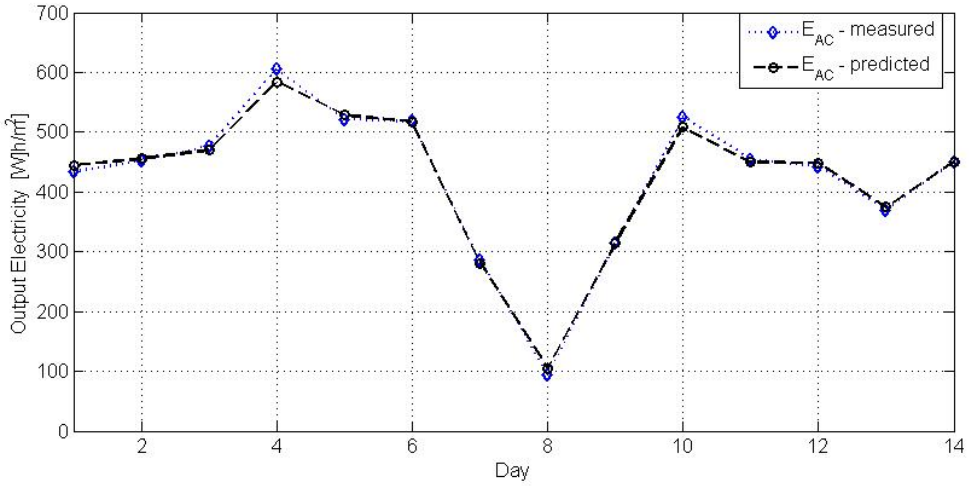


Figure. 3: The real (measured) energy and the one day-ahead forecast energy in the solar PV system (done by MLP neural network model (option 1, the net structure: 8-7-1)).

6. Remarks and conclusion

This paper presents the results of the author's research on the energy generation by solar photovoltaic system for the methods using MLP neural networks and RBF neural networks for different time periods. The forecasts of the energy production were based on eight explanatory factors (daily irradiation and seven weather factors).

For the one day-ahead forecasting, the year-round model and four seasonal models were built. The construction of seasonal models has a strong justification for the amount of irradiation and weather in Poland at any time of the year.

Almost all seasonal models provide more accurate forecasts than the year-round model. Forecasting models using MLP neural networks generate more accurate forecasts than model using RBF neural networks. However, all results of forecasted energy are reasonable for practical application. Models provide the forecast of energy production by the solar photovoltaic system with satisfactory accuracy. The forecast error for the year-round MLP model was around 3.33%. The errors of forecasting in the majority seasonal models are less than 3.30%. Only in the ‘winter’ model, option 2, the errors of forecasting are about 3.56%.

Similarly, the forecast error for the year-round RBF model was around 3.67%. The forecast errors for all seasonal RBF models were less than 3.20%. The most accurate forecasts were obtained in the seasonal 'summer' model using MLP neural networks (option 1).

References

1. Ashraf I., Chandra A.: *Artificial neural network based models for forecasting electricity generation of grid connected solar PV power plant*, International Journal of Global Energy Issues, Vol. 21, n 1-2, 2004, pp. 119-123.
2. Ciechulski T., Osowski S.: *Study on the quality of electrical power prediction using SVM, RBF and MPL neural networks*, Przegląd Elektrotechniczny, 2014, No 8, pp. 148-151 (in Polish).
3. Haykin, S.: *Neural Networks: A Comprehensive Foundation*, 2nd edition, Prentice-Hall, Englewood Cliffs, New York, 1999.

4. Huang Y. et al.: *Comparative study of power forecasting methods for PV stations*, International Conference on Power System Technology, Hangzhou, October, 2010.

5. Jie Shi , et al.: *Forecasting power output of photovoltaic system based on weather classification and support vector machine*, Industry Applications Society Annual Meeting (IAS) 2011 IEEE, Orlando, October, 2011.

6. Martin L., Zarzalejo L.F., Polo J., Navarro A., Marchante R., Cony M.: *Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning*, Solar Energy, Elsevier, Vol. 84, 2010, pp. 1772-1781.

7. Piotrowski P.: *Analysis of artificial neural network applications for short-term power forecasting and production of electricity in photovoltaic systems*, Przegląd Elektrotechniczny, ISSN 0033-2097, R. 91, No 8, 2015, pp.162-165 (in Polish).

8. Yona A., et al.: *Application of neural network to 24-hour-ahead generating power forecasting for PV system*, Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, IEEE, Pittsburgh PA, July, 2008.

Andrzej SEUL[1], Krzysztof OKARMA[2]

Texture Classification for Autonomous Cleaning Robots

In the paper the experimental results of texture classification based on Haralick features for autonomous cleaning robots are presented. Since the classification based on the application of four classical features has not led to satisfactory results, some additional features based on their variance have been considered. For verification purposes a fragment of popular ALOT texture database has been used representing the most typical textures for cleaning robots. Considering the computational cost, the experiments have also been made for decreased number of gray levels applying various color to grayscale conversion methods.

1. Introduction

The development of machine vision applications in robotics becomes more and more popular as the information amount derived by cameras allows more accurate visual classification and recognition of various objects as well as materials and textures. One of such popular areas of applications of texture analysis may be the analysis of images captured by drones. However, the classification of surfaces may be important also for mobile wheel or walking robots e.g. for autonomous cleaning robots. Knowing what kind of surface is ahead of the cleaning robot, it can apply adequate action. It could choose specific detergent, use specialized cleaning brush or set surface-specific moving velocity, or even decide if this surface is to be worked on, or skipped. The application of texture analysis for the latter ones is considered in this paper, although similar approach may also be successfully applied for some other purposes related e.g. to self-navigation of autonomous mobile robots.

2. Texture database

For the purpose of this article, the image database consisting of 700 photographs of 7 different materials has been prepared. Those images have been obtained from Amsterdam Library Of Textures (ALOT), which is a color image collection of 250 textures recorded in specific laboratory conditions. During photographing some parameters such as viewing angle, illumination angle and illumination color were varying according to the provided technical specification [1]. This resulted in 100 images for each material. The complete ALOT texture database consists of wide range of different materials - some are more of an "object" nature, and some are more of "pattern" nature. Seven materials of "pattern" nature have been chosen as the most similar to working area of cleaning mobile robot (see Figure 1). In addition, each texture image has been resized to 320×320 pixels and saved as 24-bit RGB bitmap.

[1]Ph.D. student, West Pomeranian University of Technology, Szczecin - Faculty of Electrical Engineering, ul. Sikorskiego 37, 70-313 Szczecin, POLAND, e-mail: andrzej.seul@gmail.com

[2]Assoc. Prof., West Pomeranian University of Technology, Szczecin - Faculty of Electrical Engineering, ul. Sikorskiego 37, 70-313 Szczecin, POLAND, tel. +48 91 449 53 13, e-mail: okarma@zut.edu.pl

Figure 1: Texture database used in experiments - from left: (57) cork, (74) terry cloth, (98) marble, (184) carpet, (189) hardboard, (205) carpet no.2 and (206) wood plank.

3. Standard (basic) approach

There are three basic approaches in texture analysis: statistical, geometrical and model based methods [3].

Model based methods aim to describe a texture by a mathematical model. Their nature qualifies them to be useful mostly for analyzing artificial textures. Geometric methods utilize the principle that a texture is composed of small geometrical elements trying to identify those repetitive elements and analyze them separately, either with statistical approach, or geometric methods. Statistical methods compute texture features based on spatial distribution of gray levels over the image. Because of "naturalness" of the texture database and lack of obvious geometrical patterns in images, the statistical approach has been chosen in our experiments. One of the most popular approaches in texture classification is to compute the Gray Level Co-occurrence Matrix (GLCM) of an original image, then calculate Haralick features from it and use those parameters for classification purposes. One of the motivations of its choice for our experiments is the possibility of efficient implementation in the FPGA based systems allowing fast computations with low latency as described in the paper [7].

The GLCM illustrates the spatial relations between the pixels in the specified neighborhood. For the given offset $(\Delta x, \Delta y)$ its values can be determined as:

$$C(i,j) = \sum_{p=1}^{P} \sum_{q=1}^{Q} \begin{cases} 1 & \text{if } A(p,q) = i \text{ and } A(p+\Delta x, q+\Delta y) = j \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $P = M - \Delta x$ and $Q = N - \Delta y$.

Each element of the GLCM ($C(i,j)$) represents the number of occurrences of pixels having the given luminance level i in the specified neighborhood, defined by the offset $(\Delta x, \Delta y)$, with pixels of the other chosen luminance level j. As for each image four directions can be considered: horizontal, vertical as well as both diagonals, four different matrices can be calculated for each specified distance (not necessarily equal to one pixel).

The GLCM matrix used for the calculation of Haralick features is typically symmetrical so the luminance level i of a pixel above j is considered equally as i below j for the vertical GLCM (and similarly left and right neighborhood for the horizontal matrix). It leads to even or zero values on its diagonal. Such a matrix can be additionally normalized ensuring the sum of the GLCM values being equal to 0 and therefore a reliable comparison of textures (represented by the GLCM features) can be made independently on the resolution of the original images.

In the original paper [4] 14 different features have been proposed but four of them have become the most relevant. A more detailed analysis of their influence on texture classification accuracy is presented e.g. in the paper [2]. Therefore four Haralick features shown in Table 1 shall be used as parameters for classification purposes.

Table 1: Haralick features used in experiments.

Feature	Formula
Contrast	$\sum_{i,j} \lvert i-j \rvert^2 C(i,j)$
Correlation	$\sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)C(i,j)}{\sigma_j \sigma_i}$
Energy	$\sum_{i,j} C(i,j)^2$
Homogeneity	$\sum_{i,j} \frac{C(i,j)}{1+\lvert i-j \rvert}$

The quantities in Table 1 calculated for the normalized symmetrical GLCM are: $C(i,j)$ - value of the normalized symmetrical GLCM at position (i,j); i,j - gray levels of the neighboring pixels; μ - average value of p, σ - standard deviation of p.

4. Proposed extended approach

It is possible to extend the described method to acquire more data for classification purposes. Although Haralick features were originally proposed to be calculated for the whole texture image, one can divide a texture into arbitrary chosen smaller parts and calculate those features for each part separately. Then, one can use a simple statistical measure such as variance to obtain a new set of features of the texture. For given $N \times M$ image first one has to cut it evenly across r rows and c columns into $r+c$ pieces, each of $N/r \times M/c$ resolution (see Figure 2). Then, for each of $I(r,c)$ images, calculate Haralick feature $F(r,c)$ value. Finally, one can compute the variance var[F] of such data. This method will be later referred as the VHF (Variance of Haralick features).

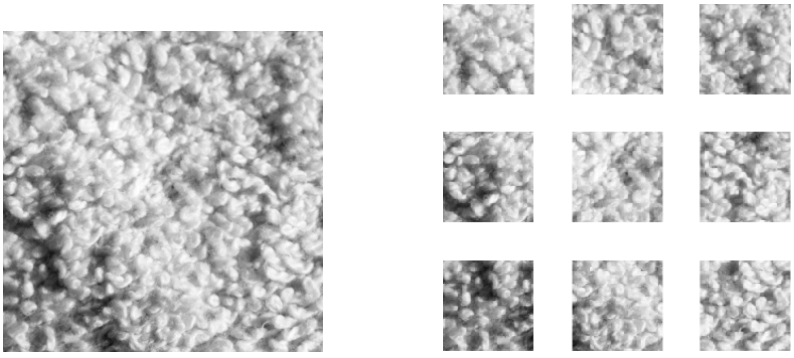


Figure 2: Illustration of the texture cutting process (left - before cutting, right - after cutting into 9 images).

For each Haralick feature the additional variance parameter shall be calculated resulting in four additional parameters used for classification. Texture classification shall be performed using this extended method, and quality of classification shall be compared to standard approach (based on four Haralick features).

5. Color to grayscale conversion

One of the purposes of this work is to determine whether using extra information about colors could lead to better classification compared to standard (BT-601 or BT-709) grayscale images, hence the need to use texture database consisting of color images. Because of that, application of any of the described methods requires previous conversion of those images from 24-bit RGB color to 8-bit grayscale images. There is no unique method to do so - one can choose one of the many conversion algorithms. Choosing a proper conversion method is very important - appropriate method will provide good quality data for later processing.

For purposes of this paper, six different methods will be used originating from various color models: RGB, HSV, HSL, HSI and CIE LAB. First two methods use simple linear combination of red, green and blue channels in RGB model to compute grayscale value of this pixel according to ITU standards for SDTV [5] and HDTV [6]. Next three operate on cylindrical-coordinate representation of color model. Grayscale image can be obtained by separating "lightness-oriented" coordinates from these models (respectively V, L, or I channel). Calculating those values is pretty simple - in every model it is just a combination of R,G,B values. The last method used on experiments is based on CIE LAB color space, where L coordinate represents lightness of a color. It is the only one method from those six that require two-step conversion. First, one has to convert RGB image to CIE XYZ color space, and then compute L coordinate. Conversion formulas for all these six methods can be found in Table 2 where Y_0 for CIE LAB space denotes the Y value of the specified white point. Texture classification shall be performed using each of those conversion techniques to see the influence of conversion technique on quality of classification.

Furthermore, the number of grayscale levels will be reduced to 7-bit (128 levels), 6-bit (64 levels) and 5-bit (32 levels), and texture classification shall be performed using those reduced grayscale images to observe how the decrease of the number of gray levels influences the classification quality.

Table 2: Color to grayscale conversion methods.

Method	Grayscale value
ITU BT.601-7	$g = 0.2989 \cdot R + 0.5870 \cdot G + 0.1140 \cdot B$
ITU BT.709-6	$g = 0.2126 \cdot R + 0.7152 \cdot G + 0.0722 \cdot B$
HSV	$g = max(R,G,B)$
HSL	$g = 0.5 \cdot (max(R,G,B) + min(R,G,B))$
HSI	$g = (R+G+B)/3$
CIE LAB	1) $RGB \rightarrow XYZ$; 2) $g = 116 \cdot \sqrt[3]{Y/Y_0} - 16$ for $Y/Y_0 > 0.008856$ or $g = 7.787 \cdot Y/Y_0 + 16/116$ for $Y/Y_0 \leq 0.008856$

6. Classification

Three different classification methods have been used to rule out false-positive (or false-negative) results, namely: linear discriminant analysis (LDA), naive Bayes classifier, and decision tree learning.

For linear discriminant analysis the pseudo-inversion has been used to invert the covariance matrix, which is the same for every class. Naive Bayes classifier has been trained using basic Gaussian distribution model. In decision tree learning, the maximum number of splits has been

limited to 15 to avoid tree overgrowing that could result in producing too many decisions, which could lead to false-positives matches of control data.

For comparison purposes, a quality criterion for classifiers has to be introduced. In this paper the resubstitution loss (RL) and F-measure will be used to compare classification results using different classification methods and different set of textural features. The RL value is adequate to percentage of misclassified observations from training data set (reversely to accuracy which is interpreted as the ratio of all correct classifications - true positives and true negatives - to all samples). F-measure is the harmonic mean of precision and recall, which are common measures used in binary classification tasks i.e. precision is calculated as true positives to all positives ratio and recall as the ratio of true positives to sum of true positives and false negatives.

7. Results

Texture database was split into two parts - training data and validation data. Firstly, classifiers were trained using set of training data of 7 categories (textures). Then, using validation data as a new observation, the quality of classification was evaluated based on whether classifier assigned new observation to correct category.

In order to assess the classification accuracy the resubstitution loss (RL) error and F-measure for different sets of classification parameters were calculated. Firstly, we have examined how ”good” is each of four Haralick features and each of four VHF’s to be used as a single parameter for classification purposes. We have calculated those parameters (as described in sections 3. and 4.) for grayscale images (used the same method to convert all images), calculated the RL errors presented in Tables 3 and 5 and then calculated F-measures presented in Tables 4 and 6. As can be clearly seen, using only one parameter for the classification is not very effective. Also, we can see that ”tree” classifier seems to be working best of those tree. It also looks like Haralick features work better in one-parameter classification than VHFs.

Table 3: RL error using only one Haralick feature.

	contrast	correlation	energy	homogenity
LDA	0.636	0.629	0.681	0.629
Bayes	0.627	0.619	0.677	0.604
tree	0.560	0.511	0.569	0.544

Table 4: F-measure using only one Haralick feature.

	contrast	correlation	energy	homogenity
LDA	0.1426	0.1473	0.1183	0.1473
Bayes	0.1476	0.1526	0.1222	0.1590
tree	0.1830	0.2149	0.1811	0.1945

Table 5: RL error using only one VHF.

	var. contrast	var. correlation	var. energy	var. homogenity
LDA	0.764	0.684	0.753	0.729
Bayes	0.753	0.707	0.707	0.747
tree	0.599	0.617	0.564	0.601

Table 6: F-measure using only one VHF.

	var. contrast	var. correlation	var. energy	var. homogenity
LDA	0.0823	0.1179	0.0867	0.0989
Bayes	0.0867	0.1081	0.1081	0.0904
tree	0.1637	0.1529	0.1822	0.1597

After that, we have examined the impact of color to grayscale conversion method used to obtain grayscale images. We have used six different methods described in section 5. and calculated Haralick features as well as the VHF features for those images, trained classifiers using either only Haralick features or those features together with the VHF features, and finally calculated RL errors shown in Tables 7 and 9 and F-measures shown in Tables 8 and 10. We can observe some kinds of correlation between the chosen method, chosen classifier and the quality of classification. Some methods work better with some classifiers, so to achieve the best classification, one should match color to grayscale conversion method with the type of the classification method. In general, we can see that the HSL method seems to be working best for most cases although there are few exceptions. Nevertheless, this method has been applied also in further experiments.

Table 7: RL error for different conversion methods (Haralick features only).

	BT-601	BT-709	HSV	HSL	HSI	LAB
LDA	0.417	0.420	0.430	0.390	0.391	0.414
Bayes	0.420	0.423	0.400	0.383	0.409	0.417
tree	0.291	0.304	0.280	0.267	0.321	0.307

Table 8: F-measure for different conversion methods (Haralick features only).

	BT-601	BT-709	HSV	HSL	HSI	LAB
LDA	0.2878	0.2871	0.2788	0.3089	0.3081	0.2892
Bayes	0.2871	0.2816	0.3000	0.3163	0.2956	0.2878
tree	0.4104	0.3966	0.4235	0.4418	0.3767	0.3944

Table 9: RL for different conversion methods (Haralick features + VHF).

	BT-601	BT-709	HSV	HSL	HSI	LAB
LDA	0.414	0.409	0.417	0.387	0.380	0.404
Bayes	0.381	0.379	0.344	0.350	0.361	0.384
tree	0.189	0.196	0.220	0.180	0.196	0.200

Table 10: F-measure for different conversion methods (Haralick features + VHF).

	BT-601	BT-709	HSV	HSL	HSI	LAB
LDA	0.2892	0.2956	0.2878	0.3139	0.3179	0.2978
Bayes	0.3171	0.3223	0.3539	0.3467	0.3360	0.3155
tree	0.5541	0.5418	0.5032	0.5655	0.5418	0.5333

Finally we have checked how the decrease of the number of grayscale levels influences the quality of classification. We used four different bit depths (256, 128, 64 and 32 levels) and calculated Haralick features and VHF features for those images. Then we have trained the classifiers using either only Haralick features or those features + VHF features, calculated the RL errors presented in Tables 11 and 13 and F-measures presented in Tables 12 and 14 . The classification quality is roughly the same for all those methods. Because there is no significant difference in RL errors, the computational complexity of our algorithm can be reduced by decreasing number of grayscale levels to 32 without a strong influence on the classification results.

Table 11: RL error for different grayscale levels (Haralick features).

	256	128	64	32
LDA	0.417	0.414	0.407	0.407
Bayes	0.420	0.414	0.411	0.413
tree	0.291	0.300	0.293	0.290

Table 12: F-measure for different grayscale levels (Haralick features).

	256	128	64	32
LDA	0.2878	0.2892	0.2963	0.2963
Bayes	0.2871	0.2892	0.2906	0.2899
tree	0.4104	0.4000	0.4092	0.4116

Table 13: RL error for different grayscale levels (Haralick features + VHF).

	256	128	64	32
LDA	0.414	0.393	0.386	0.369
Bayes	0.381	0.379	0.367	0.370
tree	0.189	0.193	0.199	0.171

Table 14: F-measure for different grayscale levels (Haralick features + VHF).

	256	128	64	32
LDA	0.2892	0.3073	0.3147	0.3316
Bayes	0.3171	0.3223	0.3325	0.3273
tree	0.5541	0.5455	0.5382	0.5804

8. Conclusions

Analyzing the results one can conclude that introducing the VHF parameters in addition to the standard Haralick features improves the quality of classification for the proposed texture database in all considered cases. It can also be observed that the reduction of the number of grayscale levels from 256 to 32 does not decrease the quality of classification, so one could safely use the smaller number of levels to reduce computational complexity of the algorithm. Finally, color to grayscale conversion method should be chosen in pair with classification method. Although the HSL method not always leads to the best results, in most of the cases it allows a proper classification of textures leading to promising results. One of the directions of our future work will be the implementation of the proposed method using the experimental robot in order to verify the lighting conditions in real environment, considering the highly demanding nature of the ALOT database used mainly for the verification of general purpose texture classification algorithms. Another considered direction of research is the combination of the GLCM features with some other texture analysis methods such as e.g. Structural Texture Similarity [8].

References

1. *Amsterdam Library of Textures*, `http://aloi.science.uva.nl/public_alot/`.
2. Beliakov G., James S., Troiano L.: *Texture recognition by using GLCM and various aggregation functions*, Proc. 2008 IEEE Int. Conf. on Fuzzy Systems (IEEE World Congress on Computational Intelligence), pp. 1472-1476, Hong Kong, 2008.
3. Chen C.H., Pau L.F, Wang P.S.: *Handbook of Pattern Recognition and Computer Vision*, *World Scientific Publishing Co. Pte. Ltd.*, 1993.
4. Haralick R.M., Shanmugam K., Dinstein I.: *Textural Feature for Image Classification*, IEEE Transactions on Systems, Man, and Cybernetics, vol. 3, pp. 610–621, 1973.
5. ITU Recommendation BT.601-7: *Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios*, 2011.
6. ITU Recommendation BT.709-6: *Parameter values for the HDTV standards for production and international programme exchange*, 2015.
7. Komorkiewicz M., Gorgoń M.: *Foreground object features extraction with GLCM texture descriptor in FPGA*, Proc. 2013 Conf. on Design and Architectures for Signal and Image Processing (DASIP), pp. 157–164, Cagliari, Italy, Oct. 2013.
8. Žujović, J., Pappas, T.N., Neuhoff D.L.: *Structural texture similarity metrics for image analysis and retrieval*, IEEE Transactions on Image Processing, vol. 22, no. 7, pp. 2545–2558, July 2013.

Dariusz SOBCZYŃSKI [1]

PSIM Model of Single-Phase Inverter for IGBT Transistor Power Loss Analysis

Renewable energy sources become more important day by day. There are many different inverter topologies without transformers for green energy sources. This paper presents a comparison of semiconductor losses of H5 inverter and NPC inverter used in renewable energy sources. Concept of the power losses calculation was described. Power losses have been calculated by thermal module provided by PSIM software. Based on the largest and smallest power losses, in order to verify validity of the presented converter topology the comparative analysis of individual converters has been made.

1. Introduction

Recently many of energy economies become independent of traditional energy sources like: coal, oil and natural gas. Fossil fuels are running out, the renewable energy resources will never run out. Green energy sources in most cases require transformation into electricity, so small scale converters become the focus of researchers [8]. There are topologies of converter with and without transformer. Even though transformer has an important role in safety and leakage current eliminating, it has disadvantages such as the increased the cost and the power losses, so the efficiency of the system is reduced [3][4][14][15]. Hence, transformer-less inverters are nowadays one of the most popular power electronic devices used to connect green energy sources to the electrical grid due to their excellent performance in the energy conversion process: they are smaller, lighter, highly cost effective, and more efficient [5][6][9][11]. Nevertheless, the lack of transformers leads to worse safety issues. Leakage currents can be dangerous for human, as well as for the converter and whole energy system. In order to solve the problem, many of new power electronics converters have been proposed [13]. The important property is the reduction of current harmonics contents, so control mechanisms have been proposed to regulate the inverter output current that is injected into the utility grid [1][2].

Many of the converter topologies in literature show the efficiency in the range of 96%-98%, but the power losses distribution of individual semiconductor components in the converter are not described. The new idea of this paper is intended to show the semiconductor devices loss distribution of selected inverters, that is important during the design process of the converter. The simulation results have been obtained by using PSIM software [10].

2. Calculation of power electronics semiconductor losses in PSIM

The methodology of the power losses calculation of the discussed converter was explained in this section. Power losses have been calculated by thermal module provided by PSIM software. It is possible to calculate semiconductor losses quickly based on power electronics switch datasheets using PSIM software.

[1] Ph.D., Rzeszow University of Technology, Powstańców Warszawy 12, 35-959 Rzeszów, POLAND, tel. +48 17 865 1974, e-mail: dsobczyn@prz.edu.pl

2.1. IGBT

The switch model consists of transistor and antiparallel diode. Power losses of IGBT transistor and antiparallel diode are divided into: blocking (leakage) losses, static losses also called conduction losses, and the transient losses known as switching losses [12]. In general in comparison to other IGBT's and anti-parallel diode losses, leakage losses are low and can be neglected. Therefore, the switch total power losses can be described by the formula:

$$P_{STot} = P_{TTot} + P_{ADTot} \approx P_{Tcon} + P_{Tsw} + P_{ADcon} + P_{ADsw} \tag{1}$$

Where, P_{STot} are the total switch losses, P_{TTot} are the transistor total losses, P_{ADTot} are the antiparallel diode total losses, P_{Tcon} and P_{ADcon} are sequentially transistor and antiparallel diode conduction losses. P_{Tsw} and P_{ADsw} are sequentially transistor and antiparallel diode switching losses.

PSIM calculation of IGBT conduction losses mainly depends on required average I_{cAV} and the RMS values I_{cRMS} of a load current, the transistor voltage u_{ce0} and a collector-emitter on-state resistance r_C.

$$P_{Tcon} = u_{ce0} I_{cAV} + r_c I_{cRMS}^2 \tag{2}$$

If the average diode current is I_{ADAV}, and the RMS diode current is I_{ADRMS}, the average diode conduction losses are:

$$P_{ADcon} = u_{AD0} I_{ADAV} + r_{AD} I_{ADRMS}^2 \tag{3}$$

The ratio of switching energies (the turn-on energy E_{Ton} and the turn-off energy E_{Toff}) and the switching frequency f_{sw} determines power losses during switching periods. The switching losses of the transistor are calculated by formula:

$$P_{Tsw} = (E_{Ton} + E_{Toff}) f_{sw} \tag{4}$$

The loss calculation for device integrated with the anti-parallel diode is the same as was presented above, however due to the small amount of turn-on energy, the following relationship is assumed:

$$P_{ADsw} \approx E_{ADoff} f_{sw} \tag{5}$$

2.2. Power diode

Diode conduction and switching losses are calculated following the same approach described previously for the switch consisting of IGBTs and anti-parallel diode. The power diode conduction losses depend on:

$$P_{Dcon} = u_{D0} I_{DAV} + r_D I_{DRMS}^2 \tag{6}$$

As mentioned in the previous subsection, diode turn on energy is neglected, so switching losses are calculated as:

$$P_{Dsw} \approx E_{Doff} f_{sw} \tag{7}$$

3. Transformer-less inverter topologies

The H-bridge converter topology is well known and widely used in both DC-DC and DC-AC inverters. This topology is also known as H4 circuit [7]. There are two main modulation strategies commonly used for AC output voltage generation. Bipolar and unipolar modulation strategy differs in current ripple, leakage current, electromagnetic interference

troubleshooting, passive elements losses and efficiency. When full H-bridge transformer-less converter is used, a common-mode voltage appears as an undesirable result. A varying common-mode voltage can generate a leakage current, flowing by circuit consisting of the stray capacitor between DC source and the ground.

This section presents a modified converter topologies that have been selected based on their main features, among which is the low number of semiconductors, which by nature can reduce the power losses and costs. First, the one phase H5 topology, which consists of only 5 semiconductors was analysed. Second topology considers circuit of the NPC converter with 6 semiconductor devices.

3.1. H5 inverter topology

H5 inverter was made by an extra switch (IGBT5) between the DC circuit and full H-bridge (H4) addition. This topology is shown in Fig. 1.

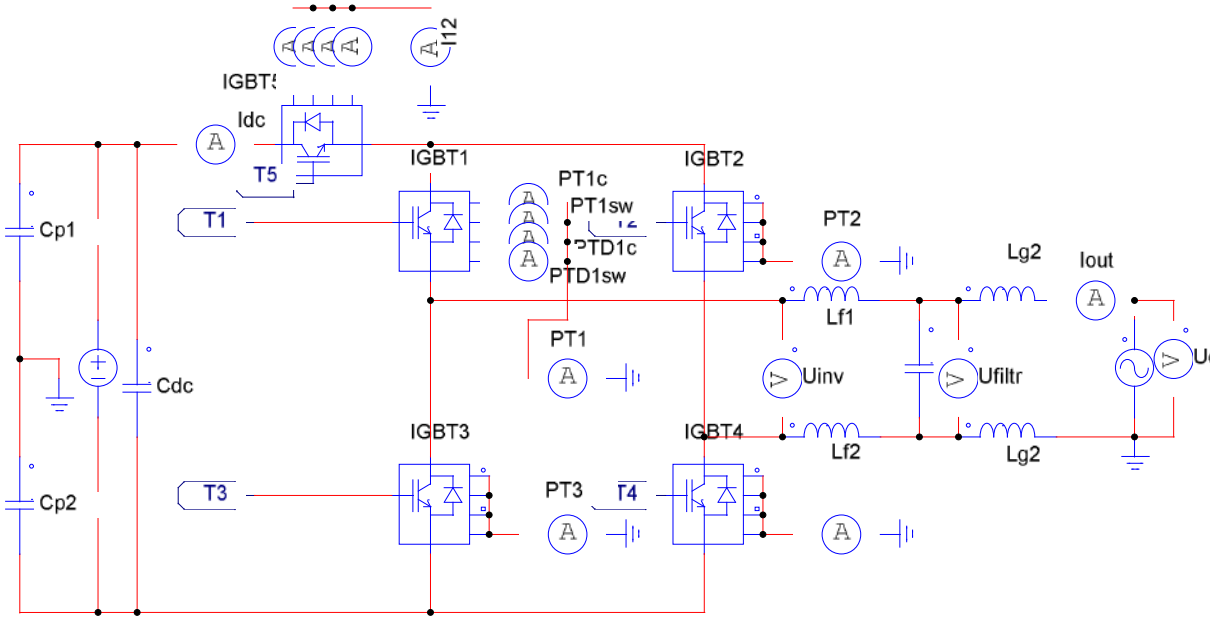


Figure. 1: PSIM model of H5 inverter topology.

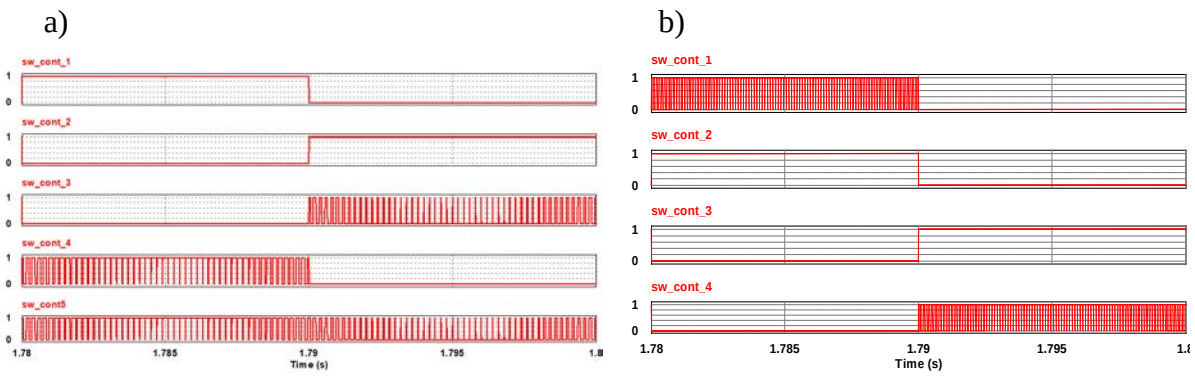


Figure. 2: Control signals for a) H5 inverter, b) NPC inverter.

The H4 converter's switches (IGBT1-IGBT4) are not complementarily commutated, two switches are high frequency and two are low frequency operated. When the switches IGBT3

and IGBT4 are operated with high switching frequency, the IGBT1 and IGBT2 are turned on complementarily. Than current flows through IGBT1, IGBT4, and IGBT5. The zero voltage vectors are achieved when IGBT4, and IGBT5 are turned off, then the current paths flow through IGBT1 and IGBT3. Because the current path flows through three semiconductor elements, H5 converter has higher conduction losses then H4 topology. Control strategy for H5 inverter has been presented in Fig. 2a.

3.2. NPC inverter topology

NPC inverter was shown in Fig. 3. That topology consists of 6 semiconductors (four transistors and two diodes). Because of the power transmission to the grid, the DC link voltage must always be higher than the output voltage amplitude. The NPC's transistor switching control signals have been presented in Fig. 2b, and can be described as follows. During the first half of the cycle, IGBT2 remains turned-on, transistor IGBT1 is operating at switching frequency, then IGBT3 and IGBT4 are turned-off. So, when IGBT1 is turned-on, current increases and its path flows through IGBT1 and IGBT2. While IGBT1 is turned off, current decreases flows through DD1 and IGBT2. During the second cycle, IGBT3 remains turned-on while IGBT4 changes at the switching frequency and IGBT1 and IGBT2 are off. When IGBT4 is turned-on current is increasing, flows through IGBT3 and IGBT4. When IGBT4 was turn-off, current is decreasing, flows through IGBT3 and DD2.

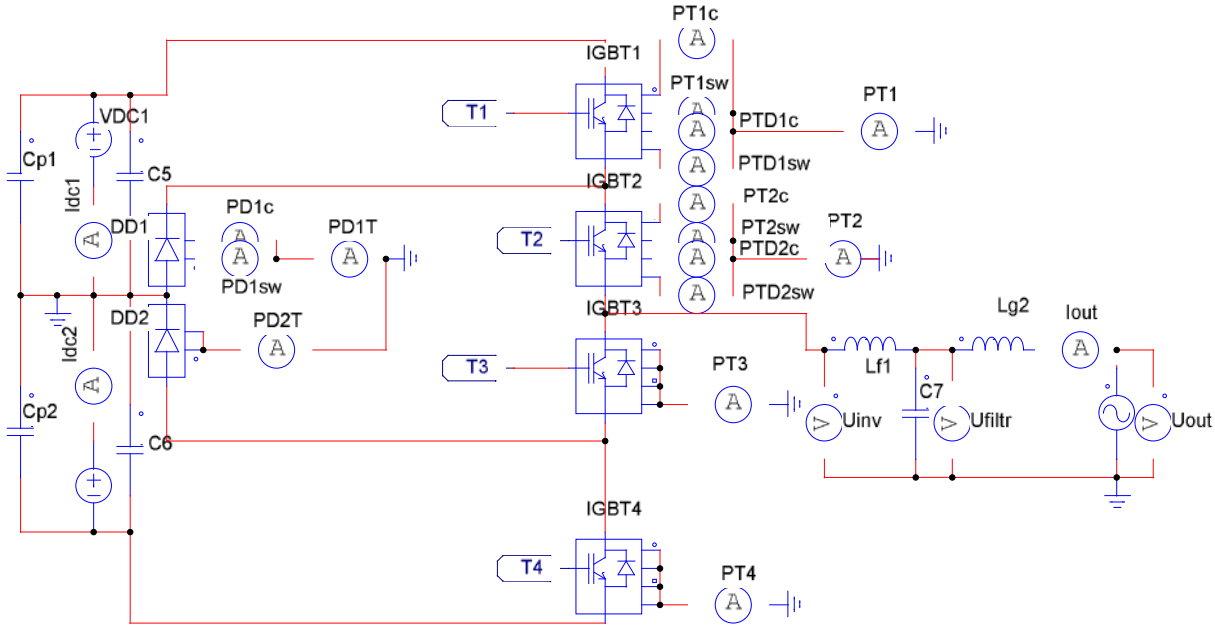


Figure. 3: PSIM model of NPC inverter topology.

4. Power loss analysis

Use of the thermal model from the PSIM library allows to calculate the power loss of semiconductor devices. Thermal models of semiconductor devices include both static and dynamic properties such as conductor voltage drop or dynamic resistance. Computer simulation was performed in the PSIM program. The IGBT MH1000HA-24H transistor thermal model was used for the study. Simulation tests were performed for the following parameters. The DC circuit voltage was set at 350V for both H5 and NPC inverters. However, the NPC adopts the two times higher inverter voltage, compared to those of the previous

topology. The converters load is the grid with L_g=1.8mH, the voltage U_{max}=325V and the frequency f=50Hz.

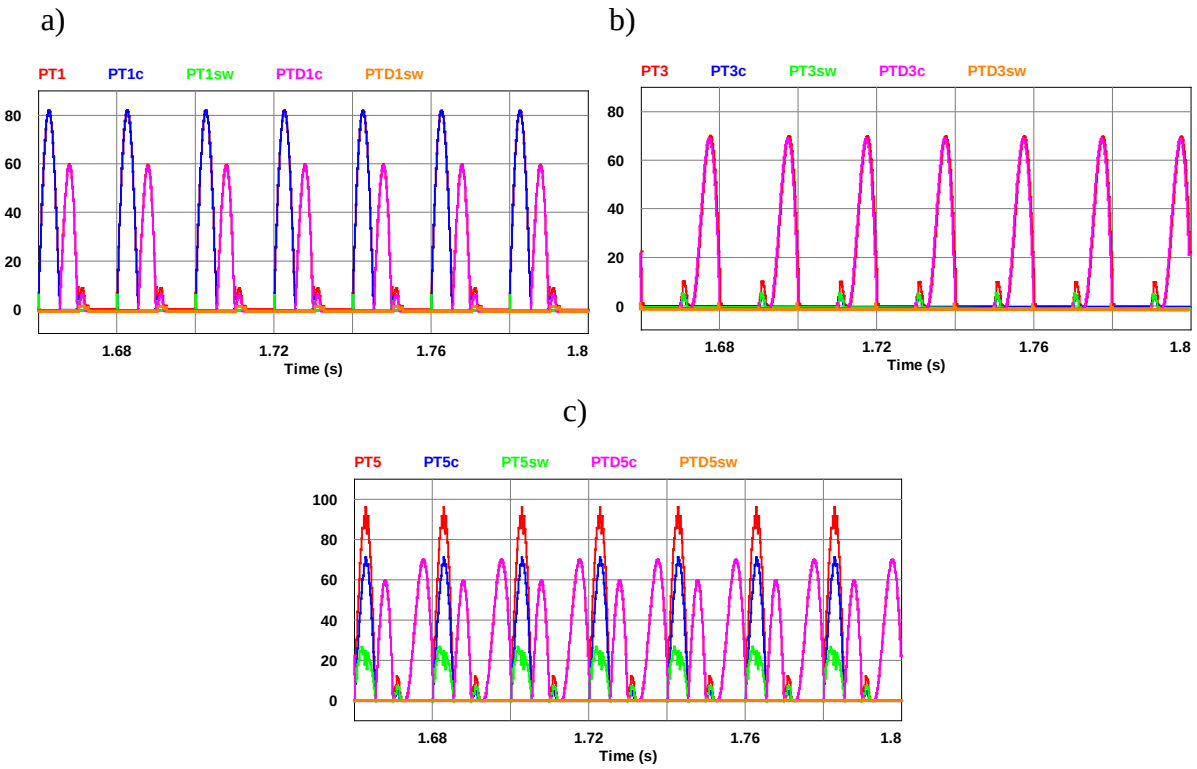


Figure. 4: H5 inverter with - power losses distribution of switches a) IGBT1, b) IGBT3 c) IGBT5, where: PT1, PT2, PT5 are respectively total losses of switches; PT1c, PT3c, PT5c - conduction losses of transistors number 1, 3 and 5; PT1sw, PT3sw, PT5sw - switching losses of transistors number 1, 3 and 5; PTD1c, PTD3c, PTD5c - conduction losses of anti-parallel diode 1, 3 and 5; PTD1sw, PTD3sw, PTD5sw - switching losses of anti-parallel diode 1, 3 and 5.

Simulations have been made for several instances - calculation of switching, conduction and total losses of all semiconductor components (transistors and diodes) in both H5 and NPC inverter's topologies. The Fig. 4a to 4c present simulation results for H5 inverter. Simulation of switching and conduction losses for IGBT1, IGBT3 and IGBT5 was done. As shown in Figure 4, the high conduction losses were recorded for transistor IGBT1 because of its square control signal (Fig.2a). The IGBT3 has a higher switching frequency, so and accordingly higher switching losses occur. Due to the longest conduction times and high switching frequencies, the highest total power loss was recorded on IGBT5. For each semiconductor switch, power losses in anti-parallel diodes are significant, but only in conduction state. Switching parts of diodes' power losses are neglected.

The semiconductor devices losses of NPC converter were shown in Fig. 5. Because of the high switching frequency, the higher power losses are noted in switches IGBT1. Switch IGBT2 has less power losses because of low switching frequency. The smallest losses characterized diode DD1 due to losses associated with conduction state only. Total losses of all semiconductor switches in H5 and NPC converter are presented in Fig. 6. The figure shows that the NPC is characterized by lower losses than the inverter H5. It can be observed that where the frequency of semiconductor devices is big, the switching connected losses

dominate. Transistor and diode conduction losses are comparable but the switching losses of power diodes are practically close to zero. The waveforms shown in Fig.4-6 illustrate in a good way uneven load on each of semiconductor element.

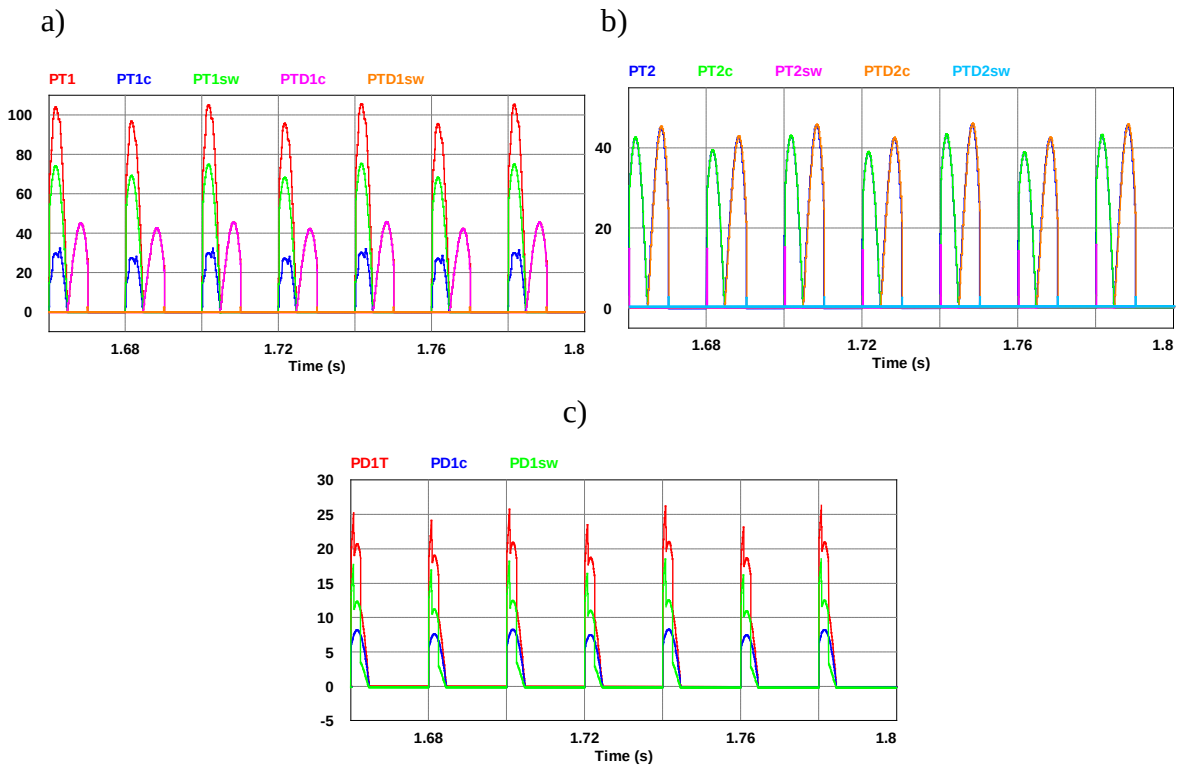


Figure. 5: NPC inverter with - Power Losses distribution of switches a) IGBT1, b) IGBT2 c) diode DD1, where: PT1, PT2, PD1T respectively total losses of switches; PT1c, PT2c - conduction losses of transistors 1 and 2; PT1sw, PT2sw - switching losses of transistors 1 and 2; PTD1c, PTD2c —conduction losses of anti-parallel diode 1 and 2; PTD1sw, PTD2sw - switching losses of anti-parallel diode 1 and 2, PD1c - switching losses of diode DD1, PD1sw - conduction losses of diode DD1

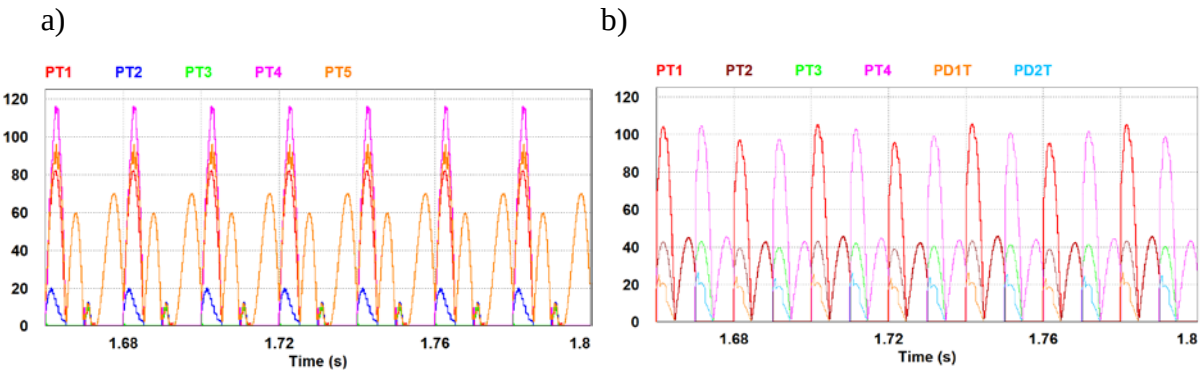


Figure. 6: a) H5 inverter with power losses distribution of switches IGBT1- IGBT5, where: PT1, PT2, PT3, PT4, PT5 denote a total losses of listed switches, b) NPC inverter with power loss distribution of switches IGBT1 – IGBT4 and diodes PD1T, PD2T where: PT1, PT2, PT3, PT4, PTD1, PTD2 denote a total losses of listed switches.

The power losses have been illustrated by the instantaneous power with high maximum value, but the mean values are lower. Therefore, in the following in Fig.7a and Fig.7b, the average values of the instantaneous power of the selected semiconductors were presented for the discussed topologies of converters. The simulation results for all semiconductor components occurring the H5 and NPC circuits have been used to develop the graphs in Figure 6. The calculated active power values are divided into conduction, switching and total losses.

a) b)

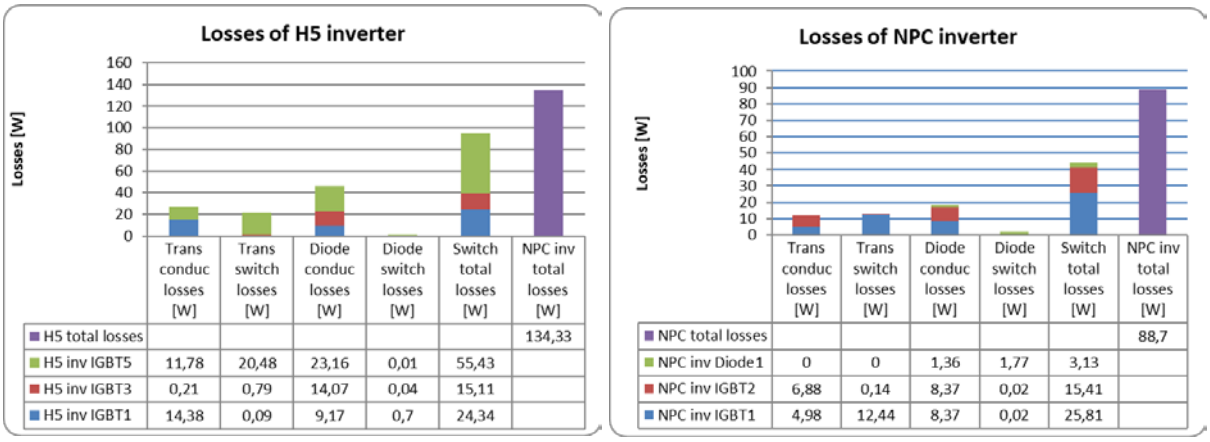


Figure. 7: Comparison of power loss on individual components of a) H5 inverter , b) NPC inverter.

As was shown in Fig. 7, the lower power losses over both discussed converter topology has the NPC inverter, although the NPC inverter consist six semiconductors – four transistors and two diodes. The H5 converter is characterized by higher power losses because of five activ active elements and control scheme while the total number of semiconductor is lower compare to NPC.

5. Conclusion

The power loss analysis of transformer-less inverters for renewable applications has been performed considering parameters of switches. The analysis of the characteristics of those converters, was done. Thermal Module for a real semiconductor device model from PSIM was used. The main advantages of these systems are the small power losses as shown in the paper, low cost and small size. Simulation results show the differences in loss dissipation achievable in each converter. Important to know is the power dissipated by each transistor to calculate the right parameters and ensure reliable operation of semiconductors.. For all discussed converters the power losses have been calculated in the combinations: IGBTs - anti-parallel diodes and power diode. The analysed NPC topologies have significant lower power dissipation so these topologies are suitable for high-performance renewable energy systems. Further investigations will be performed in the future by implementing and testing real prototypes for each of discussed topologies.

References

1. Bartman J.: *Accuracy of reflecting the waveforms of current and voltage through their spectrum determined by the standards regulating measurements*, Revue Roumaine des Sciences Techniques-Serie Électrotechnique et Énergétique vol. 61, pp.355-360, 2016.

2. Binkowski T., Malska W.: *The utilization of three phase V- connected inverter for supply high-speed motor*, Przegląd Elektrotechniczny, vol.87, Issue: 8, pp.34-37, 2011.

3. Buczek K., Malska W., Penar S.: *Use of PSIM software for modelling a small solar power station*, in Przegląd Elektrotechniczny, vol. 87 Issue: 8, pp. 42-47, 2011.

4. Calais M., Myrzik J., Spooner T., Agelidis V. G.: *Inverters for single-phase grid connected photovoltaic systems - An overview*, Proc. IEEE PESC'02, vol. 2, pp. 1995-2000, 2002.

5. Gonzalez R., Lopez L., Sanchis P., Marroyo L.: *Transformerless Inverter for Single-Phase Photovoltaic Systems*, IEEE Transactions on Power Electronics, vol. 22, Issue 2, pp.693 – 697, 2007.

6. Islam M., Mekhilef S.: *A New High Efficient Transformerless Inverter for Single Phase Grid-tied Photovoltaic System with Reactive Power Control*, 2015 IEEE Applied Power Electronics Conference and Exposition (APEC) pp.1666-1701, 2015.

7. Kjaer S. B., Pedersen J. K., Blaabjerg F., *"A review of single-phase grid-connected inverters for photovoltaic modules,"* IEEE Trans. Industry Applications, September – October 2005, vol. 41, no. 5, pp. 1292-1306, 2005.

8. Kerekes T., Teodorescu R., Liserre M., Klumpner C., Summer M.: *Evaluation of three-phase transformer less photovoltaic inverter topologies*, IEEE Trans Power Electron, 24 (9) p. 2202 -2211, 2009.

9. Meneses D., Blaabjerg F., Garc´ıa O., Cobos J. A.: *Review and comparison of step-up transformerless topologies for photovoltaic ac-module application*, IEEE Trans. Power Electron., vol. 28, no. 6, pp. 2649–2663, 2013.

10. Powersim Inc. "PSIM, User's Guide" Version 10.0, Release 5, January 2016, Copyright 2001-2016 Powersim Inc.

11. Suan F. T. K., Rahim N. A., Ping H. W.: *Modeling, analysis and control of various types of transformerless grid connected PV inverters*, in Proc. IEEE Clean Energy Technology, pp. 51–56, 2011.

12. Wintrich A., Nicolai U., Tursky W., Reimann T.: *Application Manual Power Semiconductors* , ISBN 978-3-938843-83-3, SEMIKRON International 2015.

13. Xiao H., Xie S.: *Transformerless split-inductor neutral point clamped three-level PV grid-connected inverter*, IEEE Trans. Power Electron., vol. 27, no. 4, pp. 1799–1808, 2012.

14. Xue Y., Chang L., Song P.: *Recent Developments in Topologies of Single-phase Buck-Boost Inverters for Small Distributed Power* Generators: An Overview, The 4th International Power Electronics and Motion Control Conference, IPEMC 2004., pp. 1118-1123, 2004.

15. Yang T., Koutroulis E., Sangwongwanich A., Blaabjerg F.: *Minimizing the Levelized Cost of Energy in Single-Phase Photovoltaic Systems with an Absolute Active Power*, in IEEE Industry Applications Magazine PP(99): pp.1-10 · December 2016 .

Tomasz BINKOWSKI[1],

Space Vector Modulation with One-Phase Asymmetry Compensation

The paper presents the space vector modulation strategy with one-phase load asymmetry compensation used in three phase three wires voltage source inverter. The asymmetry compensation is elaborated based on the projective geometry. The scaling function was derived and the rotation operators were introduced for different cases of an asymmetry location. The whole analysis was conducted in the two dimensional Cartesian coordinate system by the output current behaviour analysis. The transformation functions were defined and verified by simulation. Based on the elaborated transformation, the control strategy was implemented in the vector type modulator in order to eliminate the current asymmetry.

1. Introduction

The voltage source inverter (VSI) of three phase, three wires topology (3P-3W) is used in many power applications. It is a base topology of three phase induction motor (IM) drives [1] and the BLDC motor drives [9]. The dynamically developing renewable energy often uses this converter to convert the produced energy and to deliver it to the power network [10,7]. In many cases, the quality of the energy is insufficient as a result of improper control or unpredictable disorder. The one factor which results in an energy quality indicators reduction is the load asymmetry coefficient. As an unbalanced load of 3P-3W VSI effect, the output current trajectory changes and ceases to be circular [2, 8]. In order to limit this deformation of the current trajectory several methods can be proposed. In [4] the fuzzy logic system used in the correction of the VSI output trajectory was presented. The scaling method of the modulating functions was investigated in [5] as a way leading to the circular form of current trajectory. As an effect of the modulating functions scaling the non-sinusoidal currents were observed at circular trajectory in case of load asymmetry. The similar effects have had the use of double-modulation in 3P-3W VSI with load asymmetry described in [6].

Assuming the target of the VSI control in the form of a current circular trajectory the unbalanced load effect must be corrected. The power converter operation at unbalanced load is investigated by many researchers. In the paper the control strategy was introduced. Based on the mathematical model of load asymmetry of 3P-3W VSI the control strategy was defined and verified using the popular Power Sim software [3,11]. The paper is limited only to the one phase load asymmetry analysis. The other cases will not be considered.

2. Space vector modulation

The three phase, three wires VSI can be in one of the eight states. The two states connect the load terminals to the positive or negative pole of the DC source. The other states allow energy flows between source and load. These states can be illustrated in $\alpha\beta$ coordinate system

[1] Ph.D., Rzeszow University of Technology, Al. Powstancow Warszawy 12, 35959 Rzeszow, POLAND, tel. +48 178651974, e-mail: tbinkow@prz.edu.pl

as the active vectors, which position describes the matrix **U,** defined in the projective geometry (1).

$$\mathbf{U}_{active} = \mathbf{U}_a = \frac{2}{3} U_d \cdot \begin{bmatrix} 1 & \frac{1}{2} & -\frac{1}{2} & -1 & -\frac{1}{2} & \frac{1}{2} \\ 0 & \frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} & 0 & -\frac{\sqrt{3}}{2} & -\frac{\sqrt{3}}{2} \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, \tag{1}$$

where U_d is the input DC voltage. As an effect of the active voltage vectors generation, in case of the balanced resistive load, the geometrical form of the load currents trajectory have the same character, scaled by the load resistance value. In case of the load asymmetry, the output current vectors have different position then the voltage vectors. Such a situation effects the current trajectory as a deformation. The observed deformation connects the current vectors with the input voltage vectors by the scaling transformation. Such a transformation changes the position of the current vectors in relation to the proper voltage vectors. The asymmetry localization determines only a direction of scaling process. Thus, the output current trajectory form is different at the same asymmetry factor in relation to the asymmetry localization.

Taking into account the one phase asymmetry, the general form of the transformation matrix can be expressed as follows

$$\mathbf{T} = \mathbf{Rot}(\alpha)\mathbf{S}\,\mathbf{Rot}(-\alpha), \tag{2}$$

where α is the rotation angle relative to the *z* axis, **S** describes the general scaling in relation to the asymmetry localization and can be expressed as (3).

$$\mathbf{S} = \begin{bmatrix} \frac{3Z}{2Z_A + Z} & 0 & 0 \\ 0 & \frac{3Z}{2Z_x + Z} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

The coefficients and angle selection for **S** matrix can be taken from Table 1 in relation to the asymmetry localization.

Table 1. Parameters of **S** matrix

Asymmetry localization	Z_x	α
Phase A	1	0
Phase B	Z_B	$\pi/6$
Phase C	Z_C	$-\pi/6$

The output current space vector in case of unbalanced load can be calculated as the matrix operation using transformation matrix **T** and the current space vector calculated for symmetrical load $\mathbf{I}_{\mathbf{sym}}$ as shown in (4).

$$\mathbf{I} = \mathbf{T}\mathbf{I}_{\mathbf{sym}} \tag{4}$$

An idea of the unbalanced load compensation in case of the one phase asymmetry was implemented in the VSI control strategy. Taking into account the transformation matrix shown as (2), the new position of the active vectors was calculated. Based on it the vector modulation strategy was updated, using new values of the vectors position. The general

scheme of a modulation technique for the load asymmetry compensation in relation to the classical SVM method is presented in Fig. 1.

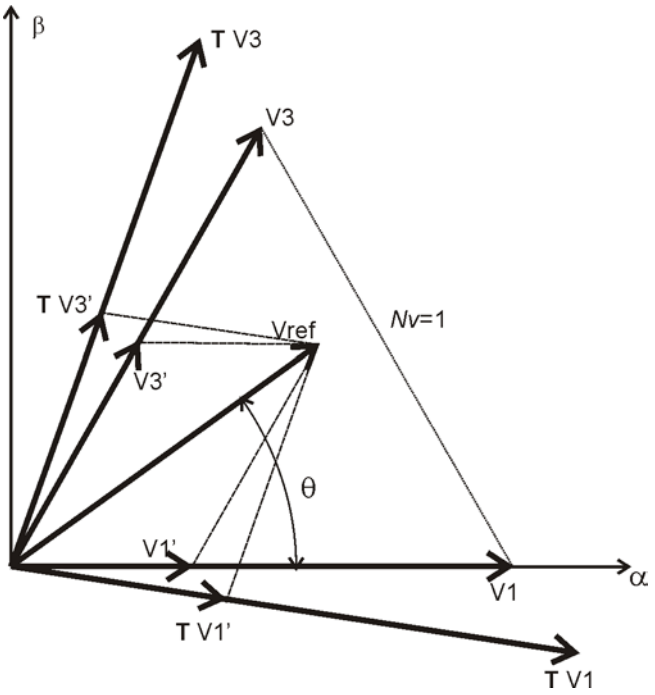


Figure. 1: The general scheme for the reference vector synthesis on the basis of V1 and V3 active vectors and on the **TV1** and **TV3** transformed active vectors for asymmetry compensation.

The new positions of the transformed active vectors, shown as **T**V1 and **T**V3 for sector number Nv=1, were used in the fill factor calculation for **T**V1' and **T**V3' vectors achievement using standard vector calculation methods for sinusoidal output waves. The number of active sector selects the proper VSI states, realized by vectors switched-on during time per each VSI tact calculated as (5), (6) and (7) for two active and one zero vector respectively. The k_l and k_r coefficients are the scaling factors of the active vectors, T_p is the modulation period (inverter's tact).

$$\mathbf{V}3' = \mathbf{V}3\frac{t_{a2}}{T_p}, \quad \frac{t_{a2}}{T_p} = k_l m \frac{\sin(\theta)}{\sin(\alpha_{N_v})}, \quad N_v = 1,2,3,4,5,6 \tag{5}$$

$$\mathbf{V}1' = \mathbf{V}1\frac{t_{a1}}{T_p}, \quad \frac{t_{a1}}{T_p} = k_r m \frac{\cos(\theta) - \sin(\theta)}{\tan(\alpha_{N_v})}, \quad N_v = 1,2,3,4,5,6 \tag{6}$$

$$\frac{t_0}{T_p} = 1 - \frac{t_{a1}}{T_p} - \frac{t_{a2}}{T_p} \tag{7}$$

3. Investigations

The main objective of performed investigations is to show the specific cases of the space vector modulation usage for asymmetry compensation. An application area description was deliberately omitted in order to confirm the effectiveness of single phase asymmetry compensation idea. Thus, the three characteristic features were investigated. The first of them refers to the static asymmetry compensation, where the currents are balanced. The proposed vector base modulation with an asymmetry compensation was investigated using Power Sim

software. The output currents were obtained in case of the classical vector modulation strategy and proposed modulation technique in order to achieve symmetrical load currents. The output currents in case of the unbalanced and balanced load are presented for different cases. Figure 2 shows the transient output currents of the VSI in case of an asymmetrical load where the classical space vector modulation strategy for the sinusoidal functions was used.

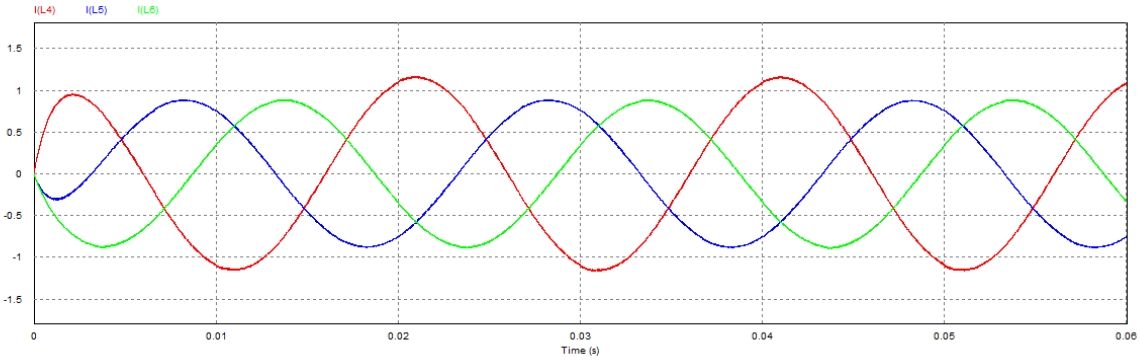


Figure. 2: The output currents waveforms of VSI in case of a load asymmetry.

The load asymmetry causes in unbalanced load. In order to achieve balanced currents the modified modulation strategy was used. As an effect of the transformed active vector usage in modulation strategy, the balanced currents were observed, as in Figure 3.

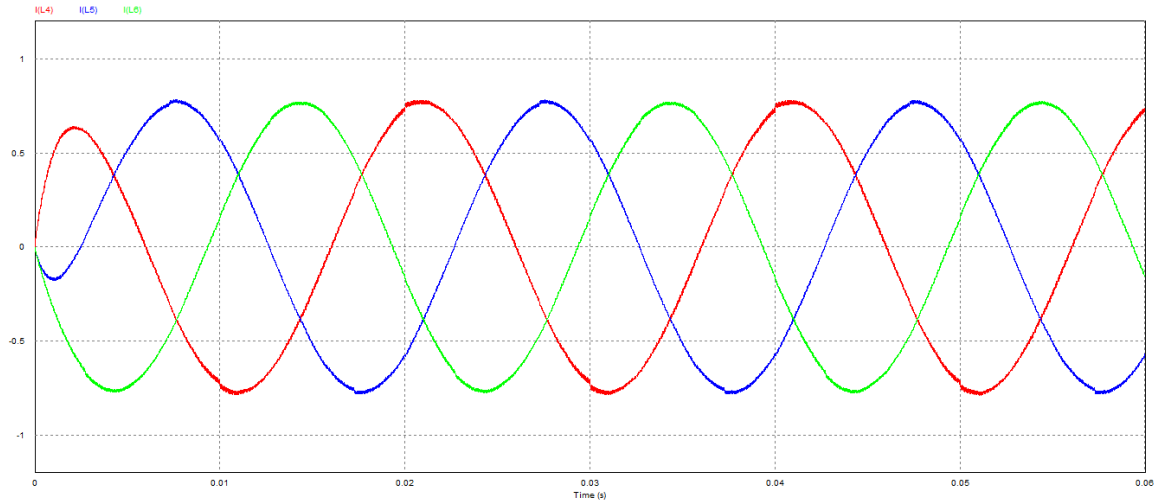


Figure. 3: The output currents waveforms of VSI in case of a load asymmetry using the space vector modulation with an asymmetry compensation.

The second investigated case refers to the situation, when the symmetrical load (R=1Ω, L=1mH) changes after 30 ms to the asymmetrical. The asymmetrical load is achieved as an effect of a partial short circuit in phase A or the VSI supply the three phase line with one phase switched loads. After this short circuit the resistance in phase A is R=0.5Ω and inductance is L=0.5mH. The R and L parameters in phases B and C were not changed. This case was simulated and transient waveforms for the classical sinusoidal modulation were presented in Figure 4. Figure 5 shows waveforms for the space vector modulation with an asymmetry compensation. Based on the results shown in Figure 3 and Figure 5 the static and

dynamic asymmetry compensation features of the proposed modulation were confirmed. They are very important futures when the load needs balanced currents to proper operation. For example, as an effect of an electro-mechanical resonances, which take place in a high speed sleek drives, the dynamically corrected currents can reduce rotor rubbing effect, when the motor is started.

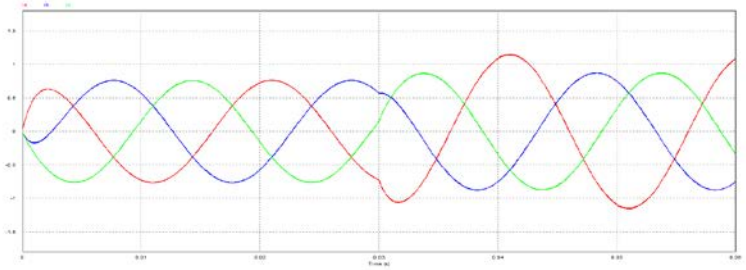


Figure. 4: The output currents waveforms of VSI in case of a load asymmetry occurring after 30ms.

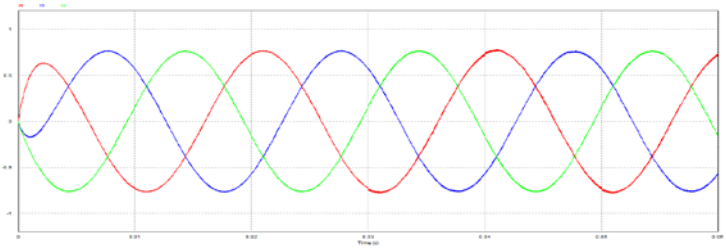


Figure. 5: The output currents waveforms of VSI in case of a load asymmetry occurring after 30ms using the space vector modulation with an asymmetry compensation.

The last presented case refers to the short circuit in the phase A after 30ms. Then the resistance and inductance in phase A equals zero. This feature of an asymmetry effect compensation is the current limitation in emergency situations. As an effect of short circuit, the current magnitude in a phase A is growing. The analysis of such a situation is shown in Figure 6. The usage of the presented modulator with asymmetry compensation eliminates an effect of short circuit, making the VSI currents become symmetrical. The simulation results for short circuit in the phase A are presented in Figure 7. The short circuit moment is observed as the increased high-frequency components.

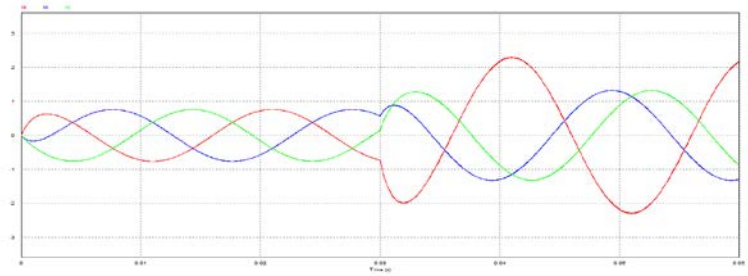


Figure. 6: The output currents waveforms of VSI in case of a short circuit in phase A after 30ms.

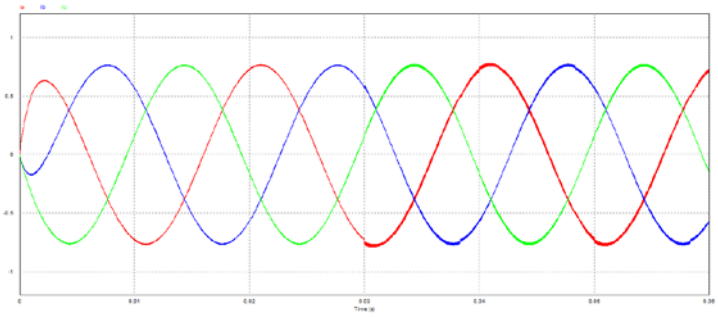


Figure. 7: The output currents waveforms of VSI in case of a short circuit in phase A after 30ms using the space vector modulation with an asymmetry compensation.

Figures 2, 4 and 6 show that the VSI legs have different loads resulted from asymmetry, extremely during short-circuit. The shorted load may damage transistors used in the VSI as an overcurrent effect. When the modulation with asymmetry compensation is used the output currents are symmetrical, switches are prevented and the control of a VSI maintain the supply making currents symmetrical as a part of a short-circuit protection system. This and other features will be tested in the future.

4. Remarks and conclusion

Presented solution refers to the asymmetry effect limitation by modulation process correction. Such a correction results from transformation matrix which was elaborated as an effect of the current first harmonic trajectory description in x-y coordinate system in case of the one-phase load asymmetry. The presented idea of the space vector modulation is the same as for widely known space vector modulation. The same active vectors were used as in the classical vector modulation, but the voltage sectors and the vector fill-factors were calculated in the different way. Such a situation makes the modulator as an open-loop control system. The same effect can be obtained using more complicated closed-loop controls, for example current control using d-q system, fuzzy correctors or dual-modulation method. Using the current correction in the modulation level, the control system treats the asymmetrical load as a symmetrical, even at short-circuit. The proposed modulation technique implemented in the classical vector control technique eliminates the asymmetrical load influence making the currents become symmetrical. If the proposed space vector modulation technique will be used to control of VSI which supplies the electrical motor, the supply maintenance with symmetrical currents prevents it in case of shorted load. The main advantages of elaborated modulation method in comparison to other known methods which compensate the load asymmetry are the simple structure of the modulator and simple calculations. Proposed load model based on the projective geometry allows to better understand the processes which take place at unbalanced load.

References

1. Ayyar K., Mohanapriyaa N., Nirmalkumar A.: *Comparative Study of Controller Design for Inverter Fed Indirect Vector Controlled Induction Motor Drive Systems, ”*, Journal of Testing and Evaluation, Volume: 44, Issue: 3, may 2016

2. Babaei S., Bhattacharya S.: *A control structure for PWM-controlled static synchronous compensators under unbalanced conditions and grid faults*, International Journal Of Electrical Power & Energy Systems, Oct 2015, pp. 160-173

3. Bartman J.: *Accuracy of reflecting the waveforms of current and voltage through their spectrum determined by the standards regulating measurements,* Revue Roumaine des sciences techniques – serie electrotechnique et energetique, Volume: 61 Issue: 4 Pages: 355-360 Published: oct-dec 2016

4. Binkowski T.: *Influence of Current Sampling Frequency on Voltage Source Inverter Fuzzy Correction, "*, 13th Selected Issues of Electrical Engineering and Electronics, Poland, May 2016

5. Binkowski T.: *Current trajectory correction of VSI at asymmetrical load*, PES, ISBN/ISSN: 978-83-927653-7-0, Koscielisko, Poland 2014

6. Binkowski T.: *Application of Double-Modulation to Correct Currents Trajectory of Voltage Source Inverter*, Conference on Progress in Applied Electrical Engineering (PAEE),1 Koscielisko, Poland, 2016

7. Buczek K., Malska W., Penar S. : *Use of PSIM software for modelling a small solar power station*, Przegląd Elektrotechniczny, Volume: 87 Issue: 8 Pages: 42-47, 2011

8. Choorak C., Kinnares V.: *Speed Control of Asymmetrical Two-Phase Induction Machine Using Three-Leg Space Vector PWM Voltage Source Inverter*, 16th International Power Electronics and Motion Control Conference and Exposition, Sep 2014, pp. 264-269

9. Kim MK, Bae HS, Suh BS: *Comparison of IGBT and MOSFET inverters in low-power BLDC motor drives*, 2006 IEEE Power Electronics Specialists Conference, Vol 1-7, Book Series: IEEE Power Electronics Specialists Conference Records, 2006

10. Sobczynski D.: *Simulation Study of Small Isolated Power Electronic Converters for PV System*, Progress in Applied Electrical Engineering (PAEE), Koscielisko, Poland, Date: Jun 26-Jul 01, 2016

11. Sobczynski D.: *Space vector technique in two-phase inverter*, Przeglad Elektrotechniczny Volume: 83, Issue: 2, Pages: 71-74, 2007

Arkadiusz AMBROZIAK[1], Magdalena STASIAK-BIENIECKA[1], Piotr BORKOWSKI[1]

Numerical and experimental analysis of heat fluxes in a smart building

Abstract. In the paper the simulation and measurements of temperature distribution and humidity in a room of smart building have been discussed. Initially, mathematical and simulation models have been created. For this purpose ANSYS Fluent has been used. The values obtained from measurements and simulations have been compared and errors have been analyzed. These studies are preliminary and concern of providing thermal comfort level in smart building and HVAC (Heating, Ventilation, Air Conditioning) control. In this work only the thermal comfort is considered.

1. Introduction

The sense of temperature comfort is related to several environmental factors such as lightening, humidity and air quality. The American Society of Heating Refrigerating and Air Conditioning Engineers (ASHRAE) proposed the thermal sensation scale (Table 1) with the purpose of quantifying people's thermal sensation (Predicted Mean Vote – PMV). PMV is an indicator, which allows estimating the predicted average value of rating in a seven-level scale of thermal sensations by a large group of people [1]. The index depends on six factors: metabolic rate, clothing insulation, air temperature and humidity, air velocity and the mean radiant temperature.

Table 1: The ASHRAE thermal sensation scale

cold	cool	slightly cool	neutral	slightly warm	warm	hot
-3	-2	-1	0	1	2	3

One more indicator – PPD (Predicted Percentage of Dissatisfied) is used for thermal comfort assessment. Everyone feels warm and cold in a different way – this is a subjective assessment. That is why it is important to know what percentage of people will be dissatisfied with the ambient temperature. The relationship between PMV and PPD was evaluated during the experiments and it is written as a mathematical formula (1):

$$PPD = 100 - 95e^{-0.03353PMV^4 - 0.2179PMV^2} \tag{1}$$

With this formula we can conclude that in the case of thermal indifference of the environment ($PMV = 0$), about 5% of people may be dissatisfied with the ambient temperature.

The norm [1], divides the rooms into three categories as shown in Table 2. At least Category B is recommended to ensure the thermal comfort of people in the room.

[1]Department of Electrical Apparatus, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, e-mail: arkadiusz.ambroziak@p.lodz.pl (Ph.D. student), magdalena.stasiak@p.lodz.pl (Ph.D.), piotr.borkowski@p.lodz.pl (Professor)

Table 2: Categories of comfort defined in [1]

Category	PPD	PMV
A	<6%	-0.2<PMV<0.2
B	<10%	-0.5<PMV<0.5
C	<15%	-0.7<PMV<0.7

The indoor temperature depends on many factors, such as: the outdoor temperature, relative humidity, solar radiation, wind speed, indoor human activity, etc. All of them are difficult to measure or to quantify for an ordinary household.

2. Mathematical and simulation models of the room in a smart building

The energy balance in a room is described by the differential equations (2).

$$\begin{aligned} C_p \frac{\partial T_r}{\partial t} &= \dot{Q}_d - \dot{Q}_o \\ C_p &= m_p c_p = \rho_p V_r c_p \end{aligned} \tag{2}$$

where: $\dot{Q}_d$ – heat flux supplied to the system, $\dot{Q}_o$ – heat flux emitted by the system, C_p – heat capacity of the air, T_r – air temperature, m_p – air mass, c_p – specific heat at constant pressure, V_r – volume of the room and ρ_p – air density.

Ansys Fluent takes into account the heat capacity of the walls through the specific heat parameter of the wall material. Walls are defined by 2R1C model. Its' mathematical description is given by equation (3).

$$\rho_w V_w c_w \frac{\partial T_w}{\partial t} = \frac{T_{ext} - T_w}{R_1} - \frac{T_w - T_r}{R_2} + \dot{Q}_{rad} \tag{3}$$

where: T_r – air temperature in room, c_w – specific heat of wall material, V_w – volume of the wall, ρ_w – density of wall material, T_w – temperature of the wall, T_{ext} – external air temperature, R_1 – thermal resistance from the outside to the wall, R_2 – thermal resistance from the wall to the room, $\dot{Q}_{rad}$ – heat flux related to solar radiation.

In this research, Ansys Fluent has been used to the solve the partial differential equations. Boundary conditions for external walls were mixed (convection and radiation), but for internal walls were only convective. In Ansys Fluent [2] the convection boundary condition is described by equation (4).

$$q = \alpha(T_w - T_p) \tag{4}$$

where: q – heat flux, α – convective heat transfer coefficient, T_w – wall surface temperature, T_p – local fluid temperature. In the external radiation boundary condition the heat flux to the wall is computed by (5).

$$q = \varepsilon_{ext} \sigma (T_\infty^4 - T_w^4) \tag{5}$$

where: q – heat flux to the wall, ε_{ext} – emissivity of the external wall surface, σ – Stefan-Boltzmann constant, T_∞ – temperature of the radiation source on the exterior of the domain, T_w – temperature of the wall surface.

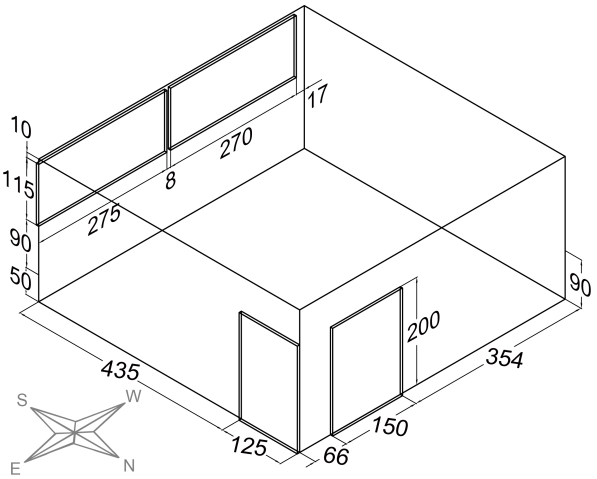


Figure 1: Dimensions of the room (in centimeters)

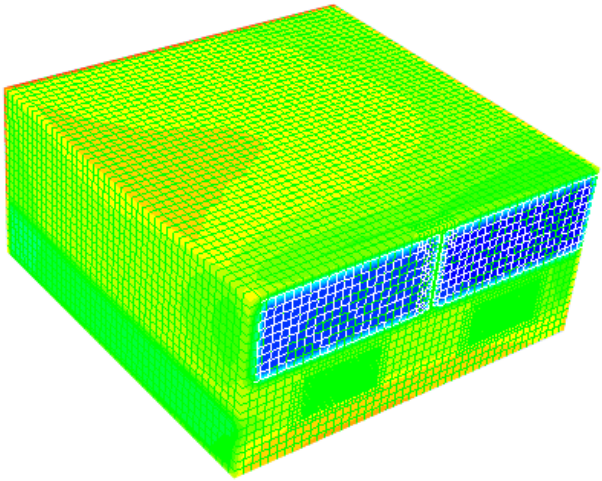


Figure 2: Mesh used in simulation

The proposed model (Fig. 1) has been discretized by hexahedron mesh in 3D, which consists of 60052 nodes and 53263 elements (Fig. 2).

Simulations have been conducted for 5 days (from 5 to 10 May), taking into account the actual weather data. In order to improve the accuracy of the simulation the procedure in C (script User Defined Function [3]) has been developed and the mesh for the window model has been thickened. The input data for this script are: month, day, and start time of the simulation, the time zone, the duration of the simulation, and the path to the weather data file. On this basis, the data needed to calculate the boundary conditions are determined: heat transfer coefficient for external walls, direct and diffuse radiation, outdoor temperature, increase of the temperature on the surface of the walls related to the solar radiation. ASHRAE papers and models have been used to determine the heat transfer coefficient [4] and the natural convection heat transfer coefficient for inner walls [5]. Stefan-Boltzmann's law has been used to calculate the increase of the temperature on the wall caused by the solar radiation.

The considered room is located on the third floor of a three-storey building at the Technical University of Lodz. The room is equipped with HVAC system. The dimensions of the room, location and size of windows and doors are shown in Figure 1.

The room is equipped with building management system produced by LCN company. Therefore all sensors used in measurements are compatible with this system. For measuring the temperature LCN-GRT sensors have been used. The resolution of these sensors is 0.1°C and the accuracy is 0.3°C (for temperatures between 15°C -30°C). For measuring the relative humidity LCN-EFS sensors have been used. Their resolution is 1% and the accuracy is $\pm$4% in the range 20%-80%.

In order to control and visualize all LCN system data an additional module from DOMIQ company has been installed. It collects all current data from every module in LCN network. Data acquisition is done via LAN. A Raspberry Pi microcomputer with own program connects to the DOMIQ module at regular intervals. All data is collected and stored in the PostgreSQL database.

The weather data needed to complete the boundary and initial value conditions for the simulations have been collected from the weather station located at Lodz Reymont Airport. The sunlight data comes from the weather station that is in the considered room. The sun sensor

Table 3: Real materials and equivalent material of western wall

	Thickenss [m]	Density [kg/m3]	Specific heat [J/kg*K]	Thermal conductivity [W/m*K]
Styrofoam EPS 70-040	0.12	20	1460	0.04
Prefabricated concrete element	0.16	2300	1000	2.3
Styrofoam	0.02	20	1460	0.045
Aerated concrete	0.12	500	840	0.17
Plaster	0.01	1700	840	0.8
Equivalent material (on the bottom)	0.6	1086.33	962.42	0.69
Equivalent material	0.43	1041.4	975.364	0.102

is connected to the intelligent building system and is located at the southern wall of the room. The sensor provides the sunlight data in luxes, and the values must be converted into W/m^2 [7], which is used in Fluent. This relation has been determined experimentally [6].

On the basis of the guidelines contained in Norm [1] and own experiments the positions of the temperature and humidity sensors have been set in the center of the western wall at 0.1, 0.8 and 1.1 meter.

Due to the complex structure of the walls, floor and ceiling in the considered room, it was necessary to determine the average thermal conductivity, density and specific heat for each wall. Table 3 present the complex structure of the selected wall.

3. Results and conclusion

The simulations and measurements have been carried out for the May for Lodz – latitude 51°43'19"N, longitude 019°23'53"E . The sunrise was about 5 AM and the sunset at 8 PM (hour from 5 AM to 8 PM – the day). Figures 3–5 present the simulated and real temperatures run over one day with day and night division.

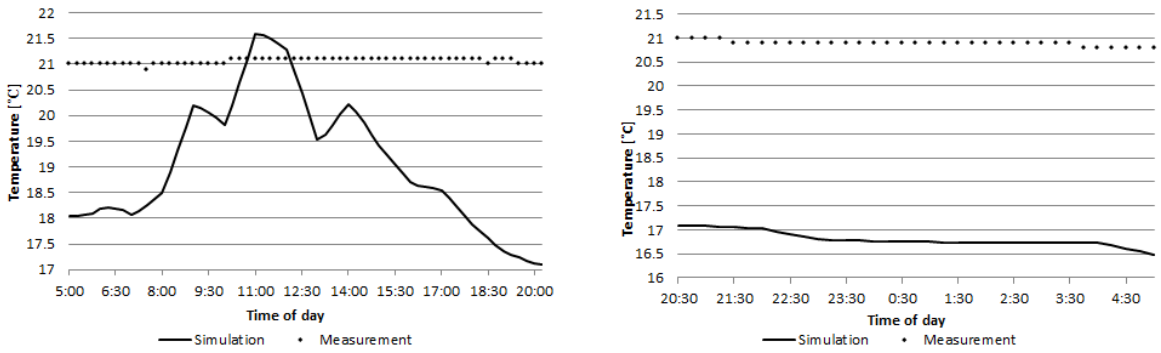
Divergences in temperature values in measurements and simulations are the result of application of many additional models, averaging and indirect values in the simulations.

In the measurement data analysis, the accuracy of the temperature sensor should be considered, it is 0.3°C.

A statistical evaluation of the measurement was made and a random error was calculated for the measurement data. The random error is described by the formula (6) – the mean standard deviation estimator [8].

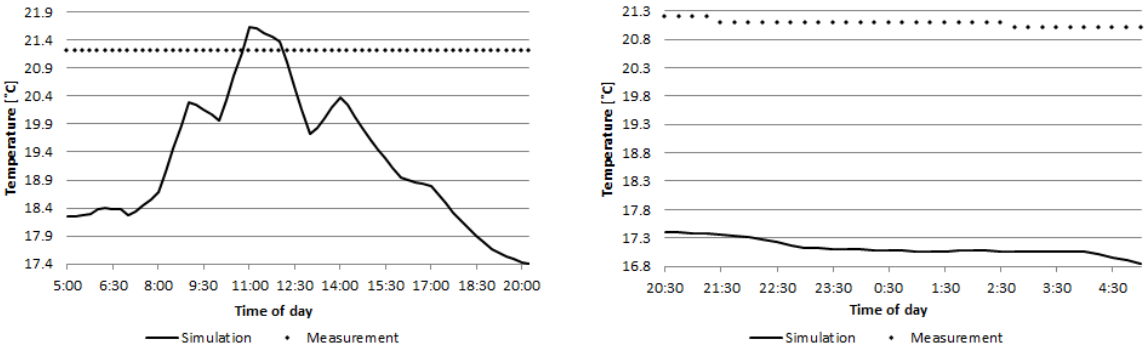
$$u(x) \equiv S(\bar{x}) = \sqrt{\frac{\sum_{i=1}^{m}(x_i - \bar{x})^2}{n(n-1)}} \qquad (6)$$
$$u(x) = 0.02$$

Taking into account sensor specifications, a systematic error connected with the accuracy of the measurements was calculated. In general, the uncertainty of measurements is described by



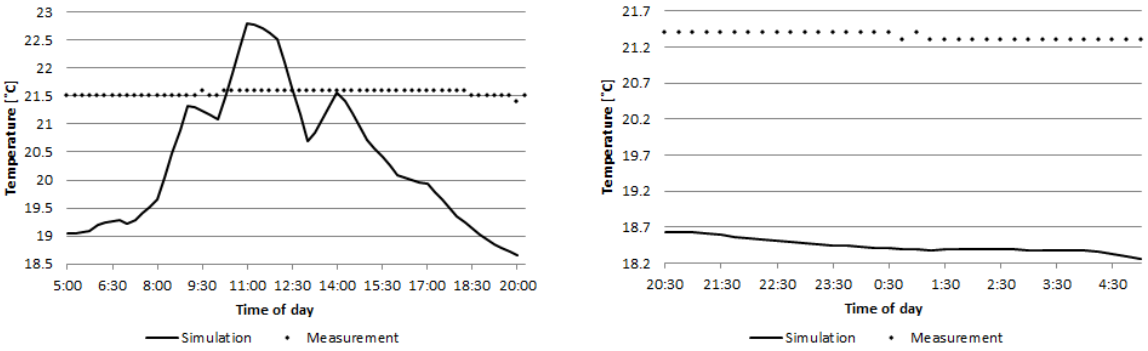
(a) Day (b) Night

Figure 3: Temperature in the room at 0.1 m



(a) Day (b) Night

Figure 4: Temperature in the room at a height of 0.8 m



(a) Day (b) Night

Figure 5: Temperature in the room at 1.1 m

the formula (7) [8]:

$$\Delta x = \sqrt{(\Delta x_L)^2 + \frac{1}{3}(\Delta x_S)^2} \tag{7}$$

where: Δx_L – random error, Δx_S – systematic error.

The uncertainty measured for the measurement data from 5 to 10 May days was 0.1745, at the accuracy to the decimal place Δx=0.2.

Based on simulation data mean radiant temperature [10], average PMV and PPD [9] values has been calculated (by UDF). The results of PMV and PPD presents table 4 (standard deviation of PMV is 0.5). Therefore it is necessary to apply HVAC system for providing of thermal comfort.

Table 4: Averege PMV and PPD values for simulated room

Average PMV	Average PPD
-1.4	46%

Simulations and measurements were made for the spring month, without considering the heating elements in the room. The future research will take into account heating elements and presence of people in the room.

References

1. PN-EN ISO 7730:2006 – Ergonomia środowiska termicznego – Analityczne wyznaczanie i interpretacja komfortu termicznego z zastosowaniem obliczania wskaźników PMV i PPD oraz kryteriów miejscowego komfortu termicznego

2. Ansys Inc.: "Ansys Fluent User's Guide", Release 14.0, 2011

3. Ansys Inc.: "Ansys Fluent UDF Manual", Release 14.0, 2011

4. ASHRAE Task Group: "Procedure for determining heating and cooling loads for computerized energy calculations: Algorithms for building heat transfer subroutines", *New York*, 1975

5. ASHRAE Task Group: "2001 ASHRAE Handbook: Fundamentals", *Atlanta*, 2001

6. http://mib.gov.pl/2-Wskazniki_emisji_wartosci_opalowe_paliwa.htm

7. http://bccp.berkeley.edu/o/Academy/workshop08/08%20PDFs/Inv_Square_Law.pdf

8. Andrzej Zięba: "Analiza danych w naukach ścisłych i technice", Wydawnictwo Naukowe PWN, *Warszawa*, 2014

9. Sudoł-Szopińska I., Chojnacka A.: "Określanie warunków komfortu termicznego w pomieszczeniach za pomocą wskaźników PMV i PPD", *Bezpieczeństwo Pracy*, 2007, Vol. 1, nr 5

10. Catalina T.,Virgone J., Kuznik F.: "Evaluation of thermal comfort using combined CFD and experimentation study in a test room equipped with a cooling ceiling", *Building and Environment*, 2009, Vol. 44, nr 8

Grzegorz SYPNIEWSKI[1], Krzysztof OKARMA[2]

Highly Effective Number Recognition System Dedicated for Industrial Usage

In this paper an experimental system allowing an automatic recognition of digits and numbers has been presented. Strictly fitted character of the system and its industrial application is related to the requirements limiting the use of some typical OCR solutions e.g. based on Tesseract or application of the HOG descriptor. Obtained experimental results are promising providing high recognition accuracy and allowing to increase the effectiveness of a production line in a packaging company.

1. Introduction

The industrial need for quality and certification policy in one of the packaging manufacturing companies was the primary motivation of the presented work. In such a company, a system generating a post-production report is implemented within a production line. The final report consists information about several parameters from several dozen production cycles. It is created and printed using a dot-matrix printer by the automatic control system based on the PLC.

The lack of information on the minimum, maximum, and average process data values in such a report has resulted in the necessity of manual typing dozens of rows and several columns into a spreadsheet. It is obvious that such an operation may cause errors and generate additional costs. In addition, after the implementation of an automatic system, a highly qualified laboratory worker would be able to deal with other duties at that time.

The first attempt was the modification of the production file format. For some reasons, the PLC code could not be changed to supplement the report with the required data. Therefore it was decided to create a system allowing to calculate the missing parameters and archive them digitally. Several boundary conditions have been set. Firstly, the system should have an open code based on publicly available software with the appropriate license with the possibility of installation in several locations. It should also allow an easy change of character recognition parameters. Another important requirement is its very high performance and accuracy so that nobody should over-control it and make adjustments. Nevertheless, the recognition of letters is not demanded for its proper work. However, in addition to the recognition of digits, the system should be able to convert them into numbers, sort the results, place them in the appropriate columns, and calculate statistics. It means, the system has to know that the number in position X,Y represents e.g. density of material in third cycle.

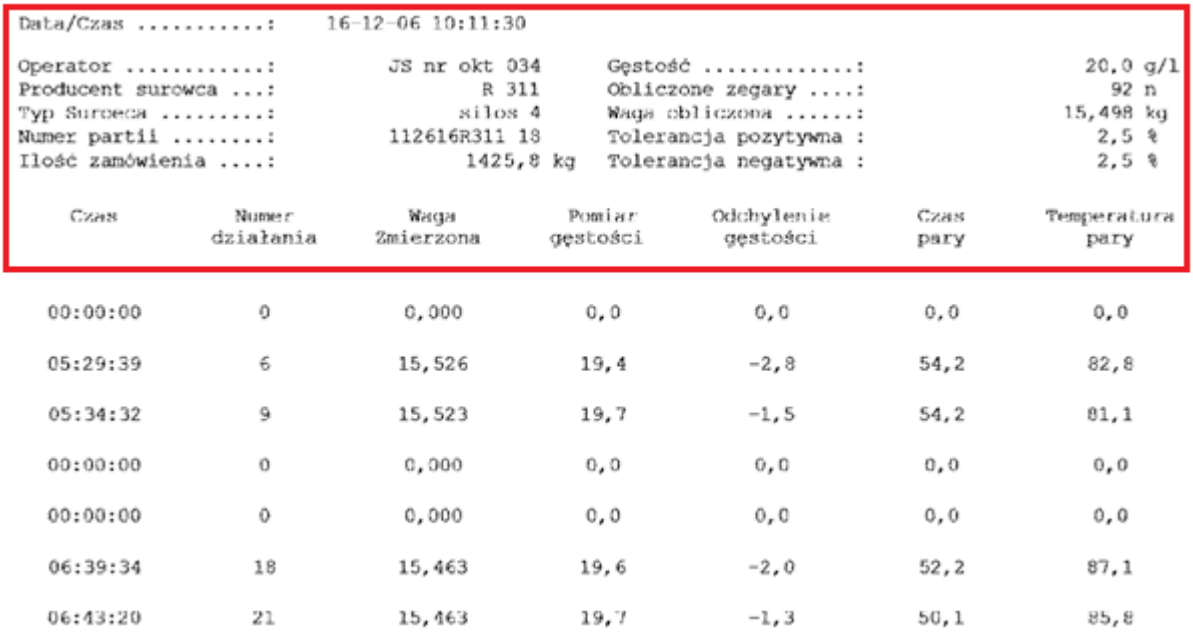
For this purpose, considering the licensing issues, the experimental environment consisting of SCILAB version 5.5.2 with SIVP and IPD libraries has been chosen. The program, after some modifications, works in OCTAVE 4.2 as well.

[1] Ph.D. student, West Pomeranian University of Technology, Szczecin - Faculty of Electrical Engineering, Sikorskiego 37, 70-313 Szczecin, POLAND, e-mail: Grzegorz.Sypniewski@zut.edu.pl
[2] Assoc. Prof., West Pomeranian University of Technology, Szczecin - Faculty of Electrical Engineering, Sikorskiego 37, 70-313 Szczecin, POLAND, tel. +48 91 449 53 13, e-mail: okarma@zut.edu.pl

2. Proposed solution

The production machine prints a report using a dot-matrix printer on a special paper that is troublesome for scanning. In addition, there is always the possibility of report loss, dirt or destruction making proper scanning impossible, not to mention that it takes time. To prevent this, a device emulating the printer has been used that stores the PLC data sent to the printer as a PRN file. It is in fact not a text file, but a code in the language ESC / P EPSON [1].

The next step has been related to processing the PRN source file into an image in TIFF format by a dedicated script. Considering the emulation of a monochrome printer, it has been assumed that the image is binary or grayscale (subjected to further binarization). This has been done utilizing, inter alia, the ImageMagick graphics package [2]. A fragment of the original report is presented in Figure 1 where some quality issues originating from the limited resolution and the used dot-matrix printer can be easily observed.



Data/Czas .....: 16-12-06 10:11:30

Operator .....:	JS nr okt 034	Gęstość .....:	20,0 g/l
Producent surowca ...:	R 311	Obliczone zegary ....:	92 n
Typ Surowca .....:	silos 4	Waga obliczona .....:	15,498 kg
Numer partii .....:	112616R311 18	Tolerancja pozytywna :	2,5 %
Ilość zamówienia ....:	1425,8 kg	Tolerancja negatywna :	2,5 %

Czas	Numer działania	Waga Zmierzona	Pomiar gęstości	Odchylenie gęstości	Czas pary	Temperatura pary
00:00:00	0	0,000	0,0	0,0	0,0	0,0
05:29:39	6	15,526	19,4	-2,8	54,2	82,8
05:34:32	9	15,523	19,7	-1,5	54,2	81,1
00:00:00	0	0,000	0,0	0,0	0,0	0,0
00:00:00	0	0,000	0,0	0,0	0,0	0,0
06:39:34	18	15,463	19,6	-2,0	52,2	87,1
06:43:20	21	15,463	19,7	-1,3	50,1	85,8

Figure 1: Fragment of the original report.

The source image is converted into the negative for the convenience of further processing and binarized (in case of the greyscale source file as mentioned earlier). Observing the numbers, some imperfections of their shapes, influencing the recognition accuracy, can easily be noticed. A typical example is shown in Figure 2 where the result of weighting of the raw material is presented.

Figure 2: Example of a number subjected to recognition.

In Figure 3 it can be observed that even some identical digits can be randomly changed. The application of some typical morphological operators such as erosion or dilation has not led to satisfactory results. Better results have been obtained using a median filter allowing to smooth the shapes of digits due to the elimination of single pixel distortions. What is significant, the best result has been achieved using uncommon square mask of 2×2 pixels, instead of default 3×3 mask. An exemplary result of such pre-processed image is shown in Figure 4.

Figure 3: Illustration of differences among the shapes of the same digit.

Figure. 4: Results of image preprocessing.

Another task has been related to the identification of the Region of Interest (ROI) as the recognition of the table header has not been required and only a table with numbers has been important.

Firstly, it is necessary to count how many rows are in the picture. Since the size of the header (marked by the red frame in Figure 1) is constant it can be easily removed and then only the table may be further processed. A simplification of this task is the fact that the source image is almost always very similar. At this stage picture is treated as the binary image represented as the matrix containing logical zeros and ones. The number of "ones" is counted for each row of the matrix. Grouping several rows of matrix containing ”ones”, it can be treated as a line of text. An exemplary line of text is shown in Figure 5.

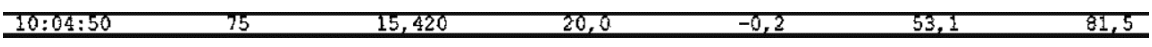


Figure 5: Illustration of an extracted single line of numbers.

Grouping individual characters can be made similarly to rowing. However, at this point there is a risk that there are still single pixels left so they have to be detected and rejected. It is important that the characters should not come into contact with each other as they might be improperly recognized. It may be assumed that single characters have a fixed height but they can have different width. The selected character is processed by a special function (discussed in the next section) that recognizes which given mini-array is the most similar. The output is recorded as a number representing the most similar pattern.

3. Classification

The recognition algorithm utilizes the fact that each digit has its own characteristics. Observing the shape of the digit "8" it can be seen that there are some areas where the pixels should be zero (representing the background of the binary image) whereas in the centre of the digit the pixels should be equal to 1. The digit "3" is similar to "8" but the pixels on the left side should be equal to zero. The digit “1” should have a vertical line in the central part of the ROI. The comma occurs only in the lowest part of the ROI array. A very important simplification of the recognition is the assumption of only one font occurring in the printed document. Moreover due to avoiding scanning and the use of a virtual printer a lot of typical errors do not appear. The rest of the orphan pixels is removed by the median filter mentioned in section 2 and the possible small rotations can be determined analysing the slope of each detected line of text or connecting the dots visible in the header of the document.

The algorithm is based on the extraction of the areas specific to a given digit (exemplary areas are shown in Figure 7) and calculation of the arithmetic mean of the coordinates of logical "ones" (in fact representing the centre of gravity of the binary sub-image representing a single digit) which should fall within the predefined limits. For example the local average of the area inside the representation of the digit "0" should be low although it may be greater than zero, because there can always be some single pixels differing from the background. The algorithm of character recognition is presented in Figure 6. Each matrix containing a single digit is tested step by step using consecutive logical conditions. A good example is the picture with the digit “three” shown in Figure 7. First it is checked for being similar enough to the digits 0, 1 and 2. When the condition is fulfilled, the program returns the recognized digit "3". If none of the conditions is fulfilled, the program returns the value "255" interpreted as "digit unknown".

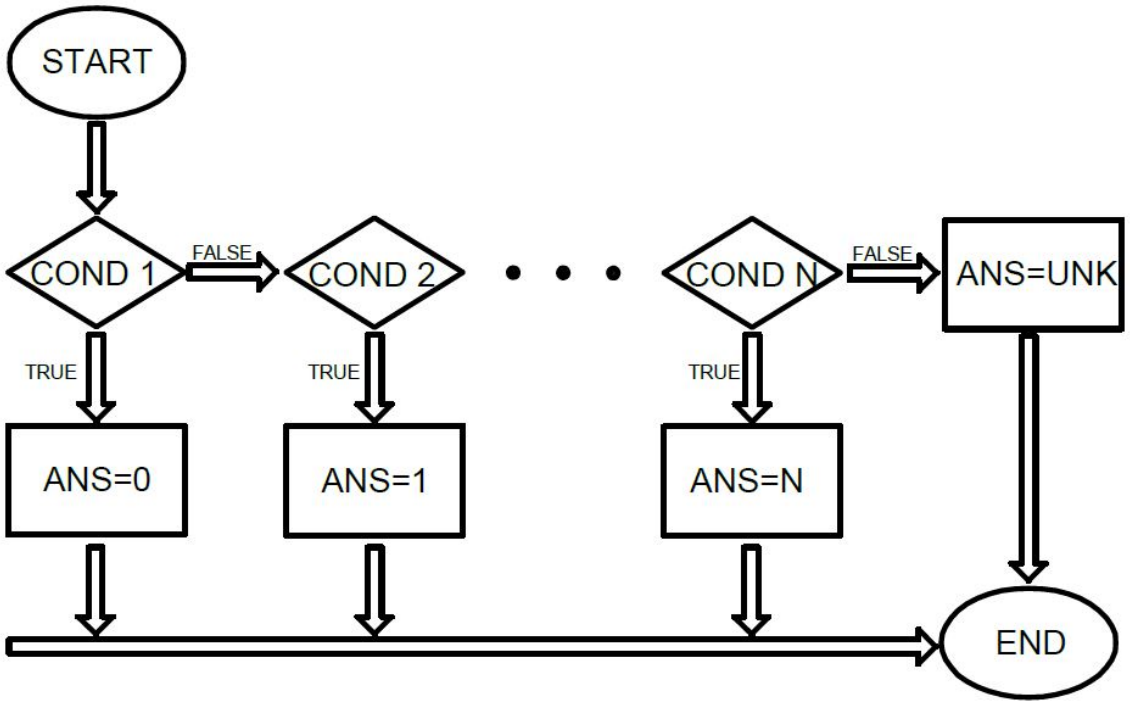


Figure 6: Algorithm of digits recognition.

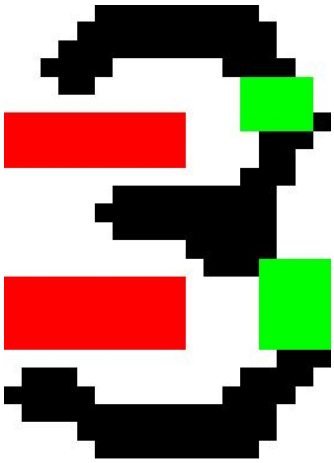


Figure 7: Example of logical condition – red color denotes the areas with expected low values (mainly "zeros" representing the background) and the green color the areas with expected high values (mainly "ones").

To correctly recognize a character, three or four conditions should be used which can be expressed as follows in the fragment of code in Scilab (for the digit "3" as shown in Figure 7 assuming the origin of the coordinate system in the determined centre of gravity of the digit):

```
If mean(test1(7:9,1:10))<0.2 & mean(test1(16:19,1:10))<0.2 &
mean(test1(5:7,$-4:$-1))>0.55 & mean(test1(15:19,$-3:$))>0.6 then
    answer=3  //it is ‘3’
```

During this part of the task the appropriate trigger levels had to be selected experimentally. Most of the characters have been quite easy to classify and the most attention has been devoted to the digit "8", which caused the biggest problems due to the similarity to the numbers "0", "3" and "9". As shown above, the exemplary digit "3" is recognized using the four conditions, however some of the digits may be recognized using only three conditions.

After processing the entire line, single characters (digits) can be converted to numbers and assigned to appropriate values. For this purpose a special function has been prepared in which characters of the char type are grouped into one string depending on their position in the image. The string is converted into a number and written to the appropriate place in the matrix. Finally, the resulting matrix is presented on the main screen. Having recognized the numbers, displaying the statistics is just a few lines of code. As long as the input report is prepared correctly and there are no empty production cycles, it is possible to calculate easily some additional parameters such as weight and total volume.

4. Comparison with alternative solutions

Probably the most obvious question is why our own character recognition system has been created seeing that the OCR technology has been developed for a very long time and usually leads to excellent results. To verify its usefulness during the production tests, the Tesseract [3] package has been used as a well-known system developed for many years by Google. The principle of its operation is discussed in the paper [4].

After several experiments, the Tesseract's performance in recognizing numbers from a production report can be estimated at around 92%. The most common mistakes are the digit "0" converted to "8" and the "minus" sign converted to "7". A frequent mistake is also the substitution of the comma with a dot. Unfortunately, the application of this solution has not been effective enough. There have also been some problems of exporting the file to a spreadsheet and arranging the data in tables. The time spent on these activities is still significant, although the application of the Tesseract OCR has simplified the work related to manual typing.

To compare the effectiveness of the presented algorithm, the effectiveness of OCR recognition implemented in the Matlab package has been checked as well. This comparison has led to very interesting results and the OCR function implemented in Matlab is very versatile and turned out to be much more effective than Tesseract with its effectiveness about 97.2%. However, it has been assumed that there have been no letters to recognize. It is interesting that it has been mistaken the most often for the digit "7". Sometimes "0" has been wrongly recognized. In addition, some commas, dots and spaces appear making it difficult to properly process the numbers.

The next solution that has been compared analyzing the quality of the digit recognition has been the application of the Histogram of Oriented Gradients (HOG) [5] and the SVM (Support Vector Machine)[5] used for classification. The idea of HOG is based on dividing the image into small cells, calculating gradients and then computing histograms of these gradients. These features keep the representation of an object and allow for a little variations in shape as well. The second part is training of the SVM classifier with a large amount of

training data. Such a solution gives very good results, comparable with the algorithm presented, but the amount of work needed to implement this algorithm is significant for commercial usage and the calculation time is also very long. Testing script has been written in Matlab, with use of existing commands concerning HOG and SVM methods.

The most common mistake in the proposed algorithm is improper recognition of the digit "8". Sometimes there have been some problems with the digit "3" and the overall effectiveness is about 97.4%. The results obtained by recognition of ten samples are presented in Table 1. Each report used in experiments has about 775 signs in the Region of Interest.

Number of sample	Developed algorithm	Tesseract	HoG/SVM	Matlab OCR
1	1	3	2	2
2	2	0	3	2
3	2	8	3	2
4	0	1	4	2
5	2	10	3	1
6	3	8	3	5
7	3	14	2	2
8	3	6	2	4
9	2	10	3	1
10	2	1	3	1
Sum	**20**	**61**	**28**	**22**

Table 1: Number of errors in recognizing digits.

For the purpose of reliable comparison of time of operation the proposed industrial algorithm has been implemented additionally in Matlab and all calculations have been made using the same platform. The average results of comparison obtained for several runs for each algorithm are shown in Table 2.

	Developed algorithm	Tesseract	HOG/ SVM	Matlab OCR
Average time of calculation [s]	4.1	6.1	106.6	7.9

Table 2: Comparison of the average time of calculation for the set of images used in experiments.

5. Remarks and conclusions

The number recognition system presented in the paper is highly effective. During its normal usage its accuracy is nearly 100%. The common weakness is related to recognition of the digit "8", but this is quite a rare case. The other tested solutions have not been resistant to small artefacts, and treated them as dots.

In the near future, a solution similar to presented system is planned to be implemented for the next production line. One of the other directions of our future work may be related to the

automatic prediction of the recognition accuracy e.g. in relation to the assessment of the binarization results [6] in similar industrial systems without the use of virtual printers where the analysed image may be obtained by a camera. Another possibility of further extension of the developed application is the use of shape descriptors [7] allowing a more reliable recognition of different types of various characters also in presence of distortions.

A disadvantage of the proposed solution is its specificity limiting its usage and much lower universality in comparison to some other general purpose OCR algorithms. Nevertheless, in industrial applications it is compensated by its fast work. In the applications where some limitations such as e.g. available font shapes or the overall organization of documents are known, the proposed solution may perform better than the universal OCR solutions not only in terms of speed but considering the recognition accuracy as well. Another advantage is fully automated work allowing a proper computation of all necessary data as some other solutions usually have problems working with columns. Although there is a significant number of different similar solutions proposed by many authors, they are concerned on deep learning approach [8] or OpenCV library [9] and could be quite problematic in usage for end users.

References

1. https://files.support.epson.com/pdf/general/escp2ref.pdf
2. ImageMagick. https://www.imagemagick.org/script/index.php
3. Tesseract Open Source OCR Engine. https://github.com/tesseract-ocr/
4. Smith R.: *An Overview of the Tesseract OCR Engine,* Proceedings of the 9th IEEE International Conference on Document Analysis and Recognition ICDAR, pp. 629-633, 2007.
5. Dalal N., Triggs B.: *Histograms of Oriented Gradients for Human Detection*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 886-893, 2005.
6. Lech P., Okarma K.: *Prediction of the Optical Character Recognition Accuracy based on the Combined Assessment of Image Binarization Results,* Elektronika ir Elektrotechnika, vol. 21 no. 6, pp. 62-65, 2015.
7. Frejlichowski D., Forczmański P.: *General Shape Analysis Applied to Stamps Retrieval from Scanned Documents*, Proceedings of the International Conference on Artificial Intelligence: Methodology, Systems, and Applications AIMSA, Lecture Notes in Computer Science, vol. 6304, pp. 251-260, Springer, 2010.
8. Wu R., Yan S., Shan Y., Dang Q., Sun G.*: Deep Image: Scaling up Image Recognition*
9. Kumari S., Gupta L., Gupta P.: *Automatic License Plate Recognition Using OpenCV and Neural Network,* International Journal of Computer Science Trends and Technology Volume 5 Issue 3, May-June 2017

Kazimierz KIEŁKOWICZ[1]

Automatic unsupervised machine learning in latent concept space

Paper introduced unsupervised machine learning method based on modified Bat Algorithm. Modified Bat Algorithm has automatically adjusted control parameters using fuzzy inference system. Learning is performed in so-called latent concept space induced using SVD decomposition. A centroid-based Bat solution encoding is introduced allowing optimization process to search for both optimum number of clusters as well as optimal centroids position in search space. Two cost function were used to evaluate solution CH-index and DB-index. Presented algorithm was tested on well-known Iris dataset.

1. Introduction

Unsupervised learning also known as cluster analysis or classification is a process of exploring the unknown structure of the data by separating finite data set into so called clusters. In particular, partitinal unsupervised learning divide data object into some predefined number of clusters. Formally in mathematical terms it can be stated as follows [11]. Given a set of m objects $X = \{x_1, \dots, x_j, \dots, x_m\}$, where $x_j = (x_{j1}, x_{j2}, \dots, xj_n) \in R^n$, find a K-partition of X, $C = \{C_1, \dots, C_K\}$ $(K \leq m)$ such that $C_i \neq \emptyset$ for $(i = 1, \dots, K)$ and $\bigcup_{i=1}^{K} C_i = X$ and $C_i \cap C_j = \emptyset$ for $i, j = 1, \dots, K$ and $i \neq j$. The problem of unsupervised learning can be stated as optimization a predefined criterion function so that data objects belonging to cluster C_i are more similar then objects belonging to cluster C_j. In principal, such optimization can be performed by so called brute force methods, but in practice it's often infeasible. An alternative is to use a heuristic algorithm like K-means. However, such a hill-climbing-like algorithm suffer from being sensitive to initial starting point and are likely to get stuck in local minima. To overcome these limitation, it's advised to use more sophisticate metaheuristic algorithm.

In principal metaheuristics algorithms can be divided into few groups, e.g. algorithms based on evolutionary approach that models evolutionary process or algorithms exploring phenomena of a Swarm Intelligence [6]. Others approach, such as algorithms for modelling response of a human immune system (e.g. Artificial Immune System algorithms) might be considered as separate category due to their multiplicity of proposed solutions.

Metaheuristics methods which are focused on exploring models of a natural evolution are (mostly but not limited to) as follows: Genetic Algorithms (GA) [10], Genetic Programming (GP) and Differential Evolution (DE) [21]. Algorithms based on Swarm Intelligence are broadly presented by Particle Swarm Optimization (PSO) [13], Ant Colony Optimization (ACO) [4] or some of its modifications.

In recent years, unsupervised learning based on a natural inspired metaheuristic algorithm, including PSO [7][14][22] and Ant algorithms [5] has attracted attention thanks to it demonstrated effectiveness in solving complicated optimisation problems

[1] M. Sc. EE., Department of Computer Science, Faculty of Electrical and Computer Engineering, Cracow University of Technology, Cracow, Poland,
tel. +(48-12) 628-25-23, fax: +(48-12) 628-25-23, e-mail: kkielkowicz@pk.edu.pl

Recently introduced method, based on population of solutions which explore phenomena of Swarm Intelligence was presented by Yang [23] in 2010 and it is called Bat Algorithm (BA). In [23] by modelling the behaviour of bats hunting for prey and by exploring phenomena of their echolocation capabilities, author managed to incorporate methods for balancing the exploration phase as well as exploitation phase of a modern Swarm Based Algorithms.

Bat Algorithms had already been applied to solve numerous hard optimization problems such as multi-criteria optimization [24] or optimization of topology of microelectronic circuits [9]. Growing popularity of the Bat Algorithm has encouraged researchers to focus their work on its further improvements. Most work has been done within the area of hybridization of Bat Algorithm with other metaheuristics or local search methods [8]. Some other solutions were involved within the area of adding self-adaptability capabilities to algorithm [1]. Some works has also been in area of adaptation of standard Bat Algorithm for binary problems [20] and on modifying scheme of acceptance of a new solution [16].

Unfortunately, most of these modifications not only improves the quality of obtained solutions, but also increases the number of control parameters that are needed to be set to obtain solutions of expected quality. This makes such solutions quite impractical. More recent work [15] introduces fuzzy logic control system build on Mamdami-Type inference method to control the exploration and exploitation phase of an evolutionary system based on modified Bat Algorithm [16]. Application of fuzzy logic to control the exploration and exploitation phase frees the user from explicit specifying control parameters and only require defining expected behaviour in a human readable knowledge-base form of *if-then* sentence.

This paper introduced modified Bat Algorithm with automatic adjusting control parameters thru fuzzy inference Mandami-type system for a problem of unsupervised learning in so-called concept (or semantic) features space induced from raw data point input space. Presented algorithm learns the optimal clustering of the data as well as optimal number of clusters in dataset.

Section 2 describe Modified Bat Algorithm used as optimization algorithm. Section 3 provide information about incorporating fuzzy logic controller to dynamically adjust the behaviour of a Modified Bat Algorithm and define linguistic variables used in “if-then” sentences in knowledge-base. Section 4 introduce solution encoding as Bat positing in search space appropriate for both searching optimal clusters as well as optimal number of clusters in the context of clustering N data poins into K clusters with cluster centroids. Section 5 describe used cost functions (DB and CH Index) that are used to evaluate given solution. Section 6 provide information about inducing semantic feature space from raw data point space using singular value decomposition (SVD). Section 7 provide experimental results. Section 8 concludes these paper.

2. Modified Bat Algorithm

Bat Algorithm is recently proposed bio-inspired metaheuristics method for solving hard real valued optimization tasks. It tries to mimic behaviour of bats hunting for their prey. Algorithm was introduced by Yang in 2010 [23]. Bat Algorithm is based on population of bats, which by flying thru solution search space explore it in order to find interesting areas. Each single bat represents one solution in multi-dimensional search space. Solutions are evaluated in terms of their fit value by provided fit function. Full discussion about Bat Algorithm, it shortcoming as well as some proposed modification to form a Modified Bat Algorithm (MBA) can be found author other work in [16]. Modified Bat Algorithm is presented as pseudocode in figure 1.

1:	Randomly initialize position x_i and velocity v_i of i-th bat in population. Bat is a encoded solution described in section 4
2:	Initialize pulsation frequency $Q_i \in [Q_{min}, Q_{max}]$, pulsation r_i and loudness A_i of i-th bat in population
3:	**while not** termination conditions are satisfied: Q_i=fuzzyInferenceSystem(diversity, iteration) Q adaptation using FLC described in section 3
4:	**for_each** bat in population:
5:	$v_i(t) = \alpha_i v_i(t-1) + Q_i(x^* - x_i(t-1)) + Q_i(x^*_{ever} - x_i(t-1))$ $x_i(t) = x_i(t-1) + v_i(t)$
6:	**if randn(0,1)** > r_i^t: $x_i' \leftarrow$ generate new solution around current bat x_i
7:	**if** $f(x_i') < f(x_i)$ **or randn(0,1)** < A_i^t: $x_i \leftarrow x_i'$ Update values of pulsation and loudness, respectively r_i^t and A_i^t as: $A_i^{t+1} \leftarrow \alpha A_i^t;\ r_i^{t+1} \leftarrow r_i^t(1 - \exp(-\gamma t))$
8:	Evaluate bats population using of of the fit function f described in section 5
9:	Find best bat in population and mark him as x^*
10	**if** $f(x^*) < f(x^*_{ever})$:
11	$x^*_{ever} \leftarrow x^*$

Fig 1. Pseudocode of Modified Bat Algorithm (MBA).

where:
r_i^t, A_i^t - pulsation and loudness of i-th bat,
γ, Q_{min}, Q_{max} - constants.

Bat velocity and position update equations:

$$v_i(t) = \alpha_i v_i(t-1) + Q_i(x^* - x_i(t-1)) + Q_i(x^*_{ever} - x_i(t-1)) \quad (1)$$

$$x_i(t) = x_i(t-1) + v_i(t) \quad (2)$$

where:
$v_i(t)$ - real-valued velocity vector of i-th bat,
$x_i(t)$ - real-valued position vector of i-th bat,
Q_i - pulsation frequency of i-th bat,
α_i - cognition coefficient of i-th bat,

Modifications to Bat Algorithm introduced by Kiełkowicz and Grela in [16] are twofold: scheme of acceptance of a new solution, and velocity equation (eq. 1 and eq. 2) is modified to overcome some limitation of an original Bat Algorithm introduced in [23]. Memory of the best solution found during the process of optimization by the algorithm is also introduced.

3. Parameters auto-adaptation using fuzzy logic controller

The dynamic of Modified Bat Algorithm is defined by position and velocity update equations (1) and (2). Pulsation frequency Q_i was chosen to be dynamically adjusted using fuzzy logic Mamdami-Type inference type system since this parameter has most influence on the movement of bats in the flock. Dynamical changes of parameter can improve overall performance of algorithm.

In these paper, fuzzy logic Mamdami-Type inference system is used to control exploration/exploitation phase of a Modified Bat Algorithm thru dynamic modification of a pulsation frequency Q_i, as it was introduced and analyzed in author other publication [15].

Mamdami-Type inference system required to define input values range, define linguistic variable and knowledge base in form of an *if-then* sentence as well as output values range (with their defuzzification method). In these paper as it was introduced in [17] we use diversity of the flock and the iterations number as input parameters. As our output parameter, we choose Q_i. We expect all our input and output parameters to be in range of $[0, 1]$.

The diversity (dispersion) of the flock is defined by following equation (3):

$$diversity(t) = \frac{1}{n}\sum_{i=1}^{n}\sqrt{\sum_{j=1}^{D}\left(x_{ij}(t) - x_j^*(t)\right)^2} \quad (3)$$

Diversity to be considered as input to fuzzy inference system needs to be normalized before, since input must be in $[0, 1]$. Equation (4) was used to normalize diversity:

$$normalizedDiversity(t) = \begin{cases} if\ minDiversity = maxDiversity, 0 \\ if\ \ minDiversity \neq maxDiversity, \dfrac{diversity(t) - minDiversity}{maxDiversity - minDiversity} \end{cases} \quad (4)$$

For iteration to be considered as input to fuzzy logic inference system it needs to be normalized, we used formula (5):

$$Iteration\ = \frac{currentIteration}{maximumNumberOfIteration} \quad (5)$$

Membership functions defining input and output linguistic variables over crisp [0, 1] interval were triangular functions shown in fig. 2 and fig. 3.

Knowledge base for Mamdami-Type inference is in form of a set of an *if-then* sentence, each sentence is constructed using linguistic variables and (possibly) "*and/or*" connectors and hedges. In these paper author considers two linguistic variables diversity and iteration as input to inference system and one output linguistic variable Q. Each variable can take linguistic values from set {small, big} which are shown in fig. 2 and fig. 3. Input and output linguistic values are fuzzy sets defined on interval $[0, 1]$. Hence, we expect crisp input values and output to be in interval $[0, 1]$. Used Knowledge-base is as follow:

$$\text{KnowledgeBase:} = \begin{cases} \text{"}\textbf{If}\ \text{diversity is } small \text{ and iteration is very } small, \textbf{then}\ \text{Q is very } big \\ \textbf{If}\ \text{diversity is } small, \textbf{then}\ \text{Q is } big \\ \textbf{If}\ \text{iteration is } big, \textbf{then}\ \text{Q is } small \\ \textbf{If}\ \text{iteration is very } big, \textbf{then}\ \text{Q is very } small \end{cases}$$

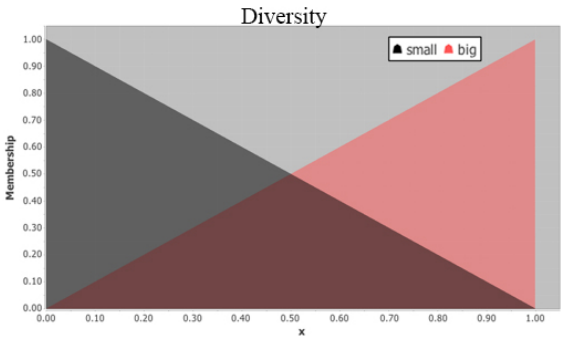


Fig 2. Diversity linguistic variable and terms {small, big}

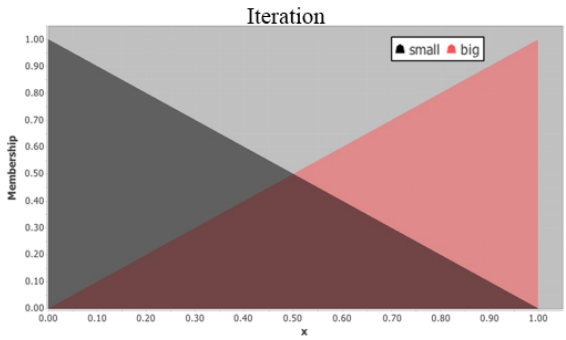


Fig 3. Diversity linguistic variable and terms {small, big}

4. Solution encoding as Bat position in R^n dimensional space

One of the major design consideration in developing unsupervised learning algorithm based on Bat optimization metaheuristic is encoding solution as Bat positing in search space. In the context of clustering N data poins into K clusters with cluster centroids $M = \{m_i\}\ i = 1, \dots, K$ ($m_i \in R^n$ for learning in n dimensional space) one way is for bat position to directly encode the centroid positions as $bat_i := (m_1, \dots, m_K)$ and assign data point to cluster based on similarity measures, eg. Euclidean distance. Which is known in literature, for other types of population based metaheuristics, as centroid-based representation [18]. However, in this paper author augment Bat representation with thresholds and encode *maximum* numbers of centroids (given as input for algorithm) in Bat position. This way optimization process search for both optimum number of centroids as well as optimal centroids position in search space regarding given cluster index (in these paper DB index was used). Solution encoded as Bat position is search space is given as $bat_i := (t_1, .., t_K, m_1, \dots, m_K)$, where threshold $t_i \in [-1,1], m_i \in R^n$. If t_i satisfied $t_i > 0$ corresponding centroid m_i is considered active. All active centroids form valid solution which is evaluated under given clustering index (discussed in section 5).

5. Cost functions

Two cost function are used to evaluate found solution, the CH-index and DB-index. Both are described below.

The CH index [2] achieved the best performance among 30 indices in Milligan and Cooper's comparison study [19]. Given a set of learning objects in form of a matrix X_{mn} with m learning data points $x_i \in R^n$assigned to K clusters $C = \{C_1, \dots, C_K\}$ with centroids m_i the CH index is defined as (6)

$$CH(K) = \frac{Tr(S_B)}{K-1} \Big/ \frac{Tr(S_W)}{N-K} \tag{6}$$

where $Tr(S_B) = \sum_{i=1}^{K} N_i \|m_i - \boldsymbol{m}\|^2$ and $Tr(S_W) = \sum_{i=1}^{K} \sum_{j=1}^{N_i} \|x_j - m_i\|^2$are the traces of the between and within cluster scatter matrices. N_i is the number of learning objects in cluster C_i and $\boldsymbol{m}$ is the total mean vector for the entire data set. Large value of CH(K) suggest a good clustering.

The DB-index [3] attempts to maximize the between cluster distance while minimizing the distance between the cluster centroid and the other data objects. By defining each individual cluster index R_i as the maximum comparison between cluster C_i and other clusters in the partition (7)

$$R_i = \max_{j, j \neq i} \left(\frac{e_i + e_j}{D_{ji}} \right) \tag{7}$$

where $D_{ij} = \|m_i - m_j\|^2$is the distance between the centroid of cluster C_i and C_j and e_i and e_j are the average errors for clusters C_i and C_j, respectively, given as $e_i = \left(\frac{1}{N_i}\right) \sum_{x \in C_i} \|x - m_j\|^2$ the DB index can be written as (8)

$$DB(K) = \frac{1}{K} \sum_{i=1}^{K} R_i \tag{8}$$

Small values of DB(K) suggest good clustering.

6. Latent Semantic analysis using SVD

Before unsupervised learning can take place, learning dataset must be transformed to vector space of reduced dimensionality, the latent space. The transformation is linear and is based on Singular Value Decomposition [12]. SVD factorise given learning data (in form of a matrix X_{mn} with m learning data points $x_i \in R^n$, $X_{mn} = [x_1; \ldots; x_m]$) to three matrices $X_{mn} = U_{mr}\Sigma_{\mathrm{rr}}V_{rn}^T$, where U and V are orthogonal matrices $U^TU = V^TV = I$ and the diagonal matrix Σ contains r singular values of X_{mn}. Approximation of X is computed by setting all by the larges k singular values in Σ to zero (=$\tilde{\Sigma}$). Which is rank k optimal in the sense of L_2-matrix norm. Once obtain the approximation $\tilde{X} = U\tilde{\Sigma}V^T \approx U\Sigma V^T = X$. Notice that the data-to-data inner products based in this approximation are given by $X\tilde{X}^T = U\tilde{\Sigma}^2V^T$ and hence the rows of $U\tilde{\Sigma}$ are defining coordinates for data points in latent space. Once latent representation of a training data set is induced unsupervised learning can take place.

7. Experimental results

For simulation experiments well known, in machine learning community, Iris dataset from UCI Machine Learning Repository was used. Iris dataset consist of three categories i.e., iris setosa, iris versicolor, and iris virginica, with each species having 50 samples represented in terms of four features (sepal length, sepal width, petal length and petal with). Iris setosa can be linearly separated from iris versicolor and iris virginica, but iris versicolor and iris verginica overlap greatly. Number of bats used in simulation was set to 500 and stop criteria was set to 200 iterations. Knowledgebase used in simulations for dynamic parameter adjustments based on current state of the flock eg,: diversity and iteration number, during optimization process using Mamdami-type fuzzy inference system was as follows:

$$\text{KnowledgeBase:} = \begin{cases} \text{"}\textbf{If}\ \text{diversity is } \textit{small} \text{ and iteration is very } \textit{small}, \textbf{then}\ \text{Q is very } \textit{big} \\ \textbf{If}\ \text{diversity is } \textit{small}, \textbf{then}\ \text{Q is } \textit{big} \\ \textbf{If}\ \text{iteration is } \textit{big}, \textbf{then}\ \text{Q is } \textit{small} \\ \textbf{If}\ \text{iteration is very } \textit{big}, \textbf{then}\ \text{Q is very } \textit{small} \end{cases}$$

Figure 4 shows solution found for Iris dataset when learning was performed in induced 1-dimensional latent space and maximum number of clusters was set to 4. Figure 5 shows solution found for Iris dataset when learning was done in induced 2-dimensional latent space and maximum number of clusters was set to 4. Used cluster index was DB-Index.

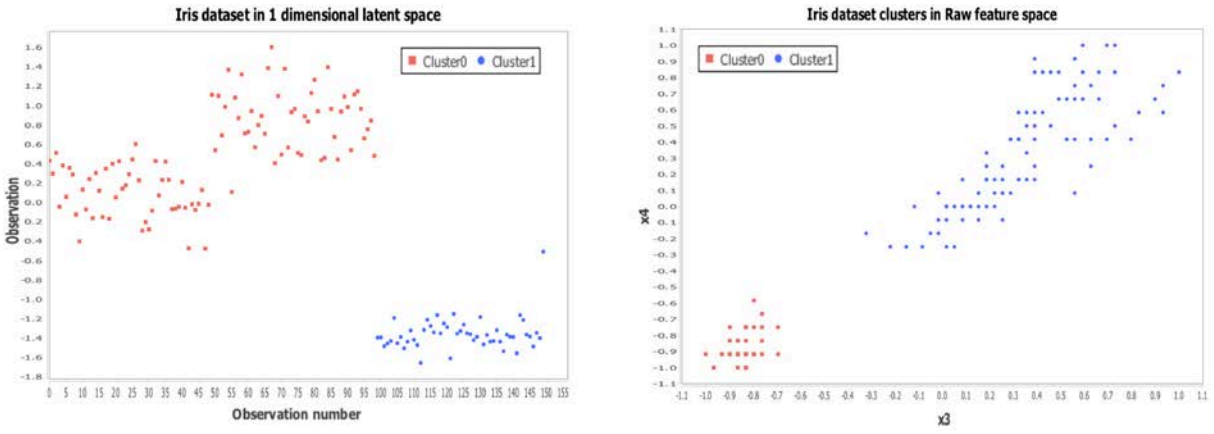


Fig 4. Clusters found in 1-dimensional latent space (left) in Iris dataset and corresponding solution in raw feature space (right). Third and fourth feature were used to depicted dataset. DB-Index was used as cost function.

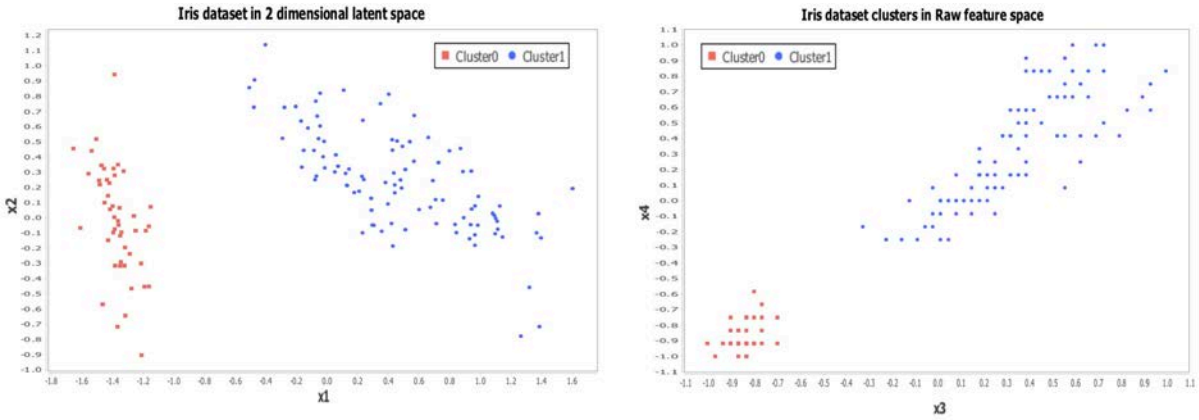


Fig 5. Clusters found in 2-dimensional latent space (left) in Iris dataset and corresponding solution in raw feature space (right). Third and fourth feature were used to depicted dataset. DB-Index was used as cost function.

Figure 6 shows solution found for Iris dataset when learning was performed in induced 1-dimensional latent space and maximum number of clusters was set to 4. Figure 7 shows solution found for Iris dataset when learning was done in induced 2-dimensional latent space and maximum number of clusters was set to 4 for CH-Index

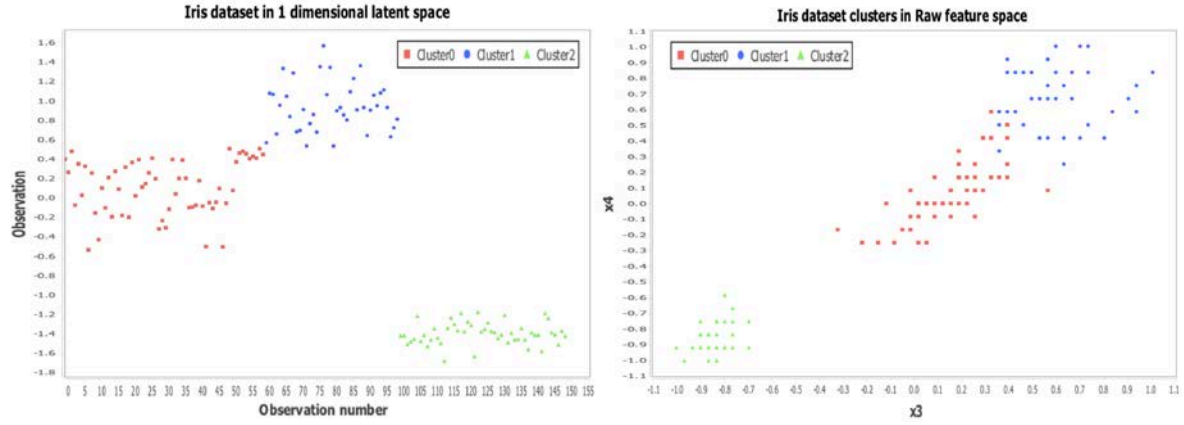


Fig 6. Clusters found in 1-dimensional latent space (left) in Iris dataset and corresponding solution in raw feature space (right) for CH-Index.

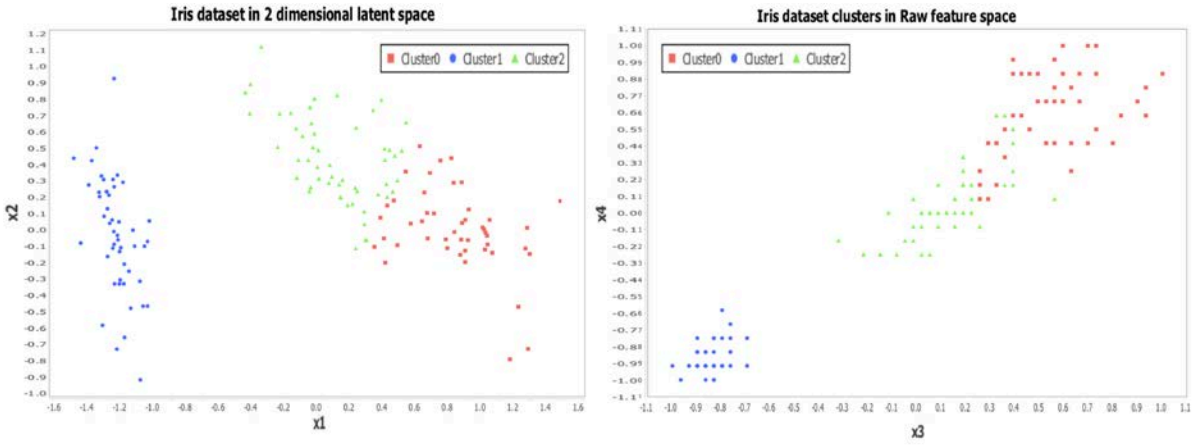


Fig 7. Clusters found in 2-dimensional latent space (left) in Iris dataset and corresponding solution in raw feature space (right) for CH-Index.

Experiments simulation show that is possible to build automatic unsupervised learning algorithm based on modified Bat Algorithm in so-called concept (or semantic) features space. Figure 4-7 (on left) show concept space induced using one and two main concept in dataset. However, the number of cluster found change depending on used cast function. Obtained results indicate that using DB-index will favour less clusters with bigger separation between them. While CH-index seems to be more suited to find cluster of finer granularity within learning set.

8. Summary

This paper introduced unsupervised machine learning method based on modified Bat Algorithm. Learning is performed in so-called latent concept space. Modified Bat Algorithm has automatically adjusted control parameters using fuzzy inference Mandami-type system. Presented algorithm learns the optimal clustering of the data as well as optimal number of clusters in concept space. Latent concept feature space is induced from raw input data-points using linear transformation method based on Singular Value Decomposition. Concept space is of a reduced dimensionality, making it suitable for large datasets and for spare dataset which typically occur in information retrieval systems. A centroid-based solution encoding was introduced as Bat position in multidimensional search space with thresholds allowing to encode *maximum* numbers of centroids (given as input for algorithm) in Bat position. This way optimization process search for both optimum number of clusters as well as optimal centroids position in search space regarding given cluster index. Two cost function were used to evaluate solution CH-index and DB-index. Presented algorithm was tested on well-known Iris dataset, as proof of concepts, in 1 and 2 dimensionally latent concept feature space. Obtained results indicate that proposed method is suitable for preforming unsupervised machine learning in latent space. Performed simulation experiments shows that using DB-index will favour less clusters with bigger separation between them. While CH-index seems to be more suited to find cluster of finer granularity within learning set.

References

1. Baziar A., Kavoosi-Fard A. A., Zare J.: *A Novel Self Adoptive Modification Approach Based on Bat Algorithm for Optimal Management of Renewable MG*, Journal of Intelligent Learning System and Application, vol. 5, issue 1, 2013, pp. 11-18

2. Caliński R. , Harabasz J.: *A dendrite method for cluster analysis*, Commun. Stat., vol. 3, no. 1, pp. 1–27, 197

3. Davies D., Bouldin D.: *A cluster separation measure*, IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-1, no. 2, pp. 224–227

4. Dorigo M., Maziezzo V., Colorni A., *The ant system: optimization by a colony of cooperating ants*, IEEE Trans. on Systems, Man and Cybernetics B, vol. 26, no. 1, 1996, pp. 29–41.

5. Dorigo M., Stützle T.: *Ant Colony Optimization*. MIT Press, 2004

6. Eberhart R. C., Shi Y.: *Empirical Study of Particle Swarm Optimization*, 1999

7. Eberhart R., Shi Y.: *Particle swarm optimization: Developments, applications, and recourses*, in Proc. Congr. Evol. Comput., 2001, pp. 81–86.

8. Fister I. Jr, Fister D., Yang X. S., *A Hybrid Bat Algorithm*, Elektrotehniski Vestnik, 2013, pp. 1-7.

9. Fong S., Yang X. S., Karamanglu M.: *Bat Algorithm for Topology Optimization in Microelectronic Application,* International Conference on Future Generation Communication Technology (FGCT), IEEE, 2012, pp. 150-155.

10. Goldberg D. E.: *Genetic Algorithms in Search, Optimization and Machine Learning*, Kluwer Academic Publishers, 1989.

11. Hansen P., Jaumard B.: *Cluster analysis and mathematical programming*, Math. Program., vol. 79, pp. 191–215, 1997.

12. Hofmann T.: Probabilistic Latent Semantic Analysis

13. Kennedy J., Eberhart R. C.: *Particle swarm optimization*, In Proc. of IEEE International Conference on Neural Networks, vol. 4, 1995, pp. 1942–1948.

14. Kennedy J., Eberhart R., Shi Y.: *Swarm Intelligence*. Academic, 2001.

15. Kiełkowicz K., Grela D.: *FLC control for tuning exploration phase in bio-inspired metaheuristic,* Annales Universitatis Mariae Curie-Sklodowska, sectio AI – Informatica (to appear in 2017)

16. Kiełkowicz K., Grela D.: *Modified Bat Algorithm for Nonlinear Optimization*, International Journal of Computer Science and Network Security (IJCSNS), 2016, pp.

46-50

17. Melin P., Olivas F., Castillo O., Valdez F., Soria J., Valdez M.: *Optimal design of fuzzy classification systems using PSO with dynamic parameter adaptation through fuzzy logic*, Expert Systems with Applications, vol. 40, issue 8, 2013, pp. 3196–3206.

18. Merwe D., Engelbrecht A.: Data clustering using particle swarm optimization, in Proc. Congre. Evol. Comput, vol. 1, pp. 215–220.

19. Milligan G., Cooper M.: *An examination of procedures for determining the number of clusters in a data set*, Psychometrika, vol. 50, no. 2, pp. 159–179

20. Mirjalili S., Mirjalili S. M., Yang Xin-She: *Binary Bat Algorithm*, Neural Computing and Applications, vol. 25, issue 3, 2014, pp. 663-681.

21. Storn R., Price K.: *Differential Evolution - A Simple and Efficient Adaptive Scheme for Global Optimization Over Continuous Spaces*, Technical Report, 1995.

22. Xu R., Xu J., Wunsch C. *A Comparison Study of Validity Indices on Swarm-Intelligence-Based Clustering*, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, vol. 42, NO. 4, 2012

23. Yang X. S.: *A New Metaheuristic Bat-Inspired Algorithm*, Nature Inspired Cooperative Strategies for Optimization, 2010, pp. 65-74.

24. Yang X. S.: *Bat Algorithm for Multi-Objective Optimization*, International Journal of Bio-Inspired Computation, vol. 3, issue 5, 2011, pp. 267-274

Volodymyr MOSOROV[1], Rafał WOJCIECHOWSKI[2], Fatma MBAREK[3]

Analysis and Comparison of Load Balancing Algorithms in Homogeneous Web Cluster

For a business system, websites are considered as an essential part, therefore it is unacceptable for them to be unavailable due to the increasing competition on the global market. To achieve a higher system availability, a greater scalability, it is important to use load balancer in Web farms or Web clusters. The load balancer system is based on several scheduling algorithms. In fact, there are not any overall analysis that demonstrates which load balancing algorithm to use, depending on the characteristic of web cluster. This paper provides a detailed comparison of some HAProxy load balancer algorithms in homogeneous web cluster.

1. Introduction

Further development of information systems emphasized organizations and companies to increase their operations based on the availability of information anytime and anywhere. That is why, it is not acceptable that their web sites are unavailable, as that can detrimental their business revenue and reputation. Thus, websites play crucial part in business aspects.

Therefore, for distributed and parallel systems, such solutions have become more complex and expensive in both architecture and infrastructure and in order to provide constant system availability. A distributed system consists of a set of loosely or tightly computers within a computer cluster. These computers connected together through fast local area network (LAN) and which coordinate their activities to share system resources. So the users access to the system as a single. To realize a higher system availability, it is necessary to take a look at components of the distributed system which provide best results with least amount of investment. Such component offers many advantages and provides scalable growth for web cluster to handle the increase of traffic. This component is load balancing.

Load balancing ensures a high availability and an efficient performance for high traffic distributed systems or computing cluster. The load balancer dispatches the requests to back end servers according to load balancing algorithms or policies. The objective of these algorithms is to minimize the imbalance between servers of web cluster in order to reduce response time or to improve the throughput. The aim of this paper is to discuss a comparison of some dynamic and static HAProxy load balancing algorithms in homogeneous web cluster.

The rest of this paper is organised as follows. Section 2 gives a review of HAProxy techniques. Section 3 describes the simulation scenario, evaluates the experimental results and probes an analytical comparison. Finally, Section 4 provides concluding remarks and future works.

[1] D.Sc. Eng., Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, tel (+48) 42 631 27 50, e-mail: mosorow@kis.p.lodz.pl
[2] Ph.D. Eng., Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, tel (+48) 42 631 27 50, e-mail: r.wojciechowski@kis.p.lodz.pl
[3] Ph.D. student, Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, e-mail: fmbarek@kis.p.lodz.pl

2. HAProxy load balancer

HAProxy or High Availability Proxy load balancer is dedicated for websites and manages the web cluster. It distributes the load among a farm of web servers. It also detects the unavailability of one of the servers. When a web server goes down, it is going out of the server farm until it is able to handle requests [1].

HAProxy works as fault tolerance system. It is being fast, efficient and stable due to its query processing and persistence. The persistence signifies the direction of client's requests to the same server during a session that needs to complete a task. The tested HAProxy algorithms for load balancing include [2][3]:

- Round-Robin (roundrobin) – This algorithm distributes all the requests sequentially among a pool of servers. Using this algorithm, all the servers in the farm are treated equally regard to the weights. Its nature is dynamic, that means the weights of servers can be altered. However, its limited number of active servers is 4095 per backend.

- Static Round-Robin (static-rr) – In this algorithm, all the requests are assigned sequentially around a pool of servers as does Round-Robin. But its nature is static, meaning that it does not allow the dynamic configuration of server's weight. That is why this algorithm is not limited regarding the number of servers, and when the server goes up, it is always immediately reintroduced into the farm.

- Least-Connection (leastconn) – It distributes more requests to servers with lower number of connections. It is recommended to use when the sessions are very long or the connection lengths are varying. But, it is not suited for protocols using short sessions, such as HTTP. This algorithm is ideal for a server group with different capacities. With its dynamic nature, server weights can set during the process.

- Source (source) – It handles out the requests to servers by hashing the source IP address. The hashed IP address is divided by the total weight of running servers in order to determine which server will get the request. The same hash result ensures that the client IP address will be served by the same server. If the hash result changes due a change in the number or weight of running servers, the session will be moved to another server. Generally, this algorithm is used in TCP mode where no cookie may be integrated. It is static by default, which means that the changing of server's weight in the process will possess no effect, however this can be adjusted using "hash-type".

- URI (uri) – With this algorithm, the distribution requests to servers can be by hashing the left part of the URI or the entire URI and divides the hash result by the weight of all running servers to determine which server will get the request. This assures that a same URI will reach the same server. This algorithm is used with anti-virus proxies and proxy caches, to maximize the cache hit rate. It can be only used in HTTP backend. Its nature is static by default, meaning that the changing of server's weight on the fly will possess no effect, but this can be altered using "hash-type". It supports two optional parameters "len" and "depth". These parameters are followed by a positive integer number and can be both helpful to balance servers based on the beginning of URI only. With "len" parameter, the URI algorithm only deems that many characters at the beginning of the URI are used to compute the hash. The "depth" parameter designates the maximum directory depth to be used for computing the hash.

- URL Parameter (url_param) – This algorithm is static by default, distributes requests to servers by looking up a particular URL parameter the query string of each HTTP GET request and performing and performing a hash calculation divided by the weight of all running

servers. If the particular parameter is missed from the URL, the default round-robin algorithm is applied instead.

- Header Name (hdr) – It is static by default, which means that the changing of server's weight in the process will have no effect, however this can be adjusted using "hash-type". The HTTP header <name> will be checked up in each source HTTP request with the equivalent ACL 'hdr()' function. If the header does not have any value or it is absent, the scheduler defaults to round-robin algorithm.

- RDP Cookie (rdp-cookie) – It is a static algorithm by default. The RDP cookie <name> (or "mstshash" if omitted) will be checked up and hashed for every incoming TCP request with the equivalent ACL 'req_rdp_cookie()' function. This method is ideal for persistence mode as it maintains session integrity by sending the same user (or session ID) to the same server. If the cookie is absent, the round-robin scheduling is applied.

3. Experimental evaluation

This section describes the entire test environment that is based on open source tools, services and operating systems. Thus, the testbed environment allows for us a greater availability and reproducibility of further analysis study and research. Then in the rest of section, we present our results with discussion.

3.1. Testbed description

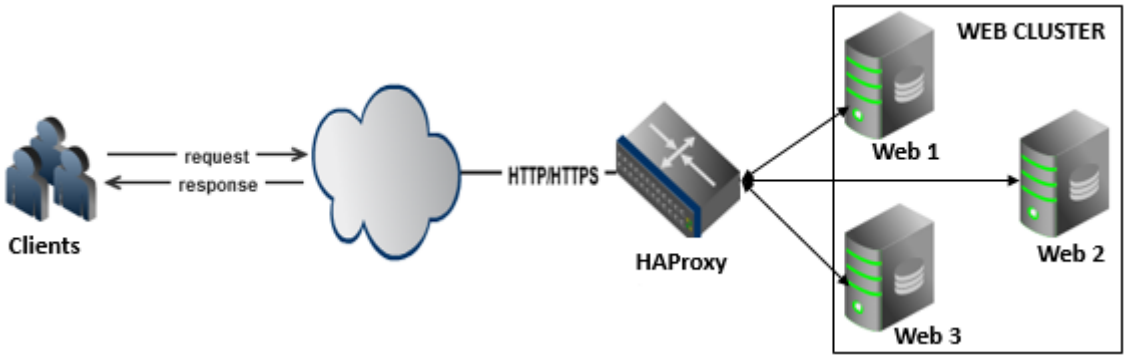


Figure 1: Test environment architecture

As shown Figure 1, our test environment consists of 14 virtual machines. One machine is acting as HAProxy load balancer and 10 others as clients while 3 machines consist one homogenous web cluster with three Apache web servers.

Server specifications:

- Web 1, Web 2 and Web 3: RAM 1GB, local hard drive 25GB.
- HAProxy: RAM 1GB, local hard drive 25GB.
- Clients: RAM 1GB, local hard drive 20GB.

Test environment specifications:

- Physical computer: AMD Athlon II X2 250 and X2 270
- OS: Ubuntu v15.04
- Virtualization platform: Oracle VM VirtualBox

- Load balancer: HAProxy version 1.5
- Scripting language: xml
- Web servers: Apache HTTP server v2.4.10
- Tsung version 1.6

Tested algorithms: roundrobin, static-rr, leastconn and source.

3.2. Experimental results

As we mentioned earlier, the testing was done for homogeneous Apache web cluster under the same scenario which consist of (number of requests in test.xml file and the number of virtual users) in each load balancer algorithm. The simulation results are illustrated in tables 1-5 and figure 2.

Table 1: Network connection statistics

Algorithm	**Connection**		
	Highest rate	*Mean rate*	*Mean*
roundrobin	2620.7/sec	2435.27/sec	1.56 sec
static-rr	2588.5/sec	2459.94/sec	1.51 sec
leastconn	2624.9/sec	2440.28/sec	1.43 sec
source	2856/sec	2470.32/sec	1.48 sec

Table 2: Network request statistics

Algorithm	**Request**		
	Highest rate	*Mean rate*	*Mean*
roundrobin	2609.2/sec	2435.04/sec	2.25 sec
static-rr	2586.1/sec	2459.71/sec	2.19 sec
leastconn	2617.2/sec	2440.07/sec	2.12 sec
source	2856.55/sec	2470.12/sec	2.16 sec

Table 3: Network session statistics

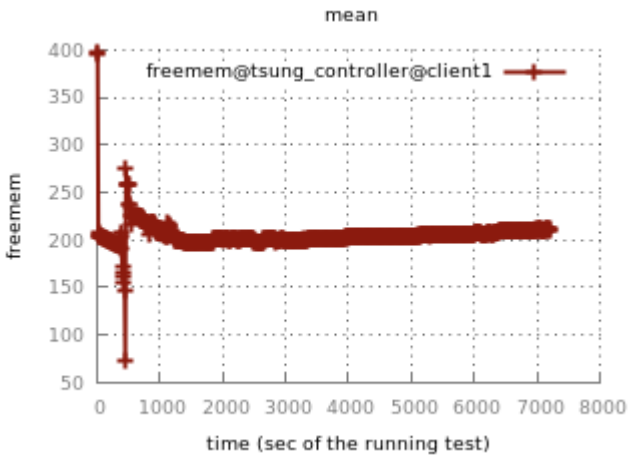
Algorithm	**Session**		
	Highest rate	*Mean rate*	*Count*
roundrobin	8.2/sec	1.85/sec	11045
static-rr	10.3/sec	1.88/sec	10947
l**eastconn**	7.9/sec	1.68/sec	11361
source	8.6/sec	1.78/sec	11255

Table 4: Network throughput statistics

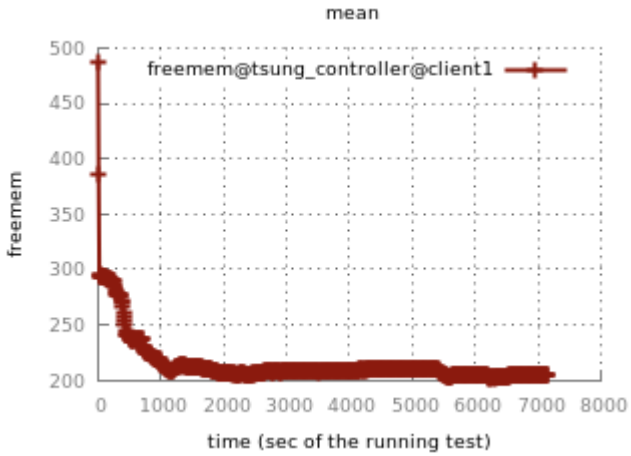
Algorithm	**Highest rate (Mbits/s)**		**Total received packets**	**Total sent packets**
	Packet received size	*Packet sent size*		
roundrobin	9.22	1.28	7.56 GB	1.05 GB
static-rr	9.14	1.26	7.58 GB	1.05 GB
l**eastconn**	9.25	1.28	8.37 GB	1.16 GB
source	10.09	1.39	8.06 GB	1.11 GB

Table 5: Clients behaviour statistics

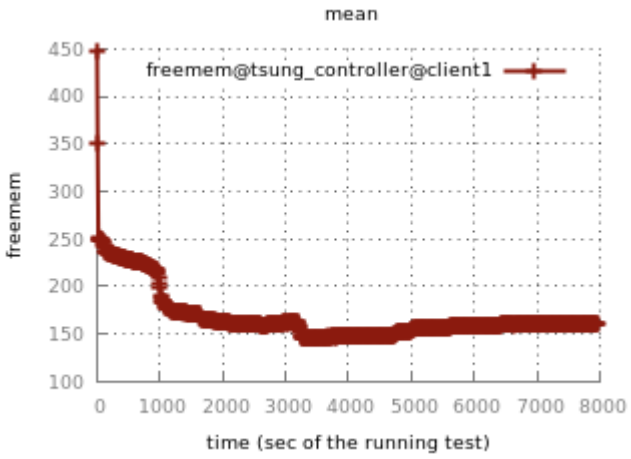
Algorithm	**Simultaneous connected users**	**Simultaneous alive users**	**Number of created users**	**Number of successfully executed requests**
roundrobin	1931	11203	14462	11057
static-rr	1959	11100	14395	10948
leastconn	1972	11202	14499	11363
source	1996	11282	14542	11256



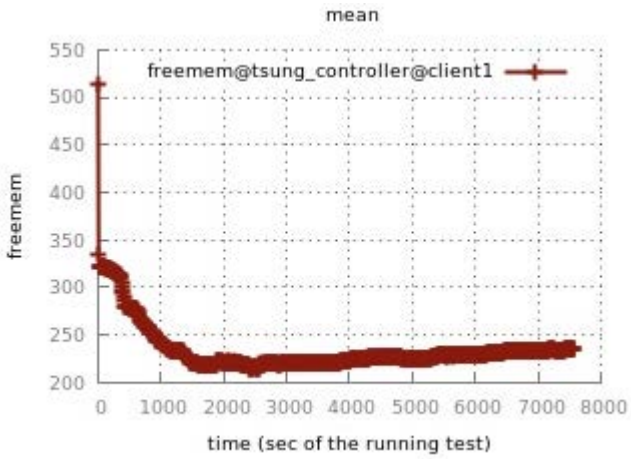
(1) Round-robin



(2) Static Round-robin



(3) Least-Connection



(4) Source

Figure 2: CPU utilization

3.3. Analytical Comparison

In this section, we discuss the comparison of above mentioned results for different load balancing algorithms. Our discussion starts by the high level comparison between load balancing types (static and dynamic) based on various parameters. Then, we prepare the comparison metrics which were explored in subsection experimental results.

A. Static vs Dynamic

In table 6, various comparisons between static and dynamic load balancing techniques are discussed using several comparative parameters [4][5].

Table 6: Static techniques vs Dynamic techniques

Parameters	***Static***	***Dynamic***
Performance	Less	More
Complexity	Less	More
Reliability	Less	More
Flexibility	More	Less
Communication Overhead	Less	More

Resource Utilization	Less	More
Stability	More	Less
Implementation	Easy	Difficult

A. Comparison between HAProxy algorithms

Various comparisons among HAProxy algorithms are listed in the following table 7.

Table 7: Comparison of HAProxy algorithms

Parameters / Algorithms	**Nature**	**Throughput**	**Resource Utilization**	**Number of Connection/sec**	**Number of Request/sec**	**Number of Session/sec**
round-robin	Dynamic	Low	Less	Good	Good	Good
static-rr	Static	Average	Less	Low	Low	Low
leastconn	Dynamic	Good	Less	Good	Good	high
source	Static	High	More	High	High	Good

4. Conclusion

In this paper, we explore a comparative analysis with experimental results. The experiment shows that some HAProxy techniques provide better results than others. When looking at throughput, resource utilization, network connection and network request, the best results are obtained by *source* algorithm, while the poorest results are provided by *static-rr* algorithm. Therefore, this algorithm has much worst results for the heterogeneous web cluster because the overall request are divided equally between the available running servers, where the server specifications are not identical to each one in server cluster. The *round-robin* scheduling and *leastconn* scheduling have also good results.

The aim of research consists to identify the load balancing algorithms which supply the best results in homogeneous web cluster. In the future, we are planning to implement a dynamic algorithm as a load balancing solution based on Ant Colony Optimization technique.

References

1. Kaushal V., Bala A.: *Autonomic Fault Tolerance using HAProxy in Cloud Environment*, INTERNATIONAL JOURNAL OF ADVANCED ENGINEERING SCIENCES AND TECHNOLOGIES, 2011, Vol No. 7, Issue No. 2, 222 – 227.
2. Alagić D., Arbanas K.: *Analysis and comparison of algorithms in advanced Web clusters solutions*, 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), IEEE, 2016.
3. Lakhe S., Shinde A., Sukthankar N., and Reddy C.: *Server load balancing using HAProxy*, JOURNAL OF INFORMATION, KNOWLEDGE AND RESEARCH IN ELECTRONICS AND COMMUNICATION, 2016.
4. Mbarek F., Mosorov V.: *A load balancing system to protect servers against DDoS attacks*, book chapter in *Algorithms, networking and sensing for data processing, mobile computing and applications*, edited by Andrzej Romanowski, Dominik Sankowski and Jan Sikora, Lodz 2016.
5. Beniwal P., Garg A.: *A comparative study of static and dynamic Load Balancing Algorithms*, International Journal of Advance Research in Computer Science and Management Studies, 2014, Vol No. 2, Issue No. 12.

Jarosław FASTOWICZ[1], Krzysztof OKARMA[2]

Application of the Hough Transform for Automatic Quality Evaluation of 3D Printed Surfaces

In this paper the method of automatic quality assessment of the flat 3D prints based on the use of the Hough transform is presented. The idea of the method is based on the assumption of high visibility of relatively large number of long straight lines in the randomly chosen fragments of high quality printed surfaces. Obtained experimental results confirm the possibility of the effective classification of high and low quality 3D prints regardless of the colour of the PLA filament.

1. Introduction

In recent years the layer-by-layer Additive Manufacturing (AM) technology is still growing part of industrial branch. Robust manufacturing of designed prototypes makes a huge step forward in the process of product development. 3D printing is a specific AM method allowing a direct production of a 3D CAD model [7,8,14]. According to [18], four most popular types of the 3D printers can be distinguished: stereolithography (SLA), inkjet 3D printing, selective laser sintering (SLS) and Fused Deposition Modeling (FDM).

Some actual trends in 3D printing such as house printing or possibilities of their use in healthcare are presented in [9,19]. In Straub's paper [16] a wide range of application possibilities of 3D print technology is also discussed.

The authors of some other publications [7,8,14,17] have stated that “the new industrial revolution” [1] generates a lot of problems to solve. The one of the biggest problems is to keep the control of the highest possible quality of 3D printed objects for various applications.

One of the first attempts of using a vision system for inspection of printing process has been made by Fang [5] in 1998. Online detection and identification of the AM defects in fused deposition ceramics have been made by comparison of the original image of 3D model and the image captured by the monitoring system. In the paper [2] the method of nondestructive inspection of the top layer before deposit of the next one has been proposed. For the inspection the boundary and surface of the each layer have been examined. The system is based on the following a “road-path” and a fuzzy model utilizing the correlation between adjacent layers. Another method for continuous inspection of the top layer has been presented by Cho [4]. In this approach a 3D sensing system using a structured light generating device has been proposed with image sensor measuring the distortions of the object printed by selective laser melting method.

An automated system for detecting two types of errors: premature job termination and lack of filament application (so called dry printing), has been presented by Straub [16]. This system is based on five cameras connected to independent Raspberry Pi processing units. Unfortunately, the printing has to be stopped many times for gathering data for vision system.

[1] Ph.D. student, West Pomeranian University of Technology, Szczecin - Faculty of Electrical Engineering, Sikorskiego 37, 70-313 Szczecin, POLAND, tel. +48 91 449 53 05, e-mail: jfastowicz@zut.edu.pl
[2] Assoc. Prof., West Pomeranian University of Technology, Szczecin - Faculty of Electrical Engineering, Sikorskiego 37, 70-313 Szczecin, POLAND, tel. +48 91 449 53 13, e-mail: okarma@zut.edu.pl

Images obtained from the cameras are compared to the reference model using the error thresholding not always leading to satisfactory results. The other problems are related to the calibration of cameras and the compensation of lighting changes. In the recent paper of Straub [17] his improved application having a better performance is presented paying also attention to the cybersecurity issues.

Discussion of using image processing for quality assessment of 3D printed surfaces has been presented in the paper [11] where the usefulness of the Gray-Level Co-occurrence Matrix (GLCM) has been examined. In the other publications the Feature Similarity [12] and Structural Similarity based image quality metrics [13] have been taken for consideration. The physical attributes of 3D prints obtained using the layer-by-layer manufacturing gives a possibility to find and measure each line in image of the scanned sample using the Hough transform which can be used also e.g. for defect detection of various objects [3]. Nevertheless, the application of this algorithm for the wooden furniture elements presented in the paper [3] is limited mainly to the detection of straight edges representing the junction of boards as well as the extraction of flat planes from the cloud of points obtained from the 3D scanner utilizing the fringe patterns approach. This method is usually applied for the extraction of the external shape of the scanned 3D objects without any information related to surfaces, their colors, textures and potential defects [10].

Long regular lines detected by Hough transform on the surface of the 3D print could suggest high quality of the tested sample. The novel approach to quality assessment of 3D prints proposed in our paper can be further utilized as a significant extension of some procedures discussed in [3] allowing the detection of various artifacts also in wooden furniture elements. However, such further modifications are out of the scope of this paper.

2. The idea of the Hough transform and its application for 3D prints

The idea of line detection using the Hough transform is based on the feature extraction approach utilizing the voting procedure applied to pixels representing an imperfect predefined shape. In the simplest case this method can be applied for detection of straight lines utilizing the slope-intercept (or rho-theta) parametrization. One of its most relevant advantages is its robustness to distortions e.g. related to slight deformations of a shape or the presence of small gaps. As the straight line $y = mx + b$ can be represented as a point (b, m) in the parameter space, it can also be defined parametrically utilizing the rho (r) and theta (θ) parameters representing the shortest distance from the line to the origin and the angle between the horizontal axis and the line connecting the origin with the closest point of the line respectively. Two or more points belonging to the same straight line are represented by sinusoids crossing at the (r,θ) point in the Hough space. Applying the voting procedure it is possible to find the local maxima in the Hough space and determine the corresponding combinations of rho and theta defining the detected straight line.

The proposed application of the Hough transform for the surface quality assessment of 3D prints is based on the assumption that for high quality prints the patterns consisting of straight lines representing the consecutive layers of filament should be easily determined. Since the Hough transform is typically applied for binary images, all the images used in our experiments have been converted to greyscale according to popular ITU Recommendation BT.601-7 and then binarized using the Otsu method as well as its adaptive version.

Therefore it is possible to define the quality factor based on the number of detected lines of a specified minimum length allowing a proper classification of high and low quality prints. In the presence of surface distortions there should be some gaps causing the shortening of the detected lines or significantly influencing the detection accuracy. To avoid problems related to

small geometrical distortions of images influencing the detection results for the whole image, the application of the modified Monte Carlo sampling has been proposed allowing a random choice of the specified number of fragments of the samples with the assumption that the whole fragment should fit into the image and consecutive fragments may overlap. The illustrations of the detected lines for an exemplary randomly chosen fragments of high and low quality 3D prints together with the results of Otsu binarization are presented in Figure 1.

As can be easily observed, the number of the detected lines for the marked fragments of both samples differ significantly. However, after some tests conducted for various colors of filaments both for scanned samples and photos, it can be noticed that the simple factor based on the number of detected lines or their total length is not enough for a proper classification.

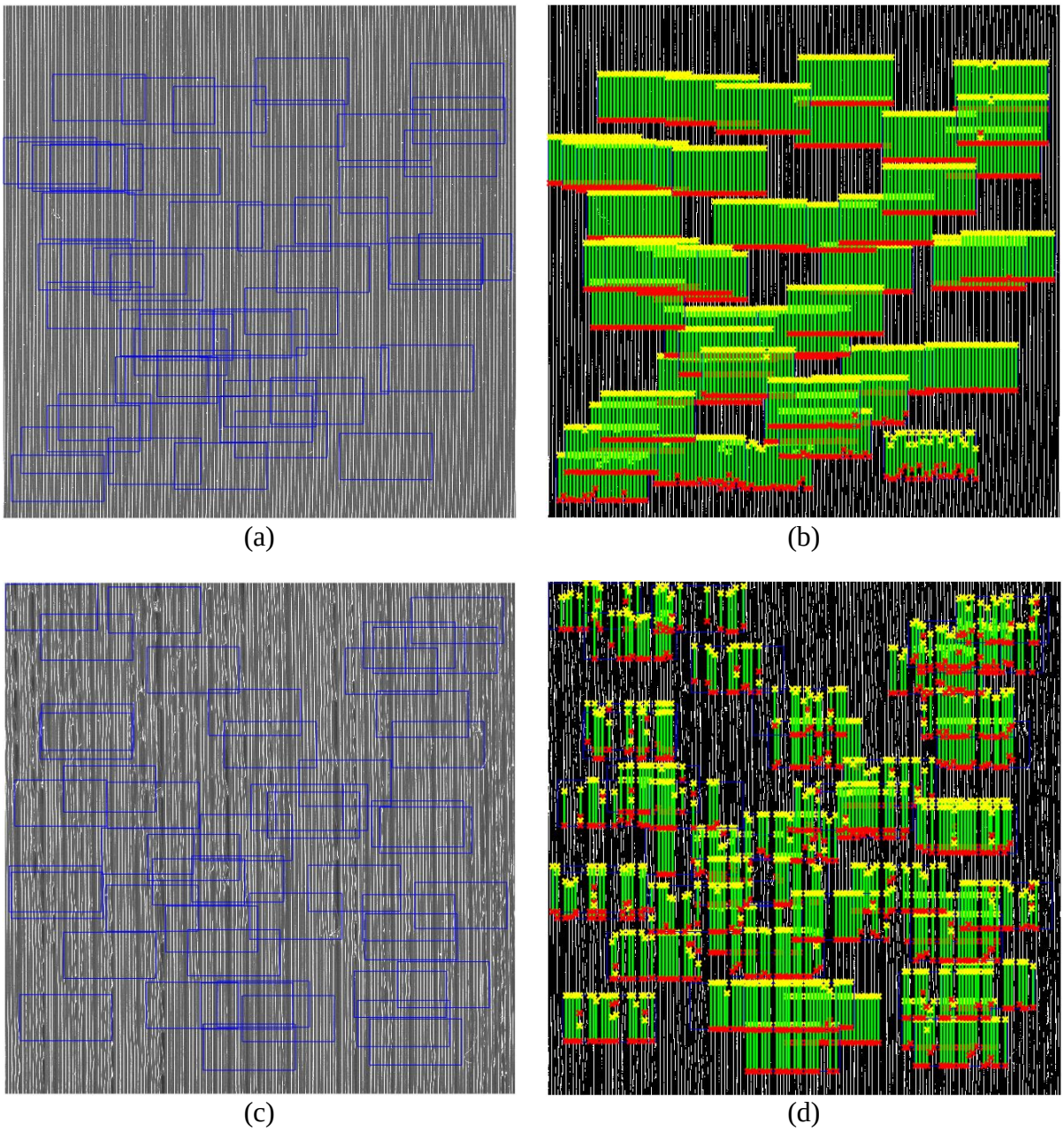


Figure 1: Randomly chosen fragments of high (a) and low (c) quality 3D prints and detected lines with marked starting and ending points detected using the Hough transform shown over the binary images (b and d) (for visualization purposes all images are rotated by 90 degrees).

3. Experimental results

The verification of the proposed approach has been made using a set of prepared 3D prints with the use of two available 3D printers, namely Prusa i3 and RepRap Pro Ormerod 2, using various colours of the PLA filaments. For each filament both high and low quality prints have been prepared forcing the surface distortions by modifications of the speed of the filament's delivery as well as changing the temperature.

The 3D prints have been obtained as small flat plates for the following colours of the filaments (in the order of results in Figure 2): white, red, silver and green. For all colours two first samples do not contain any significant distortions and the next two are lower quality prints. All these samples have been scanned using a 2D flatbed scanner with the 1200 dpi resolution and converted to greyscale and normalized using standard deviation based luminance stretching. Due to the high luminance of the white 3D prints strongly affecting the obtained results, all the randomly chosen fragments of the images with the average luminance over 128 (for 8-bit greyscale) have been converted to negatives. Further binarization required for the Hough line detection have been made using the Otsu method as well as the adaptive thresholding implemented in the OpenCV library. Next, a specified number of image fragments are randomly chosen (on our experiments 50 blocks of 300×150 pixels have been used).

For the classification of the 3D prints into high and low quality samples the following formula is proposed:

$$Q_{Hough} = \frac{avg(H) \cdot size(H)}{avg(L) \cdot |\log(avg(L))|}$$

where *H* is the vector containing the lengths of the detected lines and *avg(L)* denotes the mean luminance of the sample converted to greyscale. The above formula has been invented by analysis of relationships and trends between the length of the detected lines and luminance of samples converted to grayscale. According to conducted experiments, considering the correlation between the variables, the proposed formula has been verified as useful for colour independent classification of the 3D printed surfaces into high and low quality samples.

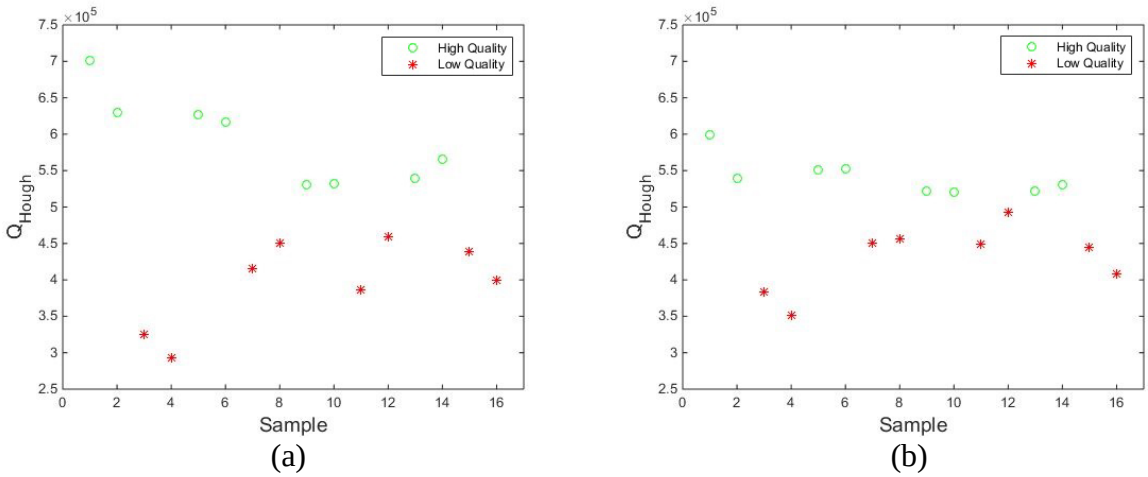


Figure 2: The values of the proposed quality factor Q_{Hough} obtained in experiments for scanned images using global Otsu binarization (a) and adaptive thresholding (b).

The results obtained for the samples used in our experiments are presented in Figure 2 using the mentioned above order of filament's colours. Green circles denote the high quality samples whereas the red stars stand for low quality 3D prints. As can be easily observed in both cases the value of $5 \cdot 10^5$ may be used as the classification criterion.

Considering the fact that in real applications related to online monitoring of the progress of the 3D printing process allowing the correction of some distortions during the printing process, the direct application of the proposed approach for scanned images is at least troublesome, the next step of our experiments is the verification of the method for the images captured by a camera for the same set of 3D prints. Due to different size of obtained images some of the parameters have to be tuned e.g. the size of the analysed window representing the image fragment. Although for the scanned samples, the application of the global Otsu thresholding leads to better discrimination between the high and low quality samples, for the images captured by the camera much better results can be obtained using the adaptive binarization. Due to the presence of some light reflections on the captured images caused by uneven lighting the application of the global Otsu thresholding leads to unsatisfactory results and therefore the adaptive thresholding has been applied by default for photos.

The results obtained for photos are presented in Figure 3 where green circles denote the high quality samples and the red stars stand for low quality 3D prints similarly as previously. Analysing the left plot it can be easily noticed that the application of the simple factor based on the average length of the detected lines does not provide a correct quality estimation and classification of the 3D prints.

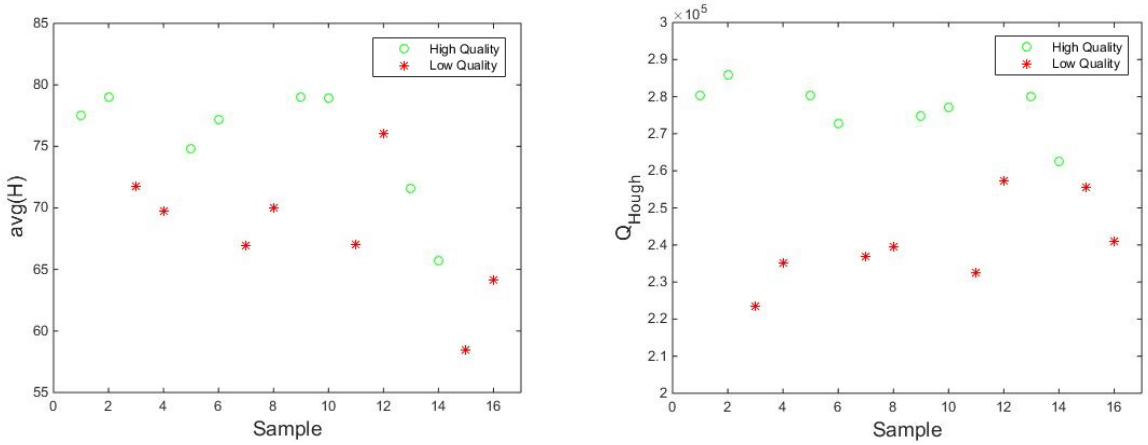


Figure 3: The values of the quality factor obtained in experiments for photos using average length of detected lines (a) and proposed quality factor Q_{Hough} for the adaptive thresholding (b).

4. Remarks and conclusions

In comparison with earlier papers [11-13] it may be emphasized that the proposed method leads to interesting results allowing a reliable estimation of the quality of the 3D printed surfaces and their appropriate classification independently on the colour of the filament. Its application for the images captured by camera may be very useful for online monitoring of the quality of the printed objects allowing immediate corrections during the printing process.

One of the directions of our future work will be related to the application of the proposed approach for the natural images obtained as the photos of the printed 3D surfaces obtained in

varying lightning conditions using more colors of the filaments as well as their different types. Another issue which should be resolved is related to the improvement of the computational efficiency by decreasing the number and size of the analyzed fragments of the image without affecting the classification results.

References

1. Berman B.: *3D printing: the new industrial revolution*, IEEE Engineering Management Review, vol. 41, no. 4, pp. 72–80, December, 2013
2. Cheng Y., Jafari M. A.: *Vision-based online process control in manufacturing applications*, IEEE Trans. Autom. Sci. Eng., vol. 5, no. 1, pp. 140–153, January, 2008
3. Chmielewski L. et al.: *Defect detection in furniture elements with the Hough transform applied to 3D data,* Advances in Intelligent Systems and Computing, vol. 403 (CORES 2015), pp. 631–640, 2016
4. Cho C. H.: *Three-dimensional measurement technology for additive manufacturing*, Proc. Conference on Lasers and Electro-Optics (CLEO), San Jose, pp. 1–2, 2016
5. Fang T., Jafari M. A., Bakhadyrov I., Safari A., Danforth S., Langrana N.: *Online defect detection in layered manufacturing using process signature*, Proc. IEEE Int. Conf. Syst., Man, Cybern., vol. 5 San Diego, pp. 4373–4378, October 1998
6. Ganganath N., Cheng C. T., Fok K. Y., Tse C. K.: *Trajectory planning for 3D printing: A revisit to traveling salesman problem*, Proc. 2nd International Conference on Control, Automation and Robotics (ICCAR), Hong Kong, pp. 287–290, 2016
7. Huang Q., Nouri H., Xu K., Chen Y., Sosina S., Dasgupta T.: *Statistical predictive modeling and compensation of geometric deviations of three-dimensional printed products,* ASME Transactions, J. Manuf. Sci. Eng., vol. 136, no. 6, pp. 061008–061018, 2014
8. Huang T., Wang S., He K.: *Quality control for fused deposition modeling based additive manufacturing: Current research and future trends,* Proc. 1st Int. Conf. on Reliability Systems Engineering (ICRSE), pp 1–6, Beijing, October, 2015
9. Liu T. L., Zahedi S., Garling R. J., Kralickk F., Harris C. A., Cheng M. M. C.: *Prosthetic arachnoid granulations using 3D printing technology*, Proc. 30th IEEE Int. Conf. on Micro Electro Mechanical Systems (MEMS), Las Vegas, pp. 542–545, 2017
10. Okarma K., Grudziński M.: *The 3D scanning system for the machine vision based positioning of workpieces on the CNC machine tools*. Proc. 17th IEEE Int. Conf. Methods and Models in Automation and Robotics (MMAR), Międzyzdroje, pp. 85-90, August 2012
11. Okarma K., Fastowicz J., Tecław M.: *Application of structural similarity based metrics for quality assessment of 3D prints,* Lecture Notes in Computer Science, vol. 9972 (ICCVG 2016), pp. 244–252, 2016
12. Okarma K., Fastowicz J.: *No-reference quality assessment of 3D prints based on the glcm analysis,* Proc. 21st IEEE Int. Conf. Methods and Models in Automation and Robotics (MMAR), Międzyzdroje, pp. 788–793, August 2016

13. Okarma K., Fastowicz J.: *Quality assessment of 3D prints based on feature similarity metrics,* Advances in Intelligent Systems and Computing, vol. 525 (IP&C 2016), pp. 104–111, 2017

14. Sabbaghi A., Huang Q., Dasgupta T.: *Bayesian additive modeling for quality control of 3D printed products,* Proc. IEEE Int. Conf. on Automation Science and Engineering (CASE), Gothenburg, pp. 906–911, 2015

15. Song L., Huang W., Han X., Mazumder J.: *Real-time composition monitoring using support vector regression of laser-induced plasma for laser additive manufacturing*, IEEE Transactions on Industrial Electronics, vol. 64, no. 1, pp. 633–642, January, 2017

16. Straub J.: *Initial work on the characterization of additive manufacturing (3D printing) using software image analysis,* Machines 3(2), pp. 55–71, 2015

17. Straub J.: *Automated testing and quality assurance of 3D printing/3D printed hardware: Assessment for quality assurance and cybersecurity purposes*, Proc. IEEE AUTOTESTCON Conference, Anaheim, pp. 1–5, 2016

18. Tourloukis G., Stoyanov S., Tilford T., Bailey C.: *Data driven approach to quality assessment of 3D printed electronic products*, Proc. 38th Int. Spring Semin. Electronics Technology (ISSE), Eger, Hungary, pp. 300–305, May 2015

19. Zukas V., Zukas J. A.: *An Introduction to 3D Printing*, Design Publishing Inc., Sarasota FL, 2015

Damian FAUSTRYJAK[1], Michał MAJCHROWICZ[2], Lidia JACKOWSKA-STRUMIŁŁO[3]

Web System Supporting Stock Prices Analysis

A web system supporting stock market investors with their decisions is presented. The developed web application gathers informations from the Warsaw Stock Exchange and from internet bank services dedicated for investors. A new approach combining technical analysis with fundamental analysis in order to achieve more effective investment support is proposed. The simulations conducted in virtual environment showed that obtained results are promising. Using the proposed technical analysis tools investors have a 67% chance that the stock price would increase by 7.54% in 25 consecutive quotations. Implementation of fundamental analysis has reduced the number of generated buying signal by 72% and improved the support effectiveness.

1. Introduction

The stock exchange is a place where two very simple transactions occur - buying and selling. These operations can only be performed for financial instruments that have been admitted for trading. Investment professionals over the last several years will share the ways of analyzing the market for two main trends:

- fundamental analysis
- technical analysis

Proponents of fundamental analysis argue that investment quality for financial assets can be classified and monitored. According to this theory, a company's valuation consists of determining its intrinsic value. The next step is to compare the valuation with the current market price. In case when the company priced by the market is below its value (the company is undervalued) then the buy signal will be obtained. On the other hand, if the company is overvalued, for example when the company's valuation on the market is too high, a sales signal will be generated. To determine the strength of a business, the investor must specify the following things: interest rates, industry growth, dividend forecasts and cash turnover [8]. Another method is the collection of various investment techniques developed over the years, which is used by many stock exchange investors, called technical analysis. Contemporary methods that are used by investors allow to analyze trends, charts and even linear formations. Additionally, by using a whole range of mathematical and statistical indicators, it is possible to visualize the market situation, which enables generation of buy and sell signals [10]. While fundamental analysis had a "task" to tell the investor what to buy, technical analysis can, with well-chosen indicators, show when to buy and when to sell the shares. Investor examines the market using chart analysis. People who use this method when analyzing a company are convinced that the price behavior of a company

[1]MSc, Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, tel: (+48) 42 631 27 50, fax (+48) 42 631 27 55, e-mail: damian.faustryjak@p.lodz.pl

[2]Ph.D., Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, tel: (+48) 42 631 27 50, fax (+48) 42 631 27 55, e-mail: mmajchr@iis.p.lodz.pl

[3]Ph.D., D.Sc. Eng., Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, tel: (+48) 42 631 27 50, fax (+48) 42 631 27 55, e-mail: lidia_js@iis.p.lodz.pl

depends on the change in demand and supply. The belief in the effectiveness of this tactic is based on the assumption that the market "knows everything". This means that the investor can investigate market dynamics by observing changes in technical analysis indices. Such a process of examining the condition of a company is based primarily on the finding of repetitive patterns, the formation of which prices are formed. Finding the most common patterns in the current quotes is the basis for forecasting the future [9] [2].

A beginner investor in order to avoid big losses in investing should use the aforementioned tools. A stock exchange game is always risky. Some of stock players rely also on the information extracted from social networks [7]. Nevertheless, experience and well-chosen tactics can allow for regular gains [6].

The aim of the work described in the paper was to develop a system that would support stock market investors with their decisions. Usually investors rely on fundamental analysis or on technical analysis [4][5], depending on their investment strategy. In this paper we propose to combine these both methods in order to achieve better results. To accomplish that technical analysis was applied. As a result number of buy signals has been reduced by the messages from the fundamental analysis.

2. System description

The presented web application was developed with web technologies such as NodeJS, HTML5, JS, and CSS. It was primarily designed to visualize results of technical analysis. These are recommendations that have been prepared by algorithms based on technical analysis. A support for generated buy signals are also messages downloaded from the stock exchange service. The application was divided into two modules. The first module is the graphical user interface. This is where the visualization of the data that was generated by the second module takes place. Reports generated in the application contain information about the current quotes of a particular company, the values of the calculated technical analysis indices, and a panel with stock market notifications. The user interface has also been enhanced with the ability to visualize historical data. The second module performs background work. This is a part of the system, which runs from Monday to Friday at 20:00. With the help of the prepared algorithms, technical analysis of all companies operating on the stock exchange is carried out. The process has been automated in such a way that the user saves as much time as possible.

The application has been divided into two main parts [Figure 1]. Green represents the areas that work as a technical analysis. Blue color was used to mark modules that work in fundamental analysis. The target item is the investor application, which is marked in orange. This is the last point of the program. At this point, the user's application is running, which visualizes the results of the technical and fundamental analysis algorithms.

The data source for the programs that carry out the technical analysis is the *http://gpw.pl/* website. In the tab of historical quotes, the user can retrieve a table from a specific date, which contains the following data:

- Company short name - Each company has its own unique abbreviation. An example for 11 Bit Studio is 11BIT. This is useful information for fundamental analysis programs. Using such a record it is much easier to find messages for a particular company
- International Securities Identification Number (ISIN)
- Currency - Defines the currency in which quotations are made.

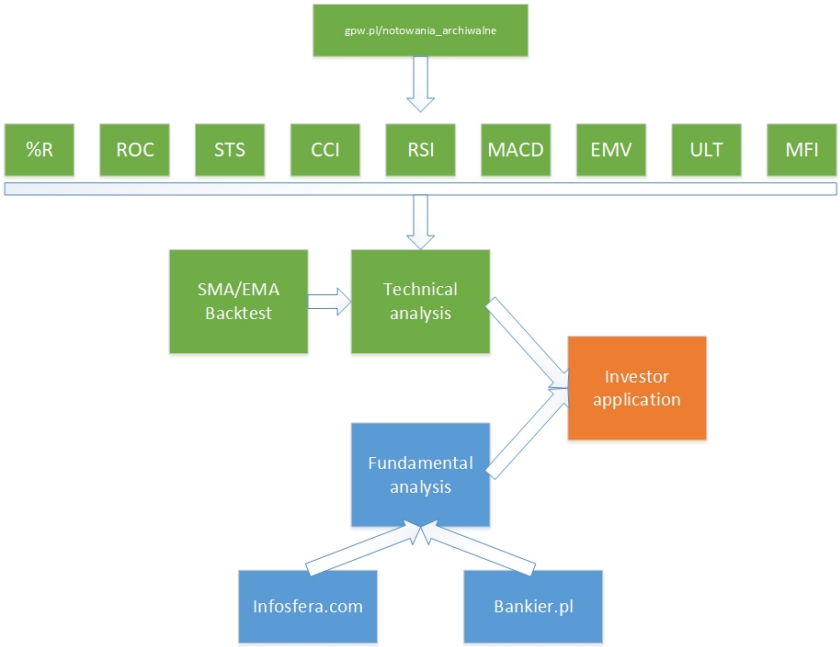


Figure 1: Overall architectural schema of the application.

- Opening price [PLN] - The price of the company that occurred at the beginning of the session.
- Maximum course [PLN] - Maximum price of the company from the session.
- Minimum course [PLN] - Minimum price of the company from the session.
- Closing price [PLN] - The price of the company that occurred at the end of the session.
- Daily change of rate [%] - Value expressed in percentages, which determines the change in exchange rate from the previous day.
- Trading Volume - total number of securities that have changed owner.
- Number of transactions - number of completed sales and purchases.
- Turnover value [in thousands] - turnover is calculated as the product of the price and the number of instruments sold.

The data source for the programs underlying the fundamental analysis are two pages, which contain the latest news about companies. Seeking fields for fundamental analysis algorithms are:

- Message title - This is the heading of the event. Most of the headlines are repeated. In that case it became possible to filter the relevant events for the company. Examples of such events may be: publishing quarterly reports or signing a significant agreement.
- Message Shortcut - Description of the event. It is useful if the investor wants to read the first three sentences of a given event. This gives a general overview of the company

- Date of the event - With the date, it becomes possible to filter news from the newest to the oldest. Only events that are not older than 5 days are selected.

All data was downloaded using *wget*. The software allows for download of a web page or file via Hypertext Transfer Protocol (HTTP), HyperText Transfer Protocol Secure (HTTPS) or FTP (File Transfer Protocol). The program runs under a free software license. Not all servers allow the download of entire site. Services can identify software that is not a browser [1]. Application *wget* allows users to mask the user agent using the `--user-agent` option. Examples of such queries are:

- `wget --user-agent = "Mozilla / 5.0 (X11; U; Linux i686; en-US; rv: 1.9.0.3) Gecko / 2008092416Firefox / 3.0.3" https://www.gpw.pl/ notowania_archiwalne_full?type=10&date=2017-05-12`
- `wget --user-agent = "Mozilla / 5.0 (X11; U; Linux i686; en-US; rv: 1.9.0.3) Gecko / 2008092416Firefox / 3.0.3" http://www.bankier.pl/ investment/ profiles/quote.html?symbol=Ciech`

The first web page is used when downloading data for technical analysis modules. The only parameter that is changed is the `date` parameter. In the example given, the page for the quotations of May 12, 2017 will be downloaded. The fundamental analysis module changes the `symbol` field, which takes the value of the shortened company names. This is the field that was downloaded from the quotation table [11]. All data that has been downloaded using the *wget* tool is a source for further analysis. They are properly converted or filtered depending on the occurrence of certain signals.

In preparing the technical analysis module, the authors used the BRIndicator (Breadth Indicator) methodology. This is a method used by investors who want to evaluate the overall condition of the company. There are several types of activities for this indicator. The data source can be calculated by general force index, Chaikin oscillator, or volume difference. The system has been programmed so that its actions are taken on the basis of previously calculated indicators of technical analysis. Each value of the oscillator is classified according to the specified rules [3]. Each oscillator has a status that describes the company. This value can be defined as a buy or sell signal for a particular instrument. The values of the calculated indices in the system are defined as:

- strong sales signal
- sales signal
- neutral signal
- buy signal
- strong buy signal

Only for certain indicators such as ROC (Rate of Change), FI (Force Index) or EMV (Ease of Movement) value can only be classified as a buy or sell signal. When the system calculates all the values of the oscillators and assigns each signal, the overall value of all the indicators will be calculated. Companies can get value from a defined range, for example strong sales signal

adds only 1 pt. For the indicator BR, while the strong buy signal already has 5 points. Figure 2 illustrates the BR scale. The arrow points to the sum of all points. In this case, many indicators met the criteria for signals that are considered positive.

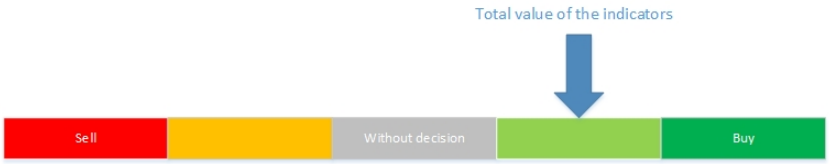


Figure 2: Visualization of the overall condition of the company.

To determine whether a given indicator value can be considered as a sales or purchase signal, two criteria have been used:

- Indicator values must be within a certain range.
- The moving average of the indicator must be greater or less than the current value.

Based on such information, the system classifies the company assets. An example of classification for the stochastic index is as follows:

- Strong buy signal: The only difference compared to buy signal is the criterion that checks if the index value is less than 45.
- Buy signal: the value of the indicator is in the range 25 - 75. The value of the moving average, which was calculated from the 5 previous quotations, is less than the current value of the indicator.
- Neutral signal: The value of the STS indicator and the average of the four previous values are in the range of 45 to 55. The value of the STS indicator is greater than 80 or less than 20. This means the market situation is overdue or sold out.
- Selling signal: the value of the indicator is in the range 25 - 75. The value of the moving average, which was calculated from the 5 previous quotations, is higher than the current value of the indicator.
- Strong Selling Signal: The only difference compared to a sales signal is a criterion that checks if the indicator value is greater than 65.

Classification for the ROC indicator is already much simpler. In this case only the value of the indicator is checked. Such criteria are as follows:

- Buy signal: indicator value is greater than 0
- Sales signal: indicator value is less than 0

The BR indicator values vary from 13 (very strong sell signal) to 65 (very strong buy signal). By using this division, the user application gets the answer in the form of buying or selling advice.

An additional advantage of the implemented system is that it combines two different types of information. Each of them comes from a combination of technical and fundamental analysis. The system detects the information published by the company that serves as an initial buy signal. Then, the technical analysis is used to pre-classify messages that were published. These moments, in which the company published its financial results and which were also recommended by the system as buy signals are shown for two exemplary companies: KGHM in Figure 3 and CD PROJEKT in Figure 4. The buy signals on the graphs are marked by blue points.

Publishing the buy message by a company will not always meet the same reaction of investors. Unfortunately, buying large amount of shares can immediately reverse the upward trend into downward one. This case took place for the KGHM company.

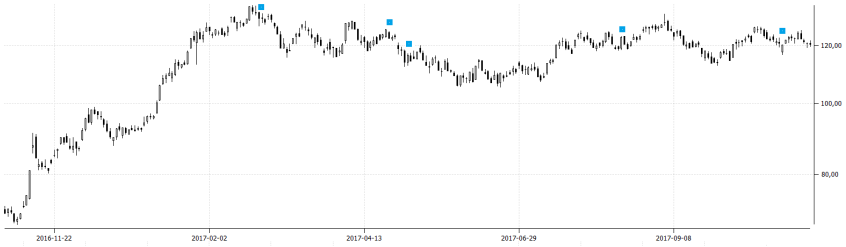


Figure 3: Generated buy signals for KGHM company.

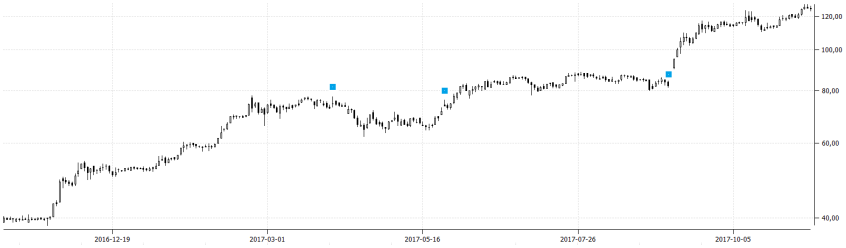


Figure 4: Generated buy signals for CD PROJEKT company.

3. Results of the research

Two methods used by investors have been presented. Using them in the preparation of the BR indicator has produced interesting results. With the BR indicator it's possible to predict whether the company's ratings are in an upward or downward trend. The calculated rolling average values for the BR ratio very well reflect the condition of the company. The simulations have been conducted in virtual environment on the basis of the real data gathered from the Warsaw Stock Exchange (www.gpw.pl) and from internet bank services dedicated for investors (www.bankier.pl). Tests for buy signals have confirmed that the investor has a 58% chance that in the next 7 quotations the company will reach a higher price than in the day, when the signal occurred. Even better results came out when the longer periods were tested. For the 25 quotations, the probability of course growth was 67%. Average growth rate was 7.54%. All back tests were performed on companies whose listing lasts more than 2 years. The BR indicator was the worst performer in the horizontal trend. Values were able to reach minima and maxima in a few days. However, the value of the moving average indicator BR oscillated around the

neutral signal. The biggest problem was checking whether the downloaded news from stock exchange websites were gossips or facts. The user himself has to classify the message and decide whether the company is worth further observation of its condition. However, the fact of retrieving messages and filtering them only to companies in which a sudden extreme change has been occurred, reduced the number of buy signals by 72%.

4. Conclusions

In this paper a web system supporting stock market investors is presented. We proposed to combine technical analysis with fundamental analysis in order to achieve more effective investment support. The conducted simulations showed, that the probability of correct buy signal decision, taken on the basis of technical analysis was about 67% and that the number of buy signals has been reduced by 72%, when the messages from fundamental analysis module were taken into account.

References

1. M. Azat, ”Node.js w praktyce. Tworzenie skalowalnych aplikacji sieciowych”, *Helion*, 2015

2. T. N. Bulkowski, ”Analiza fundamentalna i inwestowanie pozycyjne. Ewolucja gracza giełdowego”, *John Wiley and Sons Inc.*, 2016

3. J. Law, ”A Dictionary of Finance and Banking (5 ed.)”, *Oxford University Press*, 2015

4. Paluch M., Jackowska-Strumiłło L.: "The influence of using fractal analysis in hybrid MLP model for short-term forecast of close prices on Warsaw Stock Exchange", Federated Conf. on Computer Science and Information Systems, 7-10 Sep. 2014, Warsaw, pp. 111-118.

5. Paluch M., Jackowska-Strumiłło L., "Intelligent information system for stock exchange data processing and presentation", 8th International Conference on Human System Interaction (HSI). Warszawa, 25-27 June 2015, pp. 238-243.

6. B. Rockefeller, ”Technical Analysis For Dummies”, *Dummies*, 2012

7. Skuza M., Romanowski A. "Sentiment Analysis of Twitter Data within Big Data Distributed Environment for Stock Prediction", Federated Conference on Computer Science and Information Systems, 2015, pp. 1349–1354.

8. B.-S. Urszula, ”Rynek papierów wartościowych”, *C.H. Beck*, 2016

9. D. Wędzki, ”Analiza wskaźnikowa sprawozdania finansowego według polskiego prawa bilansowego”, *Wolters Kluwer*, 2015

10. Witkowska D., Marcinkiewicz E., "Construction and Evaluation of Trading Systems: Warsaw Index Futures", International Advances in Economic Research, Vol.11, 2005, 83-92.

11. A. Zaremba, ”Giełda. Podstawy inwestowania. Wydanie III zaktualizowane”, *Helion*, 2014

Michał MAJCHROWICZ[1], Paweł KAPUSTA[2], Lidia JACKOWSKA-STRUMIŁŁO[3], Dominik SANKOWSKI[4]

Acceleration of image reconstruction in 3D Electrical Capacitance Tomography in heterogeneous, multi-GPU, multi-node distributed system

Electrical Capacitance Tomography (ECT) is an effective and non-invasive visualization technique, which is used in many industrial applications. Unfortunately, image reconstruction in 3D ECT is a complex computational task requiring operations on large size matrices. In this paper, a new approach to 3D ECT image reconstruction is proposed. A new heterogeneous, multi-GPU, multi-node distributed system has been developed, with a framework for parallel computing and a special plugin dedicated to ECT. The LBP and the Landweber reconstruction algorithms have been implemented in this system. Problems of data distribution and transmission between the computing nodes are discussed in the paper.

1. Introduction

Electrical Capacitance Tomography (ECT) is a relatively new measurement technique that can be used for non-invasive monitoring of industrial processes in 2D [5], 3D [1] and even 4D dynamic mode. ECT is performing the task of imaging of materials with a contrast in dielectric permittivity by measuring capacitance from a set of electrodes placed around the investigated object.

In order to achieve a high quality of 3D image, complex reconstruction algorithms performing many matrix calculations have to be applied [11][12]. Therefore different solutions accelerating these calculation have been reported in the past by the Authors [6][8], especially these dealing with sparse matrices and Finite Elements Method [7] as well as neural networks approach [3][4].

In this work we discuss problems of data distribution and transmission between the computing nodes in a novel heterogeneous, multi-GPU, multi-node distributed system. The system network layer and its efficiency have been tested and compared to the previously developed distributed system based on the Xgrid platform.

[1]Ph.D., Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, tel: (+48) 42 631 27 50, fax (+48) 42 631 27 55, e-mail: mmajchr@iis.p.lodz.pl

[2]Ph.D., Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, tel: (+48) 42 631 27 50, fax (+48) 42 631 27 55, e-mail: pkapust@iis.p.lodz.pl

[3]Ph.D., D.Sc. Eng., Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, tel: (+48) 42 631 27 50, fax (+48) 42 631 27 55, e-mail: lidia_js@iis.p.lodz.pl

[4]Prof., Ph.D., D.Sc., Eng., Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, tel: (+48) 42 631 27 50, fax (+48) 42 631 27 55, e-mail: dsan@iis.p.lodz.pl

1.1. Image reconstruction in ECT

The scheme of image synthesis in Electrical Capacitance Tomography is called image reconstruction. It is based on solving the so called inverse problem, in which the spatial distribution of electric permittivity from the measured values of capacitance C is approximated. We can distinguish two groups of image reconstruction algorithms. Firstly, there are linear algorithms, which, because of higher temporal resolution, are used for monitoring fast-varying industrial process applications, like oil-gas flows in pipelines [1] or gravitational flows and discharging of silo and secondly, non-linear algorithms, which allow reconstructing images with a higher quality.

Image reconstruction using deterministic methods requires execution of a large number of basic operations of linear algebra, such as transposition, multiplication, addition and subtraction. Matrix calculations for a large number of elements is characterized by a high computational load. Moreover, their computational complexity class is polynomial. For example, in the case of multiplication, which in itself is an algorithm of $O(n^3)$ class can be simplified to assume that n-fold increase in the multiplied matrix dimensions will n^3 fold increase the execution time. Matrix multiplication is a key operation in ECT imaging and therefore many researchers decided even to build a custom hardware for this purpose.

In this work, an approach based on computations on GPUs is proposed, as it allows for high flexibility and relatively low application costs. Algorithms of ECT image reconstruction in a distributed multi-node, multi-GPU (Graphics Processing Units) environment were developed and tested. Two image reconstruction algorithms were implemented: the Linear Back Projection (LBP) and the Landweber algorithm. For both of them sensitivity matrices with sizes ranging from 8488x496 to 157264x496 were used.

The LBP algorithm is based on the following equation [2][12]:

$$\varepsilon = \mathbf{S}\boldsymbol{C}_m \tag{1}$$

where:
ε - electric permittivity vector (output image),
$\mathbf{S}$ - sensitivity matrix,
$\boldsymbol{C}_m$ - capacitance measurements vector.

The Landweber algorithm is based on the following iterative equation [2][12]:

$$\varepsilon_{k+1} = \varepsilon_k - \alpha\mathbf{S}^T(\mathbf{S}\varepsilon_k - \boldsymbol{C}_m) \tag{2}$$

where:
ε_{k+1} - image obtained in current iteration,
ε_k - image from the previous iteration,
α - convergence factor (scalar),
$\mathbf{S}^T$ - sensitivity matrix, transposed,
$\mathbf{S}$ - sensitivity matrix,
$\boldsymbol{C}_m$ - capacitance measurements vector.

In the case of the Landweber algorithm each iteration improves the overall quality of the output image. Nevertheless, due to its nature it is necessary to exchange the data (ε_{k+1}) in every iteration. Moreover, it can be assumed that the more iterations are performed the better image quality is obtained. As a result acceleration of image reconstruction process is a very important issue. Nowadays, industrial applications of 3D Electrical Capacitance Tomography

are mainly limited by the two combined factors: image quality and computation complexity of image reconstruction process.

2. Distributed system

Distributed systems are one of the most important technological achievements of recent years, which have had a significant impact on the development of modern computing. The scope of distributed systems applications in everyday life is very wide, from local systems such as cars, ships, aircraft to global systems of millions of nodes used for data processing services; from simple built-in systems consisting of very small and simple sensors to those containing powerful computational components; from built-in systems to those that support advanced interactive user interfaces. We can define a distributed system as a such one in which the hardware components or software on the computers in the network communicate with one another and coordinate their actions only through the transmission of messages. This simple definition covers a whole range of systems in which networked computers can be successfully used to perform tasks faster or more accurately than a single unit. The main motivation for the construction and use of distributed systems is the desire to separate resources [10]. The concept of "resources" that can be effectively shared on a computer network includes both hardware components such as disks and printers, and software such as files, databases, and other objects such as video streams that are transmitted from a digital video camera or a voice signal transmitted between mobile phones and stations.

2.1. Design assumptions

As a result of the earlier performed studies [6][9] the Authors have developed a new distributed system dedicated to ECT computations. The system is specially designed to accelerate matrix computations that are a crucial part of reconstruction algorithms used in ECT [7][9]. The architecture of the designed system is shown in Figure 1. The earlier developed solution was based on the Xgrid platform, used as a network layer. However, the analysis of this system showed the limitations of this solution, and the main conclusion from the previous research was, that the new software for the system should be developed. A special framework was designed and built that provides software tools needed both for the system architecture expansion and new algorithms development and implementation in a distributed heterogeneous environment. The proposed approach allows for a greater flexibility of the developed solutions, provides tools for their easy testing and enables further acceleration of ECT image reconstruction.

The framework was designed to ensure an efficient use of the computing power of all the devices in which the nodes are equipped. This architecture is scalable and allows users to expand the computing power of the system by adding more nodes. The above assumptions pose many challenges in the architecture of the system itself, but their application makes it straightforward to use the environment to speed up computations in existing projects, thus testing and developing new distributed algorithms is much faster. Expansion of the computing power of the system is possible through the use of "plug-in" architecture (by adding support for new devices, such as FPGAs). The basic operations of linear algebra were implemented in the system as a set of functions in the form of an API (Application Programming Interface). The use of a heterogeneous system for distributed computing in ECT required the implementation of a series of algorithms without which the proposed system would not work properly. The most important of these are as follows:

- Division of matrices between nodes
- Basic operations of linear algebra (transposition, addition, subtraction, multiplication)
- Data transfer between nodes
- Planning and division of tasks
- Support for heterogeneous devices
- Support for calculations using graphics cards
- Supports modern multi-core processors as a set of devices
- Possibility to extend existing solutions with pseudo inheritance from implemented layouts

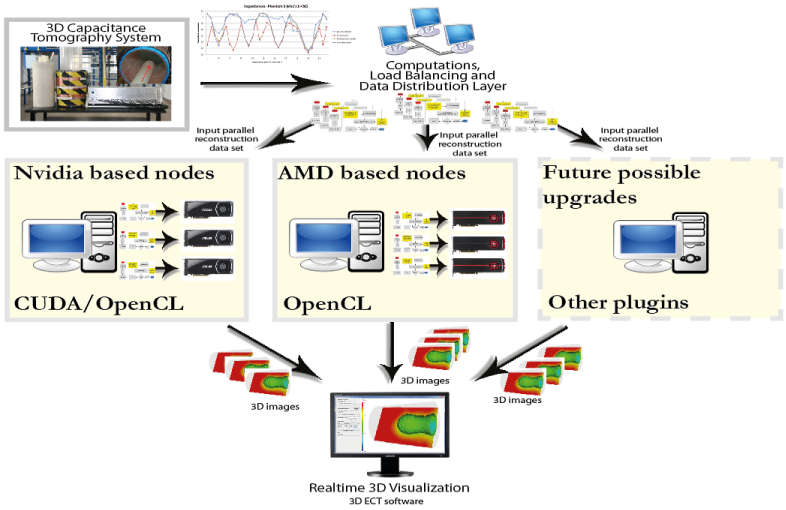


Figure 1: Schematic of a designed heterogeneous distributed computing system comprising graphics cards

3. Test results

In the design of a distributed system, problems of data buffering, division and distribution must be considered. By creating a network layer the existing solutions and networks protocols can be used or the own ones can be developed. In the previously built distributed system a ready Xgrid platform was applied as a network layer [6]. In the current work the author's system named KISDC (KIS Digital Computing) with network layer (NET) was designed and implemented. While designing the KISDC-NET layer, the existing network protocols were applied and tested in advance in order to choose the best solution. The network characteristics of the previously developed solution based on the Xgrid system was compared with the new system using other data distribution protocols: HTTP (Hypertext Transfer Protocol), FTP (File Transfer Protocol) and SMB (Server Message Block).

Based on the obtained results (see Fig. 2) it can be stated that the use of each of the tested protocols ensures a shorter data transfer time than the Xgrid-based solution. Although the results obtained with the FTP protocol are very good, the best results have been achieved for the HTTP protocol. Therefore, HTTP protocol has been selected as the best one for the KISDC-NET. Both distributed systems: the one based on the Xgrid platform and the KISDC have been extensively tested and compared. In both cases the hardware was identical, consisting of two nodes of high computing power, both using 8 thread Intel i7 930 CPUs and Nvidia GPUs (Tesla S1070 + Tesla C2070 compute devices in the first node and dual GTX 570 in the second).

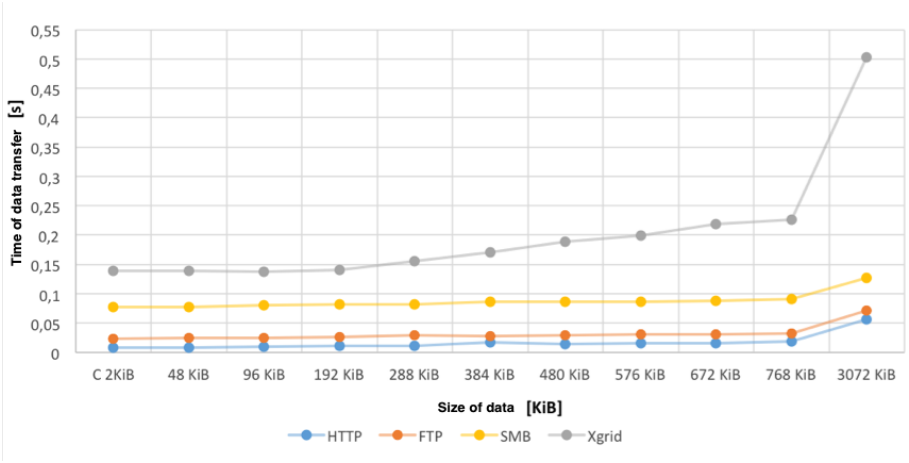


Figure 2: Comparison of average data transmission times for different network protocols

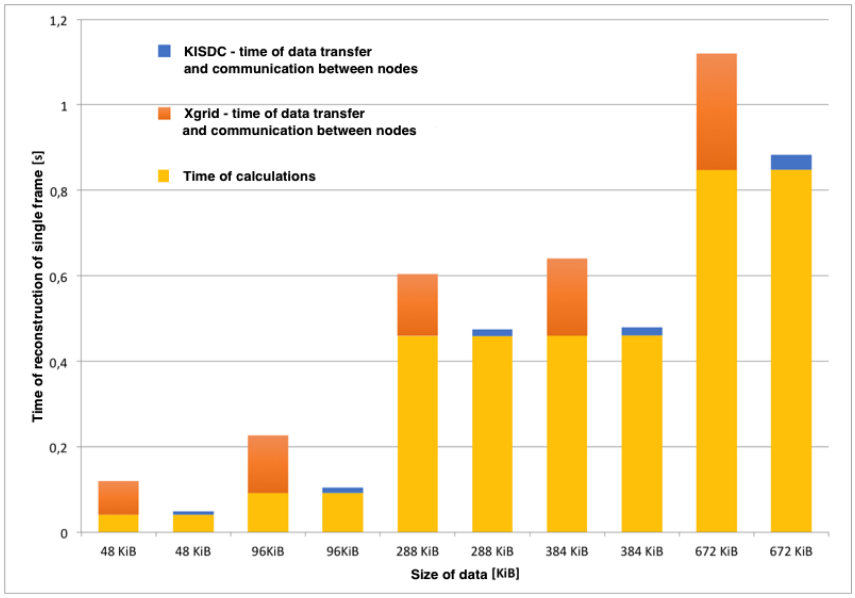


Figure 3: Comparison of image reconstruction time in Xgrid and KISDC systems for dual computer configuration

The comparison of times of a single frame reconstruction in the two node system are shown in Figure 3. Yellow color represents calculation time (the same for the both systems), blue color is related to data transfer time for the KISDC system, and orange color denotes data transfer time for the Xgrid system. For each of the analyzed data sizes, the speed up of image reconstruction expressed in the number of reconstructed frames per second was noted. The most significant acceleration was achieved for 48 KiB and 96 KiB image vectors.

4. Conclusions

A flexible, distributed computing system for tomographic image reconstruction called KISDC has been designed and developed. The work described in this paper was focused on its network layer, i.e. on improvement of data management in the distributed system and on reducing delays in the data transmission over the computer network. The comparison of times of data transfer and communication between the nodes shows very clearly that the use of the new developed system with HTTP protocol ensures much better results than with the Xgrid platform. It is also evident that the KISDC system allowed for a significant reduction of the total time of a single frame reconstruction and a major speed up in implementations of both the LBP and the Landweber algorithms.

References

1. Banasiak, R., Wajman, R., Fidos, H., Jaworski, T., Fiderek, P., Kapusta, P., Majchrowicz, M., Sankowski, D., “Fusion of three-dimensional electrical capacitance tomography and fuzzy logic inference for phases fraction and flow structures identification in horizontal and vertical gas-liquid flow pipelines.,” *7th World Congress on Industrial Process Tomography,* Kraków, 2013, pp. 818–827.
2. Cui, Z., Wang, Q., Xue, Q., Fan, W., Zhang, L., Cao, Z., Sun, B., Wang, H., Yang, W., “A review on image reconstruction algorithms for electrical capacitance/resistance tomography,” *Sensor Review*, Vol. 36 Issue: 4, 2016, pp. 429-445
3. Garbaa, H., Jackowska-Strumillo, L., Grudzien, K., Romanowski, A.: "Funnel Flow Dynamics Prediction based on Electrical Capacitance Tomography and Artificial Neural Networks", *In Proc. International Interdisciplinary PhD Workshop 2016*, Brno, 12-15 Sep. 2016, pp. 58-62.
4. Garbaa, H., Jackowska-Strumillo, L., Grudzien, K., Romanowski, A., "Application of electrical capacitance tomography and artificial neural networks to rapid estimation of cylindrical shape parameters of industrial flow structure", *Archives of Electrical Engineering*, Vol. 65(4), 2016, pp. 657-669.
5. Grudzień K., Chaniecki Z., Romanowski A., Niedostatkiewicz, M., Sankowski, D. “ECT Image Analysis Methods for Shear Zone Measurements during Silo Discharging Process,” *Chinese Journal Of Chemical Engineering*, Vol. 20 (2), 2012, pp. 337-345
6. Kapusta, P., Majchrowicz, M., Sankowski, D., Jackowska-Strumiłło, L., Banasiak, R., “Distributed multi-node, multi-GPU, heterogeneous system for 3D image reconstruction in Electrical Capacitance Tomography - network performance and application analysis,” *Przegląd Elektrotechniczny, 89 (2 B)*, 2013, pp. 339-342.
7. Kapusta, P., Majchrowicz, M., Sankowski, D., Jackowska-Strumiłło, L., “Acceleration of image reconstruction in 3D Electrical Capacitance Tomography in heterogeneous, multi-GPU system using sparse matrix computations and Finite Element Method,” *Proc. of the 2016 Federated Conf. on Comp. Science and Inf. Systems*, ACSIS, Vol. 8, 2016, pp. 679-683
8. Majchrowicz, M., Kapusta, P., Banasiak, R. , “Applying parallel and distributed computing for image reconstruction in 3D Electrical Capacitance Tomography,” *Zeszyty Naukowe AGH - Automatyka, Vol 14, Issue 3/2,* 2010, Kraków, Wydawnictwa AGH, pp. 711–722.
9. Majchrowicz, M., Kapusta, P., Wąs, Ł., Wiak, S, “Application of General-Purpose Computing on Graphics Processing Units for Acceleration of Basic Linear Algebra Operations and Principal Components Analysis Method,” *Man-Machine Interactions 3, Advances in Intelligent Systems and Computing Volume 242,* Springer International Publishing, 2014, pp. 519–527.
10. Nawrocki W.: LAN-Based Measurement Systems, In: Nawrocki, W.: *Measurement Systems and Sensors, 2nd edition*, Artech House Remote Sensing Library, pp. 339-358, 2016.
11. Rymarczyk, T., Filipowicz, S. F., Sikora, J.: "Image reconstruction with discontinuous coefficient in electrical impedance tomography", *Przeglad Elektrotechniczny*, Vol. 87(5), 2011, pp.149-151
12. Yang, W.Q., Peng, L. “Review of image reconstruction algorithms for electrical capacitance tomography, Part 1: Principles.,” *International Symposium on Process Tomography in Poland,* Wrocław, 2002a, pp. 123–132.

Abdulla JUWAIED[1], Lidia JACKOWSKA-STRUMIŁŁO[2], Artur SIERSZEŃ[3]

Improved clustering algorithm of LEACH Protocol in IoT Network

The Wireless Sensor Network (WSN) consist of a large number of sensors or nodes that are limited in energy, storage and processing. The energy of nodes in Wireless Sensor Network (WSN) is the most important thing, because it is limited by the energy of the node. Therefore, extend WSN lifetime is a significant problem. Main assumption of LEACH (Low-Energy Adaptive Clustering Hierarchy) protocol is to decrease system delay and reduce energy consumption. Application of the LEACH protocol can also extend the lifetime of Wireless Sensor Network or connected devices. In this paper, we propose an improvement of the LEACH routing protocol, which relies on choosing a cluster head node in a middle of the cluster area on the basis of the location of all nodes in the cluster and its minimum distance to the other nodes in the same cluster.

1. Introduction

One of the recently used network technologies in this century is WSN and one of the recent technology is sensing application which is used for the growth of devices with less cost and power. Wireless Sensor Network (WSN) is a collection of large number of sensors, also called nodes which are capable of sensing the environment, have communication capabilities and are able to compute the information. The information is processed locally and is sent to the point of collection using wireless links. These sensors can communicate directly with themselves or they are deployed over large area with one or more than Base Station (BS). The Base Stations have more recourses and capabilities than the nodes. The most important goal in Wireless Sensor Network (WSN) is to carry out of a data communication in this way to extend the lifetime of the network [1].

The use and evolution of WSNs depends on two main points: security problems and routing strategies which are a great research challenge in WSNs. There are many routing protocols dedicated to WSN. The important one are routing protocols known as hierarchal protocols, for example SPIN, PEGASIS [2], and LEACH [3]. In hierarchal routing protocols some nodes have a higher priority than the others for example, some nodes will divide into clusters and these clusters will choose a cluster head. These cluster head are responsible to send and receive data between all nodes in the clusters [2]. In this paper a problem of choosing the cluster head is considered. In the LEACH protocol, cluster heads are randomly taken. Some cluster heads are near to the centre of the cluster in other hand some cluster heads are in the boundaries of the cluster [5] [6]. This feature of the LEACH protocol leads to higher energy consumption and it also impacts on the performance and security of the entire network.

[1] Ph.D. student, Institute of Applied Computer Science, Lodz University of Technology, Lodz, Poland, e-mail:: ajuwaied@kis.p.lodz.pl
[2] Ph.D., D.Sc., Institute of Applied Computer Science, Lodz University of Technology, Lodz, Poland, e-mail: lidia_js@kis.p.lodz.pl
[3] Ph.D., Institute of Applied Computer Science, Lodz University of Technology, Lodz, Poland, e-mail: asiersz@iis.p.lodz.pl

Therefore, the cluster heads which are in a middle distance to the other nodes will be chosen as the cluster head. This makes the protocol more secure and energy consumption is reduced.

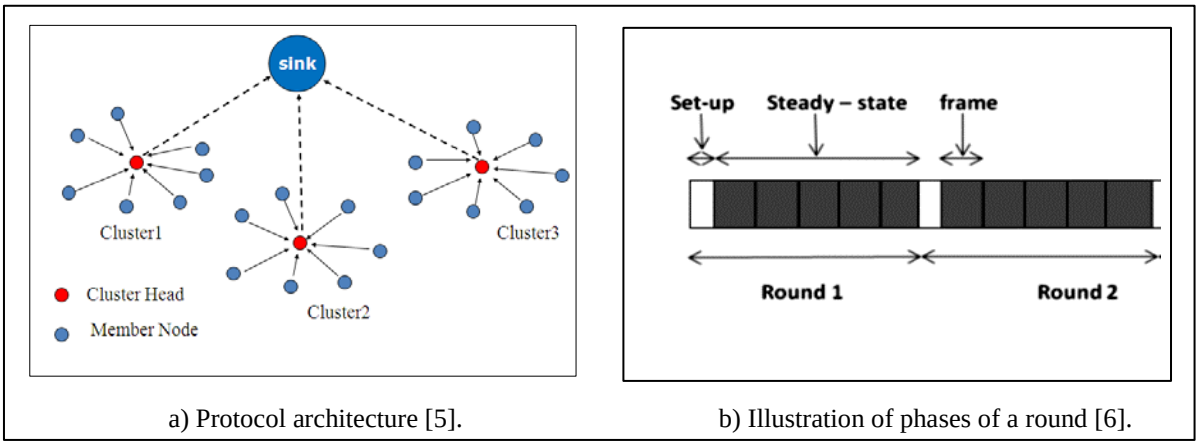
2. Internet of Things

Internet of Things (IoT) is a concept and a model that considers pervasive in the environment of a set of things or objects that through wireless and wired connections with unique addressing schemes are able to interact with each other and cooperate with other things to create new applications or services to reach goals [9]. The general idea of IoT is to connect objects to each other and also to the Internet, many properties made IoT one of the most popular technology in the field of computer science such as: advances in data analytics, computing economics, rise of cloud computing, miniaturization and ubiquitous connectivity.

3. LEACH Protocol

Low Energy Adaptive Clustering Hierarchy (LEACH) is one of major routing protocols for Wireless Sensor Network. LEACH protocol is a TDMA (Time Division Multiple Access) based MAC (Media access control) protocol. The main idea of this protocol is to expand the energy or battery lifetime of WSN through a distribution of the energy load among the sensors or nods in the network, as shown in figure 1.a. The principle aim of LEACH protocol is to order all sensors in the network into small clusters to choose one of them as cluster head, after that the cluster head will send all information received from the sensors to the base. Therefore, LEACH divides network operation into several rounds, and every round contain two stages [4] *the Set up phase* and *the Steady phase* (see Fig. 1.b).

The selection of cluster heads is random in this protocol and it does not take into account the location of the other nodes, so some cluster heads are near to the center of the cluster and the other in edge of the cluster [5, 6]. This feature of the protocol leads to higher energy consumption and it has a great impact on the performance of all sensors in the network.



a) Protocol architecture [5]. b) Illustration of phases of a round [6].

Fig. 1: LEACH protocol architecture (a) and phases (b).

3. LEACH Protocol Phases

As was mentioned previously the operation of LEACH protocol consist of several rounds where each round is divided into two phases.

$$T(n) = \begin{cases} \dfrac{P}{1 - P[r * \bmod(1/P)]} & if \quad n \in G \\ 0 & otherwise, \end{cases} \qquad (1)$$

Set up phase: in this phase nodes generate randomly numbers between 0 and 1. These nodes in the cluster are chosen as cluster heads according to the value of threshold *T(n).* The threshold *T(n)* is calculated using the equation (1) [4] [10]:

In the equation (1) *n* is a given node, *r* is the current number of rounds, *P* is the percentage of cluster nodes accounting in the total number of sensor nodes, *G* is the set of nodes that didn't become cluster heads in the *1/P* round and *mod* is the number of nodes which have been elected as cluster heads in the round *r.*

If the random number is less than threshold *T(n),* the node becomes the cluster head of the current round *r*, then common modes will join the nearest cluster. After the node is selected as a cluster head, all nodes join the corresponding cluster depending on cluster head broadcasting an advertisement message to all other nodes. The last step in this phase is cluster heads create TDMA schedule according to the number of nodes in the cluster [6]. TDMA schedule is used to switch the remaining nodes into sleep mode and also to prevent the collision of data sent by the nodes [7].

Steady state phase: the nodes in each cluster communicate only with the cluster head through a single hop transmission, then the cluster head will communicate with the base station. When the node sends data to the cluster head in the allocated time, another nodes of the cluster remain in the sleep mode, because TDMA is used as it was mentioned previously [7]. The data transfer from the cluster head and the base station is done also with the help of TDMA schedule [6]. After a certain time, which is determined a priori, the network again goes back into the setup up phase and new cluster-heads are chosen.

Each cluster communicates using different CDMA (Code Division Multiple Access) codes in order to reduce interference from nodes belonging to other clusters [8]. Time line of LEACH operation: the set-up and steady state phase is shown in figure 1.b.

4. LEACH algorithm

In this part an original algorithm of the LEACH protocol is explained, which consist of the following steps:

1. *CN=> r // check candidate node with round*
2. *If r > T(n) then, CH = CN else, goto step1 //node can be considered for cluster head*
3. *reset all nodes to non-CH*
4. *CH => G: id(CH), join adv*
5. *A(i) -> CH(j) : id(A(i)) , id(CH(j)) , join req*
6. *CH(j)-> A(i) : id(CH(j)) , < t(i) , id(A(i)) >*
7. *CH(j) = TRUE;// node j is a candidate for cluster head*
8. *CH(j) = FALSE; //node j is not a candidate for cluster head*
9. *A(i) -> CH(j) : id(A(i)) , id(CH(j)) , info*
10. *CH -> BS : id(CH) , id(BS) , aggr info*

The symbols used in the LEACH algorithm are as follows: *CN* − candidate node to become the cluster head, *x* − random variable (0> r> 1), *T (n)* − threshold value, *CH* − cluster head, *G* − all nodes in the network, *Id* − identification number, *A* − normal node, *Join_adv* − request to join the cluster, *T* − time-slot to send the sensed data, => means broadcast and -> means unicast.

5. LEACH algorithm modification

As it was mentioned in the section 3, the LEACH protocol uses a threshold *T(n)* depending on the number of nodes *n*. The sensor performance is very sensitive to the value of *n* in the network, therefore when the value of *n* is large, then many nodes will communicate with the cluster head and large distances between them will result in high loss of energy, because the nodes and cluster heads dissipate energy while transmitting the data.

Therefore, to save the total residual energy of the nodes in the cluster and extend a lifetime of the network, we propose to choose the cluster head in the network depending on a location of all nodes in the same cluster and a distances between them. The modification of choosing the cluster head in cluster area is done in three main steps:

Step 1: show the position of all nodes in the cluster depending on the coordinates x and y. The result will show a 2-dimension matrix containing the location of all nodes in the network.

```
For j=1:size(clusterData)
clusterData{j}.clusterNodes = [];
end
```

Step 2: calculate the distances between the nodes to choose the cluster head (near to middle- point comparing to the other nodes), then the nodes will join this cluster head which is the nearest to the middle position in the cluster area.

```
for j=1:size(S,2)
    x = S(j).xd;
    y = S(j).yd;
    minDist = Inf;
    H = 0;
    for p=1:size(X,2)
        dist = sqrt((x-X(p))^2+(y-Y(p))^2);
        if dist<minDist
            minDist = dist;
            H = p;
        end
    end
if H>0
        clusterData{H}.clusterNodes = ...
           [clusterData{H}.clusterNodes; [x y] ];
    end
```

Step 3: After the middle distance had been calculated in the previous step and the nodes have been ordered to the cluster head had depending on their location, the network is divided into regions based on a distances to points in a specific subset of the network, what can be done using Voronoi diagram as shown in figure 2.b.

6. Results

The simulation was conducted in MATLAB. First a set of random nodes and cluster heads in the network was generated using LEACH protocol, as shown in figure 2.a. Cluster heads are marked as dark stars. Then the modified algorithm was implemented and the resulting clusters and cluster heads positions are shown in figure 2.b. There are 9 cluster areas, each one consists of 1 cluster head and set of nodes. Most cluster heads in the figure 2.b are near to the middle point of the cluster comparing to other nodes in the same cluster.

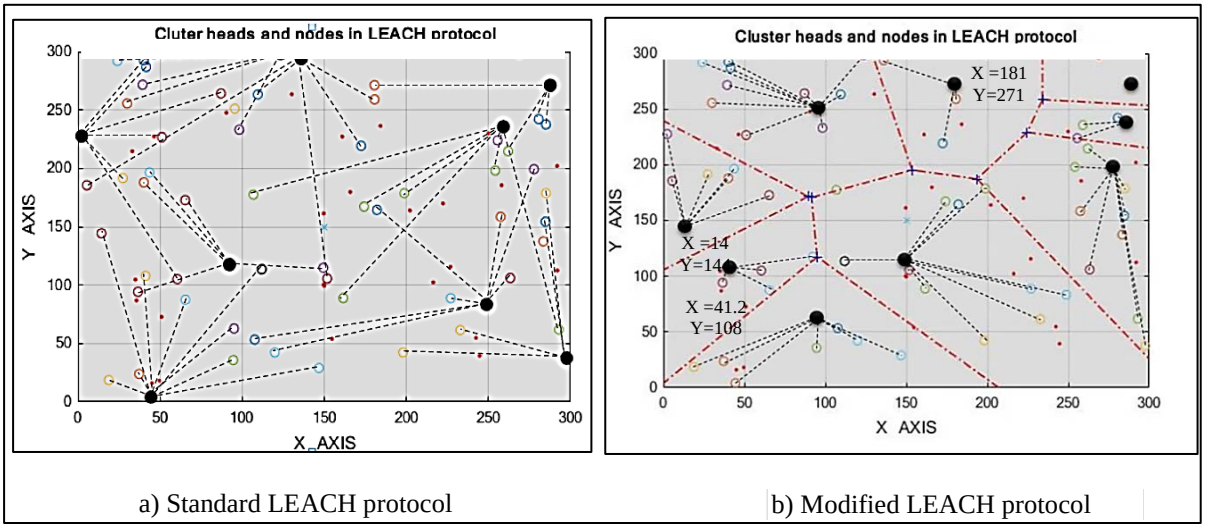


Fig. 2: Simulation of cluster heads positions and distanced between the nodes for LEACH protocol: a) - before and b) - after modification.

The results are stored in an array containing two columns with x and y coordinates of the nodes positions in the cluster area. If the values in columns are sorted ascending or descending the cluster head position will be shown close to the middle of the column, as it is shown in figure 3.

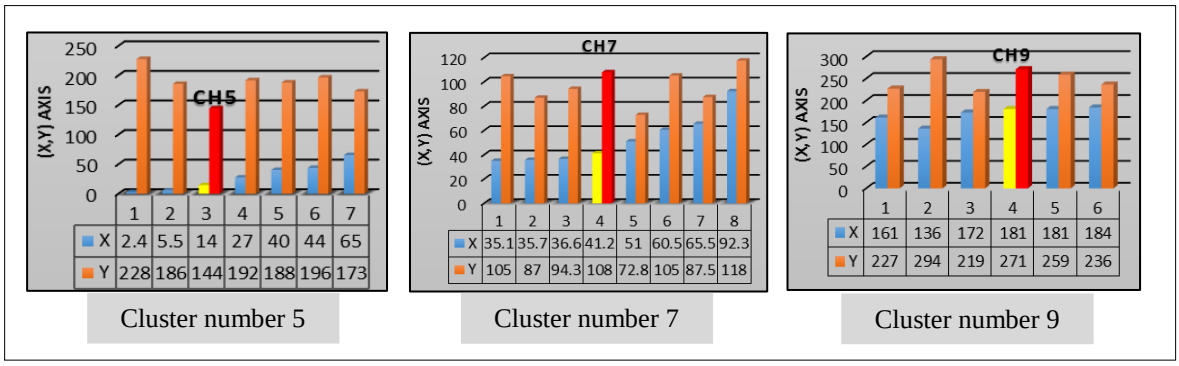


Fig. 3: Choosing cluster heads in different clusters

Consequently, the higher percentage of the cluster head nodes are always near to the middle point of the cluster. This results the less amount of energy dissipation rate, comparing with the standard LEACH protocol, where the cluster heads are randomly distributed in the clusters as shown in figure 2a. LEACH protocol does not provide clarity about the position of a cluster head in a cluster area because it only depends on a probability model. Some cluster heads are at the edge of the cluster area, some cluster heads are near to each other and some network areas do not contain any cluster heads. Comparing figures 2.a and 2.b we can notice a significant difference between the cluster heads position in the cluster area. The energy efficiency level in the cluster heads could not maximize the energy efficiency of WSN.

7. Conclusion

The energy in Wireless Sensor Network (WSN) is the most important factor in designing a protocol. Energy consumption is the main aspect to make the network more secure, therefore the energy efficiency and performance of the network depends on the position of cluster heads in the network. The aim of this paper is to evaluate the cluster heads in LEACH protocol, which is one of the most popular clustering protocols using a probability model to select cluster head. The modification of LEACH protocol is proposed, in which the cluster head in the network is chosen depending on a distance or location of all nodes in the same cluster. The simulation results were conducted for two different type of LEACH protocol: standard version and modification version. The simulation showed that the proposed modified protocol ensures better selection of the cluster heads, which are placed near to the middle point of the cluster. In the future work more factors of the Wireless Sensors Network effecting its lifetime will be considered, such as scalability, topology change, environment and power consumption.

References

1. S. Chandramathi, U.Anand, T.Ganesh, S.Sriraman and D.Velmurugan , "*Energy Aware Optimal Routing for Wireless Sensor Networks*", Journal of Computer Science 3 (11): 836-840, 2007, ISSN1549-3636© 2007.
2. S. Lindsey, C.S. Raghavendra, "PEGASIS: *power efficient gathering in sensor information systems*", in: Proceedings of the IEEE Aerospace Conference, Big Sky, Montana, March 2002.
3. Reenkamal Kaur G., Priya C. and Monika S., "*Study of LEACH Routing Protocol for Wireless Sensor Networks*", International Conference on Communication, Computing & Systems (ICCCS), India, 2014, pp.196-198.
4. Rajesh P., Sunil P., Vijay U.," *Energy and hroughput Analysis of Hierarchical Routing Protocol(LEACH) for Wireless Sensor Networks"*, International Journal of Computer Applications, Volume 20, No. 4 (April 2011), pp. 32-36.
5. Jia Xu, Ning Jin, Xizhong Lou,Ting Peng,Qian Zhou,Yanmin Chen, "*Improvement of LEACH protocol for WSN*", 9th International Conference on Fuzzy Systems and Knowledge Discovery, China, 2012.
6. Sunil Kumar S., Prabhat K., and Jyoti Prakash S., "*A Survey on Successors of LEACH Protocol*", IEEE Access,Volume: 5, 2017, pp. 4298 - 4328
7. Chen B., Yao H., Yang M., et al. *A intercluster multi-hop routing protocol improved based on LEACH protocol*. Chinese Journal of Sensors and Actuators, 2014, pp. 373 - 377.
8. Parminder K., Mrs. Mamta K., "*The Energy-Efficient Hierarchical Routing Protocols for WSN: A Review*", International Journals of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 11, ISSN: 2277 128X, India, 2012.
9. A. Iera, C. Floerkemeier, J. Mitsugi, and G. Morabito, "*Special issue on the Internet of Things*," in IEEE Wireless Communications, vol. 17, 2010, 8–9.
10. Volodymyr M., Sebastian B. and Taras P., "The *Dependence Between the Number of Rounds and Implemented Nodes in LEACH Routing Protocol-Based Sensor Networks*", International Interdisciplinary PhD Workshop, Brno 2016, pp. 111-116.

Marwah BANI SAAD[1], Wojciech BIENIECKI[2], Lidia JACKOWSKA-STRUMIŁŁO[3]

Algorithms for Automatic Open Questions Scoring in E-Learning Systems

This study deals with the automatic scoring of short answer question type in the e-learning systems. The paper presents a new algorithm of open questions scoring, which is based on the key words method with lemmatization. The algorithm is dedicated to languages with complex grammar, for which the simple key words method is not sufficient. The algorithm has been implemented for Polish language and tested experimentally on a large group of students. The results obtained for the key words method and its modification with lemmatization have been compared and presented. The paper reports on the results of statistical analysis of these evaluations.

1. Introduction and related work

Short Answer Questions (SAQ) are generally easier to develop than multiple choice questions and easier to score than short or long essay questions, nevertheless they are useful in assessing students' knowledge and their ability to remember foundational information such as terminology. Short-Answer Question can be answered in just a few words or phrases and well-designed SAQs provide students with clear assessments and provide the assessor clear marking guide.

The automatic evaluation of SAQs has not been yet implemented for most popular e-learning platforms like WebCT™, Blackboard™ and Moodle™ [1]. In [2] C-rater is mentioned, which is a module to score responses to content-based short answer questions and it is based on paraphrase recognizer. C-rater matches as many sentence-level concepts as possible between teacher and student answers and generates a canonical representation of the texts using syntactic variation, anaphora, morphological variation, synonyms, and spelling correction for recognizing the responses that have a common meaning [3]. It extracts the underlying structure of the response, resolves pronoun reference, normalizes across inflected words, and recognizes the use of similar terms and synonyms.

Our aim is to investigate a new algorithm in scoring open text responses automatically. The open text responses we are dealing with are ranging from one word up to one line. These responses are for the questions that typically ask the student to state, describe, suggest, explain, etc. These questions are from an assessment in different subjects on a web system in the local network in the Institute of Applied Computer Science, Lodz University of Technology [4]. The system is used in practice and supports academic teachers in the laborious process of tests' marking. It displays the assessment for the students in two languages, English and Polish [5, 6]. This system had been introducing open questions type and applied the Levenshtein measure for spell checking. The Levenshtein measure proved to

[1] Ph.D. student, Institute of Applied Computer Science, Lodz University of Technology, Lodz, Poland, e-mail: banisaad88@gmail.com
[2] Ph.D., Institute of Applied Computer Science, Lodz University of Technology, Lodz, Poland, e-mail: wbieniec@kis.p.lodz.pl
[3] Ph.D., D.Sc., Eng. Institute of Applied Computer Science, Lodz University of Technology, Lodz, Poland, e-mail: lidia_js@kis.p.lodz.pl

be inadequate for comparing answers consisting of a few words. Thus using standard syntactic and semantic analysis methods will be difficult.

In [6] we examined a family of approaches to automatic scoring of SAQ basing on the teacher valid responses. We assumed, that the student response would be compared to up to four model answers prepared by the examiner with the ability to determine which answer is more and which is less correct.

The proposed methods have been summarized in Table 1.

Table 1: Variants of scoring SAQ used in [6]

Variant name	*Description*
exact	if one of the acceptable answers (*acc_ans*) exactly matches the answer given by the student (*st_ans*), the task is assessed at 5 points, otherwise – 0 points.
exact_weight	if the student's response is equal to the first model answer, he receives 5 points, if it matches the second one – 4 points and so on.
Lev_exact	like *exact*, but with approximate match;
Lev_weight	like *exact_weight*, but with approximate match;
Lev_weight_compliance	like ***lev_exact*** but the score reflects the degree of compliance with any of the model answer;
Lev_exact_split	Treating student's and teacher's answers as sets of words and examining the intersection of these sets using Jaccard index [10]

Two kinds of short answer questions were implemented and tested: type 1 (T1) – with scoring by exact matching and type 2 (T2) – by approximate matching with the Levenshtein measure.

From 67 questions of T1 type processed with variant ***exact,*** 58 (88%) marks were compliant to manual assessment and from 397 answers of T2 type processed with variant ***Lev_exact,*** 371 answers (93%) were correctly assessed.

At the same time the experiment showed that the measures: ***exact*** and ***Lev_exact*** are inadequate for comparing answers consisting of a few words and numbers.

Then, all questions were assessed by ***Lev_exact_split*** method. From 67 questions of T1 type 66(98.5%) marks were compliant to manual assessment and from 397 answers of T2 type 389 (98%) answers were correctly assessed. Empty answers were qualified as incorrect by the teacher and by the system.

Currently we extend the variant ***Lev_exact_split*** with some pre-processing using dictionary created as a process of lemmatization [11].

2. Stemming and lemmatization

Word stemming is an important feature supported by present day indexing and search systems and it is a popular tool technology for quantitative text processing in Information Retrieval and Data Mining. The main purpose of stemming is to reduce the family of inflected and derived words (nouns, adjectives, verbs, adverbs, negations etc.) to its *stem* – a sequence of letters common to the whole word-forming family [7].

The following steps are done by the stemming algorithm:

1. Find a unique sequence of characters for any given word form, called a *stem*, which represents the word root.
2. Morphological forms of a word are supposed to have the same base meaning and should be mapped to the same *stem.*
3. Words that do not have the same meaning should be left as they are.

From a group of words: *dziennikarstwo*, *dziennikarz*, *dziennik*, *dzienny* we may obtain a stem *dzienn*. Note, that in this case the stem is not equal to the root for these words which is *dzień* and, actually *dzienn* is not a valid word.

In contrast, word lemmatizing goal is obtaining the *lemma* that represents the word basic form after decoding its part of speech and part of the sentence. This is usually required because some words have the same spelling but different meaning (homonym or homograph).

As analyzing the sentence may require some artificial understanding and having encoded all grammatical rules of a given language, it is a more complex process compared to stemming.

A *lemma* in morphology is the canonical form of a *lexeme*, which refers to the list of all the forms with the same meaning, so, the *lemma*, refers to the particular form that is chosen by convention to represent the *lexeme*. An algorithm that converts a word to its correct basic form, from linguistic point of view is called a *lemmatizer*, which refers to a tool that does full morphological analysis to identify the lemma for each word [7].

A valid lemma of the group: *psa*, *pies*, *psach*, *psami*, *psem*, *psie*, *psom*, *psów*, *psu*, *psy* is *pies*.

3. The algorithm

In current research, we will re-test the algorithm for calculation the similarity between acceptable answers given by the teacher and the student's response.

As the approach assigned as ***Lev_exact_split*** in [6] became the most reliable we also use it but we enhance the pre-processing phase.

Pre-processing of short answers used in [6] included:

1. separation of numbers from other words using space;
2. replacing all special and punctuation characters by spaces;
3. converting all letters to lowercase;
4. splitting text into words on spaces;
5. removing all remaining spaces (trimming);

The next new phase, proposed in this work, involves conversion of all separated words using the dictionary of words with basic form and flexion forms.

The following steps showed how we simplify the words in the student's answers:

1. if there is a positive hit for the examined word in the dictionary, it is converted to the first word in the current line (basic form);
2. if there is a number inside the student's answer, it will be separated from other letters;
3. if there is in the dictionary a positive hit for the upper-case form of the word (DVD, CD, ROM, etc.), it will converted to upper-case format

After the above steps are done, we calculate the score after the following steps are implemented:

1. Search the whole dictionary for the acronyms in upper-case to find its lemma.
2. When there is negative hit for some words, we search the whole dictionary for the nearest word (only one letter longer or shorter) and apply Levenshtein distance.
    a. If the word is so far different, it will be returned without change.
    b. If the nearest word is accepted, it will be converted to its *lemma*.

Then, after searching the lemmatization dictionary and applied Levenshtein distance to create new student model, the splitting algorithm will be implemented to score the students answers as it is shown in the listing (1).

```
score ← 0
ss ← GetEditArea( )
s_ans ← set(Split(Filter (st_ans)))
FOR EACH ans IN acc_ans
        points ← 0
        t_ans  ← set(Split(ans))
        common_words ← t_ans∩ s_ans
        points ← Size(common_words ) / Size(t_ans)
        IF points > score
            score ← points
        END IF
END FOR
```

Listing. 1: A pseudocode of Splitting scoring for Lemmatizing answers

Numbers, which appear in model teacher answers, are very important (for instance: how many bytes are in one kB?). If the number is present, it must be found in the student answer in exact form, otherwise the answer is incorrect. The function displayed in Listing 2 (Python code) is run in parallel to other testing methods. It returns:

a) -1 if some of the teacher answers contain number;

b) 0 if all teacher answers contain numbers and the student did not use one of these numbers;

c) 1 if some of the teacher answers contain number and the student used it aptly.

```
 def testNumericalAnswer(t_ans, s_ans):
   score = 0
   points = [0,0,0,0]
   st_ans_not_number = (len(re.findall(re_number,s_ans))==0)
   N = 0
   for j, ans in enumerate(t_ans):
     y = re.findall(re_number, ans)
     if  len(y)> 0:
       N+=1
     if len(y) == 0 or st_ans_not_number:
       points[j] = -1
     elif all(x in s_ans for x in y):
       points[j] = 1
     else:
       points[j] = 0

   if 0 in points:
     score = 0
   elif 1 in points:
     score =1
   else:
     score =-1
   if st_ans_not_number and N==4:
     score = 0
 return score
```

Listing. 2: Python code for the number criteria

4. Results of the experiment

A group of students has been asked to do the test of open question type and the total number of the questions was 509. Open answer questions are presented in such a way, that the possible answers (stored in the database) are not displayed to the student, who must type the answer into a given *EditArea*. While formulating the question, the teacher fills up to four possible answers (*acc_ans*). The phase of test planning is very important and the teacher should take into account all possible synonyms in the set of the four correct answers.

When the assessment is finished, the student's answers are stored in the database as an Excel file (*Result.xlsx*).

The experiment has been conducted for the student answers pre-processed in a standard way and with the dictionary converting each word to its *lemma*. The answers have been assessed manually and split into correct part (grade (2.5; 5]) and incorrect part (grade [0; 2.5)). Empty answers have not been taken into account.

Table 2: Compliance of the manual and automated assessment

Type of scoring	*Expert*	*Lev_exact_split*		*Lev_exact_split* with number pre-process		*Lev_exact_split* with number pre-process and dictionary	
		as correct	as incorr.	as correct	as incorr.	as correct	as incorr.
Correct answers	186	138 (74%)	48 (26%)	141 (76%)	45 (24%)	144 (77%)	42 (23%)
Incorrect answers	187	1 (0.5%)	186 (99.5%)	1 (0.5%)	186 (99.5%)	1 (0.5%)	186 (99.5%)

In Table 2 the summarization of the compliance between manual and automatic assessments is displayed.

The basic version of the algorithm [6] for the current dataset is shown as ***Lev_exact_split***. 26 percent of correct answers have been underestimated and one incorrect answer has been

wrongly qualified as a correct. The conclusion is that the ability to detect inappropriate responses is very good and does not require improvement.

Contrary, detection of correct answers is not satisfactory for the basic version of the algorithm. Introduction of examining numbers has improved the compliance of ratings in a few cases. The use of the lemmatization and dictionary has also brought some improvement in compliance, but not so significant as it was expected. Finally, the error rate has been reduced to 23%.

5. Conclusion

A few versions of the algorithm for automatic open questions scoring in e-learning systems are considered in the paper. All of them rely on calculation the similarity between acceptable answers given by the teacher and the student's response. The previously elaborated algorithm based on key words method and Levenshtein distance has been improved by implementing two modifications: exact recognition of numbers and word lemmatization. Using dictionary pre-process and converting each word to its *lemma* is time consuming. However, it allows recognizing different forms of the responses such as plural, and converting them to basic form for easy comparison with the acceptable answers.

The new algorithms has been implemented for Polish language and tested experimentally on a large group of students. Their ability to detect inappropriate responses reached 99,5%, what is very good result. However, detection of correct answers is not satisfactory. For the basic version of the algorithm 74% of the correct answers have been recognized correctly, while implementing two modifications improved only slightly the rate to 77% for the examined dataset. Therefore, further development of the algorithm is needed, before it could be applied in practice.

In the further research, we plan to include thesaurus for pre-processing of the student responses.

References

1. Aberdour M.: *Open Source Learning Management Systems*, Technical Producer, Epic, United Kingdom, 2007.
2. Leacock C., Chodorow M.: *C-rater: Automated scoring of short-answer questions*, Computers and the Humanities, Netherlands, November, 2003.
3. Burrows S., Gurevych I., Stein B.: *The Eras and Trends of Automatic Short Answer Grading*, International Journal of Artificial Intelligence in Education, 2015, Vol. 25, No. 3.
4. Jackowska-Strumiłło L., Bieniecki W., Brożyński T.: *Interactive computer system for checking students' knowledge in the Intranet*, X Conference on Information Technology Systems, Łódź, October, 2002. (in Polish).
5. Bani Saad M., Jackowska-Strumiłło L., Bieniecki W.: *Interactive Web Service for Checking Students' Knowledge*, Fifth International NOTICE Conference: iNOTICE'14, Slok, Poland, June, 2014 (CD ROM).
6. Jackowska-Strumiłło L., Bieniecki W., Bani Saad M.: *A Web System for Assessment of Students' Knowledge.* Proc. 8th International Conference on Human System Interaction (HIS 2015). Warszawa, 25-27 June 2015, pp. 20-26.

7. Jivani A.: *A Comparative Study of Stemming Algorithms*, International Journal of Computer Technology and Applications, 2016, Vol. 2, No. 3.
8. Eldahshan K., Tawfeek S., Elsayed E.: *Automatic evaluation technique for certain types of open questions in semantic learning systems*, Human-centric Computing and Information Sciences, 2013, Vol. 3, No. 10.
9. Pulman S., Sukkarieh J.: *Automatic short answer marking*, EdAppsNLP Proceedings of the second workshop on Building Educational Applications Using NLP, 2005.
10. Chierichetti F., Kumar R., Pandey S., Vassilvitskii S.: *Finding the Jaccard Median*, SODA '10 Proceedings of the twenty-first annual ACM-SIAM symposium on discrete algorithms, Austin, Texas, January, 2010.
11. Słownik języka polskiego – lista słów z odmianami https://sjp.pl/slownik/odmiany/

Adam R. ZYWICA[1], Marcin ZIOLKOWSKI[2]

Image Reconstruction Quality in Magnetoacoustic Tomography with Magnetic Induction

Magnetoacoustic Tomography with Magnetic Induction (MAT-MI) is a new hybrid imaging modality dedicated for non-invasive electrical conductivity imaging of low-conductivity objects such as biological tissues. The purpose of the presented study is to determine the optimal scanning resolution. In order to resolve this problem a special image reconstruction quality indicator must be introduced. Finally, the image reconstruction examples for object's complex shape have been analyzed and the Lorentz force divergence reconstruction has been achieved. Taking into account different numbers of measuring points and various image processing algorithms, the conditions allowing successful image reconstruction have been provided.

1. Introduction

Magnetoacoustic Tomography with Magnetic Induction (MAT-MI) is a new hybrid imaging modality dedicated for non-invasive imaging of internal structure of low-conductivity objects. MAT-MI is characterised by a good spatial resolution, high contrast and has a potential application in medical diagnosis. New biological tissue studies provide more and more information about their electrical properties that are correlated with their physiological and pathological properties. For example, breast cancer or liver tumour cells have significantly higher electrical conductivity compared to healthy tissue of the same gland or organ. These facts mean that receiving information about the spatial distribution of conductivity of the examined object will allow early detection of cancer lesions. Therefore, MAT-MI can become a primary method used in cancer screening [3][5].

The MAT-MI method can be divided into two main parts, namely so-called *forward problem* and *inverse problem*. The former is come down to express the acoustic pressure generated in the biological tissue. The latter consists of two steps, i.e. reconstructing the distribution of the Lorentz force divergence according to the ultrasonic signals collected by piezoelectric transducers or calculated at 'measuring points' (transducers' equivalents in a present simulation study), and finally the imaging of the electrical conductivity distribution [3][5].

MAT-MI relies on excitation of electromagnetic field method and detection of ultrasonic method. In MAT-MI technique an object to be imaged is placed in static and time-varying (pulsed) external magnetic fields. In the object, the pulsed magnetic field induces eddy currents. Consequently, the object emits acoustic waves through the Lorenz force generated as result of interaction between static magnetic field and eddy currents. The propagated acoustic waves can be acquired and used for electrical conductivity image reconstruction [3]. The illustration of MAT-MI concept has been shown in Figure 1.

[1] Ph.D. student, West Pomeranian University of Technology, Szczecin, Sikorskiego 37, 70-313 Szczecin, POLAND, tel. +(48-91) 4494791, e-mail: adam.zywica@zut.edu.pl
[2] D.Sc. Professor, West Pomeranian University of Technology, Szczecin, Sikorskiego 37, 70-313 Szczecin, POLAND, tel. +(48-91) 4494481, e-mail: marcin.ziolkowski@zut.edu.pl

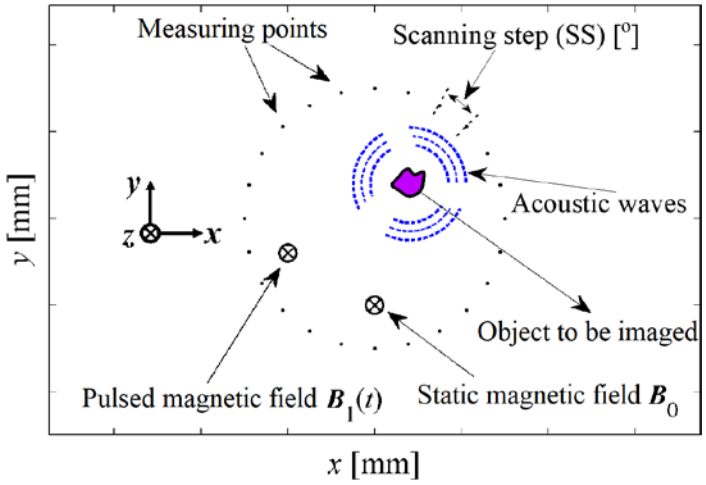


Figure. 1: Schematic diagram of 2D MAT-MI concept.

The purpose of the presented study is to determine the optimal scanning resolution which is the equivalent of ultrasonic pressure measuring points' number. In order to resolve this problem a special image reconstruction quality indicator must be introduced. Finally, the image reconstruction examples for object's complex shape have been analysed and the Lorentz force divergence reconstruction has been achieved. Taking into account different numbers of measuring points and various image processing algorithms, the conditions allowing successful image reconstruction have been provided.

2. Reconstruction quality

In order to determine the conditions allowing successful image reconstruction an assessment of the quality of the reconstructed images must be made. In this case, the image reconstruction quality should be understood by the similarity between the reconstructed image and the original image. There are two kinds of quality assessment: *subjective assessment* (a method based on estimation of image quality by human) and *objective assessment* (considered in this work) where the goal is to design and use quality indicators that can predict perceived image quality automatically. In practice, an objective image quality metric can be employed to *benchmark* and to *optimise* scanning step (SS) and the parameter settings in the simulation studies which can be used subsequently in experiments to calibrate a measuring system [4].

In tomography, various approaches and correlated indicators are commonly used in image processing approaches, e.g. an image feature extraction (shape, curve, texture, and histogram) and straight image comparison. The latter approach, based on comparing binary images globally, has been used in this work [1]. The block diagram of a quality estimation of image reconstruction is shown in Figure 2.

In order to obtain sample images, the first step of the MAT-MI inverse problem ought to be conducted. The Lorentz force divergence can be reconstructed using time reversal technique which is one of the ultrasound imaging modalities [3][5]. The resulting sample images (stored in grayscale) are subjected to a binarization and then compared with the original image of the model. The binarization process has been performed with the help of Matlab's function *imbinarization*. The function uses Otsu's method, which chooses the threshold value to minimize the intraclass variance of the thresholded black and white pixels (*global* thresholding method).

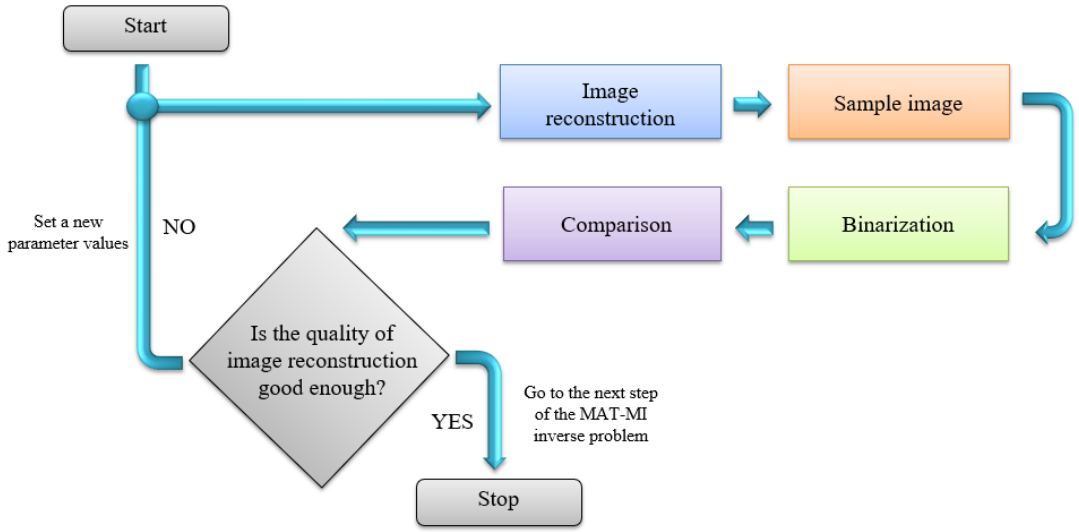


Figure. 2: Block diagram of reconstruction quality evaluation algorithm.

The most appropriate and natural ways to compare binary images is the statistical approach using such indicators as: positive predictive value (*PPV*, precision) and true positive rate (*TPR*, sensitivity) which are derivatives from so-called *confusion matrix* (error matrix) [2]. PPV and TPR can be determined by the following expressions:

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \cdot 100\%, \tag{1}$$

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \cdot 100\%, \tag{2}$$

where: TP (true positive) and TN (true negative) are the numbers of background image reconstruction pixels consistent with model pixels and inconsistent with model pixels, respectively; FP (false positive) and FN (false negative) are the numbers of object's image reconstruction pixels consistent with model pixels and inconsistent with model pixels, respectively.

Besides confusion matrix approach, very often, others measures are used to evaluate the similarity of two images, e.g. pixel-to-pixel matching (MPxP), peak signal-to-noise ratio (PSNR) and correlation (COR). MPxP is the method in which each pixel from reconstructed image is compared to the corresponding pixel from the original image. Secondly, peak signal-to-noise ratio (PSNR) is the image quality measure which is simple to calculate, has clear physical measuring but does not match well with the human perception. In practice, the higher PSNR generally the higher image reconstruction quality. The PSNR is defined as (in dB) [4]:

$$\mathrm{PSNR} = 20 \cdot \log_{10} \cdot \left(\frac{MAX_I}{\sqrt{MSE}} \right), \tag{3}$$

where: MAX_I is the maximum possible pixel of the image, *MSE* is the mean squared error.

Lastly, correlation (COR) is a related optical method that employs tracking and image registration technique for accurate 2D measurements of changes in images. COR uses the cross-correlation concept which is based on the measurement of similarity of two images as a function of the displacement of one relative to the other. The 2D discrete cross-correlation COR can be defined as:

$$\mathrm{COR} = \frac{\sum_m \sum_n \left[f(m+i, n+j) - \bar{f} \right] \left[g(m,n) - \bar{g} \right]}{\sqrt{\sum_m \sum_n \left[f(m+i, n+j) - \bar{f} \right]^2 \sum_m \sum_n \left[g(m,n) - \bar{g} \right]^2}}, \tag{4}$$

where: $f(m,n)$ is the pixel intensity or the grayscale value at a point (m,n) in the original image, $g(m,n)$ is the grayscale value at a point (m,n) in the translated image, $\bar{f}$ and $\bar{g}$ are mean values of the intensity matrices f and g respectively.

3. Results

The MAT-MI forward problem has been performed with the help of the Finite Element Method (FEM). In order to reconstruct the acoustic source as the Lorentz force divergence $\nabla \cdot (\boldsymbol{J} \times \boldsymbol{B}_0)$, the time reversal algorithm has been used. The model view and results of the image reconstruction for different number of measuring points (NP) is shown in Figure 3.

The 2D geometry of the object to be imaged which consists of inner and outer layers with the complex shapes and low conductivities has been considered. It should be noted, that the layers are acoustically homogeneous without any reflections, dispersion and attenuation. Moreover, all the studies are ideal and therefore the recorded signals are considered as noise-free.

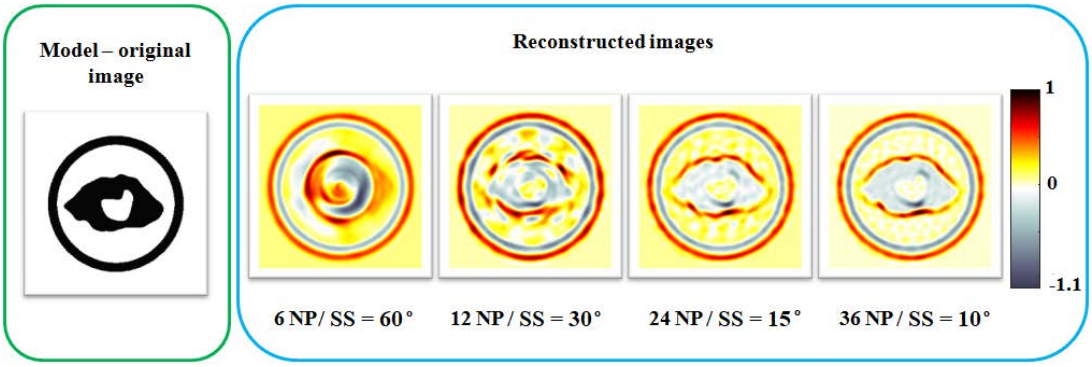


Figure. 3: The model view (left) and normalized reconstructed images for different number of NP (right).

We have made two groups of the binary images from reconstructed images (ten images in each) which were compared to the origin (image of the model). The former is made after binarization from the grayscale image reconstruction of the object with full range of values. In turn, the latter is made after binarization from the grayscale image reconstruction of the object without positive values. Results of image binarization (for the first four images) are shown in Figure 4.

In Figure 3 and Figure 4, we present the object image reconstruction for 6, 12, 24, and 36 measuring points (NP). The subjective assessment is undoubtedly pointing at a greater similarity of the images in case B to model image than in case A. For 6 NP the ring was reconstructed but not much information about the inner part of the object is revealed. With 12 NP, some of the inner basic features of the object start to emerge but still blurring artefacts around and between the inner and outer part of the object can be observed. Finally, for 36 NP the reconstructed acoustic source positions almost exactly correspond to the actual locations of the object. The greater number of measuring points, the better quality images with less time reversal noise can be observed.

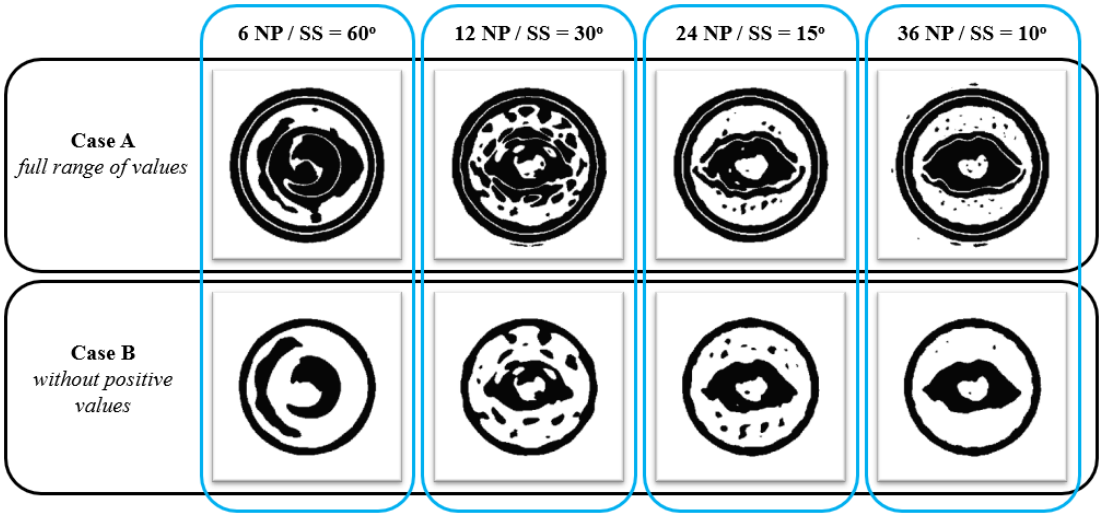


Figure. 4: Result of binarization the reconstructed images for case A and case B.

In order to reduce the computation time, the size of images has been rescaled to 300 x 300 pixels. The straight image comparison (globally) has been chosen which seems best adapted for this case. The binary images have been compared without using feature extension (feature comparison). In order to evaluate the reconstruction quality properly, some of the statistical indicators from the confusion matrix have been computed for the cases A and B. The results are shown in Figure 5, on the left and right, respectively.

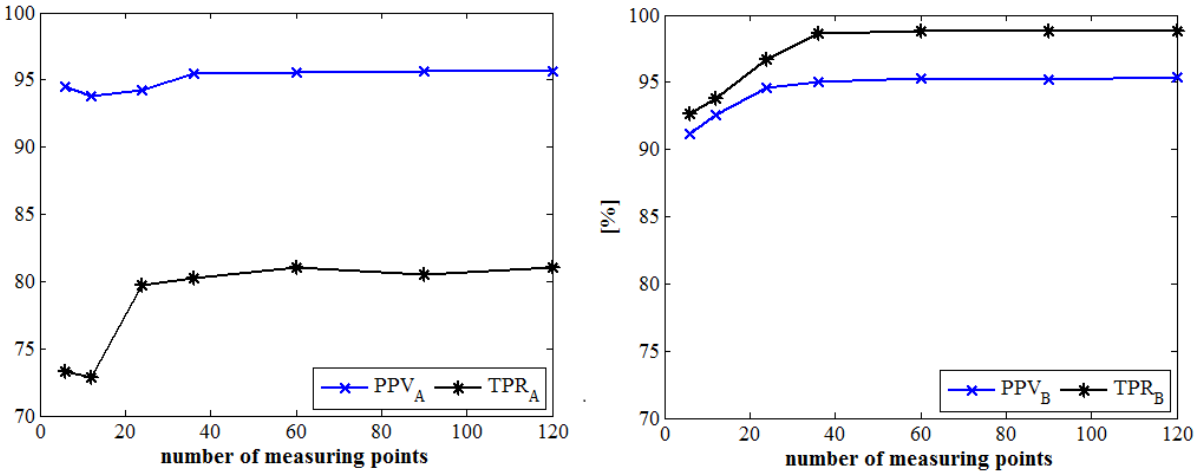


Figure. 5: Values of confusion matrix indicators obtained for the case A and case B, respectively.

The results present the high similarity of the compared binary images in both cases. However, the lesser difference between PPV and TPR for case B than case A indicates that the cut off of positive values from the grayscale image reconstruction is resulted in a distinct improvement in the match of the reconstruction to the model. Significant image reconstruction improvement takes place when the number of measurement points begins by 36 for case A and case B, respectively. What is more, further increment of measuring points' number does not change the image crucially. In addition, the binary image comparison was carried out using three other indicators i.e. pixel-to-pixel matching (MPxP), PSNR and correlation (COR). The results of the comparison have been summarized in Table 1.

Table 1: Image comparison method values for case A and case B for different number of measuring points.

		A			B		
SS [^{o}]	NP	MPxP [%]	PSNR [dB]	COR	MPxP [%]	PSNR [dB]	COR
60	6	73,32	54,32	0,496	92,61	57,08	0,624
30	12	72,88	54,18	0,475	93,77	57,82	0,684
15	24	79,69	55,21	0,548	96,74	59,78	0,797
10	36	80,22	55,49	0,583	98,67	61,10	0,850
6	60	81,04	55,65	0,594	98,77	61,37	0,859
4	90	80,48	55,56	0,590	98,81	61,35	0,858
3	120	81,00	55,66	0,596	98,80	61,43	0,861
2	180	81,11	55,68	0,597	98,81	61,46	0,862
1	360	81,11	55,69	0,598	98,80	61,43	0,861

It can be observed that the results present the great similarity of the compared binary images in both cases. As for previous situation, significant image reconstruction improvement occurs when the number of measurement points starts at 36 for both cases.

4. Conclusion

In this paper, the influence of MAT-MI scanning resolution on the image reconstruction quality has been studied in detail. In order to evaluate properly the image reconstruction quality, the set of quality indicators have been used. The main task of the indicators is comes down to predict a perceived image quality automatically allowing to benchmark and to optimise scanning step and parameter settings in the simulations studies. The information obtained after the simulation tests will allow for accurate matching of the experimental conditions and to calibrate the measuring system. Taking into account the subjective and objective assessment of image reconstruction quality, among all the indicators we used to determine the similarity between the reconstructed image and the original image, the correlation gives the best results. It has to be noted that the used indicators depend on the resolution of images being compared which is correlated with sampling time used in the MAT-MI forward problem which was not presented in this work.

References

1. Baudrier E., Nicolier F., Millon G., Ruan S.: *Binary-image comparison with local-dissimilarity quantification*, 2008, Pattern Recognition, Vol. 41
2. Lech P., Okarma K.: *Optimization of the Fast Image Binarization Method Based on the Monte Carlo Approach*, Elektronika IR Elektrotechnika, 2014, Vol. 20
3. Li X., Yu K., He B.: *Magnetoacoustic Tomography with Magnetic Induction (MAT-MI) for Imaging Electrical Conductivity of Biological Tissue: a Tutorial Review*, 2016, Physics in Medicine & Biology, Vol. 61
4. Wang Z., Sheikh H. R., Bovik A. C.: *Objective Video Quality Assessment*, 2003, Chapter 41 in The Handbook of Video Databases: Design and Applications, Furht B. and Marqure O., CRC Press
5. Zywica A.R.: *Magnetoacoustic Tomography with Magnetic Induction for Biological Tissue Imaging: Numerical Modelling and Simulations*, Archives of Electrical Engineering, 2016, Vol. 65(1)

Leszek ŚLIWA[1]

Natural Language Processing in Computer Go

This research shows how NLP techniques can be used to build the probabilistic language model of Go game. The proposed algorithm selects the potentially most promising moves that should be evaluated with Monte-Carlo simulations. A move prediction is done based on n-grams model and a noisy channel model. Statistical hypothesis testing is used for a model improvement. The game model is evaluated with a perplexity measure. The novel approach shows its advantages over a selection of most promising moves based only on confidence intervals.

1. Introduction and Overview

Strategy in the Go game, as well as the behavior of people, cannot always be described by strict rules. In addition, the use of human language is not restricted to concise and universal grammatical rules. Of course, there are rules for constructing words and sentences. Likewise, the game of Go defined criteria for correctness of a move according to the rules of the game. Players in Go often use the same schemes, but they also do new moves - just like people who constantly invent and introduce novel words and phrases to the dictionary. The more observations we collect, the more we know about the language (in the case of Go, the language of the game), and we can construct a more accurate model.

In my paper “Using Patterns in Computer Go", I presented the technique for machine retrieving patterns from a collection of game records played between human expert players and implementing patterns in a computer Go program [13]. Each pattern has been represented by a small 3×3 points piece of board with an empty point in its center, where a potential move has been considered. Every pattern identified by an integer number has been given a weight depending on its frequency. During the clustering, to avoid duplication caused by patterns variations, operations of: rotation, mirroring and color exchange have been applied. For final set, patterns with highest weight, due to the Pareto rule (80/20), have been selected. To verify how the patterns are effective and improves the playing strength of the computer program, the heuristics based on patterns has been tested in 4 contests and compared to reference tournaments. The results of experiments show that the heuristics of "good" patterns improves the playing strength on medium and large boards. Unfortunately the proposed solution didn’t scale well and applied to 5×5 and 7×7 patterns were not satisfactory. This work is a continuation of the research but proposes a different approach – the use of Natural Language Processing (NLP) techniques to predict moves in the computer Go game.

Section 2 presents related works. Section 3 is dedicated to basic NLP concepts. Next, Section 4 introduces a minimum editing distance. In Section 5, a brief explanation is given how a game language has been modeled. Section 6 presents a noisy channel model used for a move prediction. Section 7 is focused on the game model assessment. The results of experiments are presented and discussed in Section 8. The research conclusions are gathered in Section 9.

[1] Ph.D. student, The Institute of Computer Science, Faculty of Electronics and Information Technology, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warszawa, POLAND,
tel. +(48-22) 234 74 32, fax: +(48-22) 234 60 91, e-mail: l.s.sliwa@stud.elka.pw.edu.pl

2. Related Works

In recent years, a new algorithm MCTS (Monte-Carlo Tree Search) was developed [2]. MCTS was developed initially for computer Go game, but it is a general algorithm and can be applied to solve other problems. MCTS is a best-first search method guided by Monte-Carlo simulations that does not require a positional evaluation function. MCTS consists of four main steps: selection, expansion, simulation, and backpropagation. In the selection step, the tree is searched from the root node until a leaf. In the expansion step, a new node is added to the game tree. In the simulation step, random moves are played in a self-game. Finally, in the backpropagation step, the result of a simulated game is propagated from a newly added leaf node up to the game tree root.

MCTS has many variants that differ in search strategy, but most of them are based on confidence intervals and Central Limit Theorem. The confidence intervals in computer Go are used to select the potentially most promising moves that should be evaluated with Monte-Carlo simulations. Smart selection of moves for evaluation has the crucial impact on program's playing strength. The Central Limit Theorem states that the probability distribution of the sum of independent and identically distributed random variables with finite variance approaches a normal distribution. In my research, I showed that the move-selection strategy which implements calculation of the exact confidence intervals based on discreet, binomial distribution is much more effective than based on normal [14]. The novel approach shows its advantages particularly in games played on medium and large boards.

Strategies used in selection and simulations can be facilitated by applying knowledge in a form of patterns. There has been a big effort to implement patterns in a Go engine from the early beginning – Zobrist A.L. (1970) [16]. Boon M. (1990) implemented pattern matching in the computer program GOLIATH [1]. Cazenave's work (2001) consisted in automatic acquisition of patterns generated in a special tactical context [3]. Gelly et al. (2006) introduced the sequence-like simulation [5]. Wang and Gelly (2007) implemented sequence-like 3×3 patterns in their Go program MOGO [15]. Coulom (2007) searched for useful patterns in Go by computing Elo ratings for patterns, improving program CRAZY STONE [4]. Hoock and Teytaud (2010) used Genetic Programming to automatically find "better" patterns for their computer program MOGO [7].

Computer program AlphaGo, developed by Alphabet Inc.'s Google DeepMind, uses MCTS algorithm together with deep neural networks: a value network, to assess the state of the game and two policy networks – one taught under supervision and the other taught with reinforcement, to find potential strong movements. Both networks implement a strategy of probing the space of possible solutions. The value network is trained by the supervised learning method based on the records of games played between players competing at the championship level [12]. AlphaGo beat the best professional human Go player in the world, in a three-game match at the 2017 "Future of Go Summit".

NLP techniques for a move prediction were not researched in related works.

3. Basic Concepts

In natural language processing techniques, a collection of analyzed documents (books, articles, press releases, etc.) is called a *corpus* [8], [11]. The documents that make up the corpus consist of sequences of *segments* – words and punctuation. Each unique segment is called a *type*. For example, the phrase "four times, two times" consists of four segments, but only three types.

A collection of all types is called a *dictionary*. Sequence consisting of one segment is called a *unigram*. A *bigram* consists of two segments, a *trigram* of three segments, a *n-gram* of n-segments.

Table 1 shows the basic analogies between the concepts used in natural language processing and in computer Go.

Table 1. Analogies between concepts of natural language and computer Go.

NLP	**Computer Go**
Corpus	Collection of game records between players competing at the championship level. A collection that meets the criteria of representativeness and authenticity.
Language	Collection of moves' sequences recorded in the corpus.
Word, segment, type	Move made by a player. Move is coded in the form of a pattern that matches the direct context.
Dictionary	Collection of unique moves recorded in the corpus.
Sentence	Sequence of moves made in close proximity.
Word correction	Move prediction. Choice of a move based on the game language model.

The game model is described by probability distribution, which is characterized by the ability to observe a sequence of moves in the large corpus. Since there are infinitely many sequences of moves and that for each of them the probability of occurrence is to be determined, the task is not easy, for example, the solution cannot be a simple counting of moves because some of these moves may never occur even in the largest corpus.

4. Minimum Editing Distance

The minimum editing distance allows estimation the similarity of patterns. To indicate similar patterns, a measure of similarity must be defined. Transforming one move pattern into another consists of a sequence of elementary operations: insertion, removal or replacement of a single stone. Editing distance is the sum of the number of individual operations required to transform one move into another [10].

Transformation of one pattern into another can be done in many ways, and consequently, the search space is large. The search space can be modeled as a tree in which the root is the starting pattern, branches correspond to elementary editing operations, nodes to the current pattern generated by subsequent operations, and leaves to the final words. The goal of the search algorithm is to find the path that has the least cost.

The algorithm for calculating the shortest editing distance has been enhanced with scales that accept smaller values for those fields that are farther away from the midway point of the pattern (the point at which the stone is laid). Distance is calculated in Manhattan measure in 2D space. In the basic version, insertion, deletion and replacement operations have a fixed cost. Fixed values are replaced by a cost function whose value depends on the distance from the midway point of a pattern.

5. Language Model

The game model determines the probability of a sequence of moves. The sequence of moves can be formally written as $w = (w_1 w_2 \ldots w_n)$ and may be abbreviated w_1^n. The capital letter P means the probability. If w is a sequence of moves, then $P(w_i \mid w_{i-1})$ is a conditional probability of the occurrence of the sequence w_{i-1}, w_i. By using a chain rule, the probability of a sequence occurrence can be determined by the equation 1.

$$P(w_1^n) = P(w_1)P(w_2 \mid w_1)P(w_3 \mid w_1^2) \ldots P(w_n \mid w_1^{n-1}) = \prod_{k=1}^{n} P(w_n \mid w_1^{k-1}) \tag{1}$$

In the view of the considerable number of parameters used in this model, the Markov assumption is usually used according to which the sequence history depends entirely on the preceding local context (equations 2):

$$\begin{aligned} &\text{1-gram (unigram):} && P\left(w_k \,\middle|\, w_1^{k-1}\right) \approx P(w_k)\ , \\ &\text{2-gram (bigram):} && P\left(w_k \,\middle|\, w_1^{k-1}\right) \approx P(w_k \mid w_{k-1})\ , \\ &\text{n-gram:} && P(w_k \mid w_1^{k-1}) \approx P(w_k \mid w_{k-n+1}^{k-1})\ . \end{aligned} \tag{2}$$

6. Noisy Channel Model and Move Prediction

A move prediction can be compared with Shannon's game in which the player wants to predict a word in the text. Based on the current state of the game, a list of candidate moves is generated. The list only contains game rules-based moves. Because of the large branching of the game tree, only moves that are close to the previous move made by the opponent are analyzed. Then, from the set of candidates, the one that maximizes the probability value based on the noisy channel model is selected.

In the noisy channel model, the word w is given to the input of the transmission channel. In view of channel noise, the word w is distorted into a word x during transmission. The noisy channel can be described mathematically using the theory of probability. Knowing the distorted word x received at the output of the channel, a list of words V that could have appeared on input of the channel can be build. From the set of candidates, the most likely word is selected (equation 3) [9].

$$\hat{w} = argmax_{w \in V}\, P(w|x) \tag{3}$$

Using the Bayes theorem, the equation 3 can be transformed into the equation 4.

$$\hat{w} = argmax_{w \in V}\ P(x|w)P(w)/P(x) \tag{4}$$

Because $P(x)$ does not depend on w, then the equation can eventually be written as 5.

$$\hat{w} = argmax_{w \in V}\, P(x|w)P(w) \tag{5}$$

The last equation contains the product of two factors: (1) $P(x \mid w)$ – likelihood, the channel model or alternatively the error model, $P(w)$ – prior, a language model, probability of occurrence of the correct word at the input of the channel.

A move can be done if it complies with the rules of the game. Not every correct move is a "strong" strategic move. How in the sequence of moves in the game find those moves that

are potentially strong? For each move, a set of candidate moves is created. The collection of candidates includes moves within the limited editing distance.

Let's consider the sequence of moves described by the patterns: $w_1, w_2, \dots, w_n$ (equations 6). For every i move described by the pattern w_i, the candidate set $C(w_i)$ is defined:

$$\begin{aligned} C(w_1) &= \left\{w_1,\ w_1^1, w_1^2, w_1^3, \dots\right\}, \\ C(w_2) &= \left\{w_2,\ w_2^1, w_2^2, w_2^3, \dots\right\}, \\ C(w_n) &= \left\{w_n,\ w_n^1, w_n^2, w_n^3, \dots\right\}, \end{aligned} \tag{6}$$

where w_i is the test pattern, $w_i^1, w_i^2, w_i^3, \dots$ are similar patterns within the limited editing distance from the pattern w_i.

In a turn-based game (such as Go), players perform alternate moves. In practice, the prediction applies to the current move only. Previous moves have already been made and cannot be changed. Therefore, only the last movement in the sequence is checked – we are looking for the strongest move that completes the sequence.

We assume that moves $w_1, w_2, \dots, w_{n-1}$ are correct and we are looking for w_n which will maximize the probability of the sequence $P(W)$. If for one of the moves the value of $P(W)$ is clearly greater than for others candidate moves, we can recommend this move. If the probabilities of candidate moves are comparable, then it is difficult to indicate the strongest move.

For pruning random moves (even professional players make mistakes) a statistical hypothesis testing was used. With the assumption that the population distribution is normal then Student's t-test can be used, which allows to, based on the mean and variance of the sample, determine the probability of origin of the sample from the population. We select the bigram sequence w_1, w_2. We have a series of binomial random variables, 1 – sequence occurrence, 0 – no occurrence, i.e. Bernoulli's test. If frequencies $C(w_1)$, $C(w_2)$, $C(w_1, w_2)$ and the quantity of all types N are known, then probabilities $P(w_1)$, $P(w_2), P(w_1, w_2)$ can be computed. The null hypothesis is stated in the form $H_0: P(w_1, w_2) = P(w_1)P(w_2)$. A computation of the test statistics t is straightforward (equations 7).

$$\begin{aligned} \bar{x} &= C(w_1, w_2)/N, \\ \mu &= P(w_1,\ w_2), \\ s^2 &= p(1-p) \approx p \approx P\left(w_1,\ w_2\right) \\ t &= (\bar{x} - \mu)/\sqrt{s^2/N} \end{aligned} \tag{7}$$

If the value of the test statistics t is less than $1 - \alpha$ (where α is a significance level) then there is no reason to reject the null hypothesis and it can be assumed that the sequence of moves w_1 and w_2 is random.

7. Game Model Assessment

Each game model should be evaluated. The evaluation should consider whether the model prefers good moves in relation to the weak ones. In other words, whether it gives more probability to real strategic sequences, often observed or an accidental, a weak, a rarely played moves. Model parameters are selected based on the training set. The correctness of the constructed model is examined on test data that was not previously analyzed. Also, a metric is needed, which will allow us to evaluate the quality of the model.

The best way is to compare models empirically. In the case of predictions of moves in the computer Go, the result will be the accuracy of the prediction for both models, i.e. how many moves are correctly predicted. The disadvantage of empirical evaluation is a long test time.

For the construction of prototype solutions, a rough estimate based on the perplexity measurement can be used to analyze the practical utility of constructed model. The smaller is the value of the perplexity the better is the move prediction. Minimizing perplexity is equivalent to maximizing probability. The perplexity measure PP is equal to the probability P in the test set after the normalized by a quantity of moves n in the analyzed sequence of moves $W = (w_1, w_2, \ \ldots, w_n)$, equation 8.

$$PP(W) = PP(w_1, w_2, \ \ldots, w_n) = \sqrt[n]{\frac{1}{P(w_1, w_2, \ldots, w_n)}} \tag{8}$$

Using an n-gram model to calculate probabilities, the chain rule allows to calculate the non-deterministic measure. In this case, the probability of a long sequence of moves is replaced by the product of the conditional probabilities for each move in the sequence being tested (equation 9).

$$PP(W) = \sqrt[n]{\prod_{i=1}^{n} \frac{1}{P(w_i | w_1, w_2, \ldots, w_{i-1})}} \tag{9}$$

If to use a bigram based model to calculate probability, the perplexity can be expressed by the equation 10.

$$PP(W) = \sqrt[n]{\prod_{i=1}^{n} \frac{1}{P(w_i | w_{i-1})}} \tag{10}$$

The perplexity does not depend on the length of the sequence of moves being analyzed, what is the result of normalization by the sequence length. The perplexity factor depends on the model's degree. The trigram model will have a lower factor value than the bigram model.

8. Experiments and Results

For a training data, a set of Go game records played between players represented a master level of playing strength (6-9 dan) has been used. The set contained records of the games played on the K amateur Go Server (KGS, formerly known as Kiseido Go Server) [6]. For the analysis, tournaments played in 2013 have been chosen - 12,536 games and 2,447,567 moves. All games have been played on 19×19 boards. A trigram game model was built with some pruning of random moves.

To verify the effectiveness of the proposed NLP approach, three tournaments of 100 games each with colors swapped halfway were played on boards: small 9×9, medium 15×15 and regular 19×19. The "thinking time" was set to 3 seconds for a move.

Table 2 presents the results of the games between players using different strategies in MCTS algorithm to select the potentially most promising moves. An opponent, used only the confidence intervals (CI), the player – move-selection based on noisy channel model.

In case of lack of a move in the game model, a confidence intervals approach (NC+CI) was used.

Data in the W/L (Win/Lost) ratio column refers for NC+CI player. In the tournament on a small 9×9 board the player using NC+CI won 53 games and lost 47, on a medium board 15×15 won 61 games and lost 39, on a regular board 19×19 won 66 games and lost 34.

Table 2. Comparison of effectiveness of NC+CI and CI strategies.

Player's strategy	Opponent's strategy	W/L Ratio		
		9×9	15×15	19×19
NC+CI	CI	1,13	1,56	1,94

9. Summary

This paper presents how NLP techniques can be used to build the probabilistic model of Go game. The model was refined from random moves with a tool using a hypothesis testing. The effectiveness of the novel approach was empirically verified in the tournament between the player using the game model and confident intervals and the opponent using just confident intervals. The result of a competition shows the game model improved a move-selection strategy.

References

1. Boon, M.: *A Pattern Matcher for Goliath. Computer Go*, Vol. 13, pp. 13-23, 1990
2. Browne, Cameron B., et al.: *A survey of Monte Carlo tree search methods*, IEEE Transactions on Computational Intelligence and AI in games 4.1, pp. 1-43, 2012
3. Cazenave, T.: *Generation of Patterns with External Conditions for the Game of Go*, Advances in Computer Games 9 (eds. H.J. van den Herik and B. Monien), pp. 275-293, Universiteit Maastricht, Maastricht, The Netherlands, 2001
4. Coulom, R.: *Computing Elo Ratings of Move Patterns in the Game of Go*, Int. Comp. Games Assoc. J., vol. 30, no. 4, pp. 198-208, 2007
5. Gelly, S., et al.: *Modifications of UCT with Patterns in Monte-Carlo Go*, Technical Report 6062, INRIA, 2006
6. Görtz, U.: Game records in SGF format, http://www.u-go.net/gamerecords
7. Hoock, J. and Teytaud, O.: *Bandit-Based Genetic Programming*. In Proc. Euro. Conf. Gen. Prog., vol. 6021, Istanbul, Turkey, pp. 268-277, 2010
8. Jurafsky D., Martin J. H.: *Speech and Language Processing: An Introduction to Natural Language Processing*, Speech Recognition, and Computational Linguistics, 2nd edition, Prentice-Hall, 2009
9. Kernighan M. D., Church K. W., Gale W. A.: *A Spelling Correction Program Based on a Noisy Channel Model*, In Proceedings of COLING-90, The 13th International Conference on Computational Linguistics, Vol. 2, 1990
10. Levenshtein V. I.: *Binary codes capable of correcting deletions, insertions, and reversals*. Technical Report 8, 1966

11. Manning C. D., Schutze H.: *Foundations of Statistical Natural Language*, The MIT Press Cambridge, Massachusetts London, England Processing, Second printing with corrections, 2000

12. Silver, D., Huang, A, Maddison, C, et al.: *Mastering the game of Go with deep neural networks and tree search*. Nature, Vol. 529, pp. 484-489, 2016

13. Śliwa, L.S.: *Using Patterns in Computer Go*, Pattern Recognition and Machine Intelligence: 6th International Conference, PReMI 2015, Warsaw, Poland, pp. 565-574, 2015

14. Śliwa, L.S.: *The Confidence Intervals in Computer Go*, Artificial Intelligence and Soft Computing: 15th International Conference, ICAISC 2016, Zakopane, Poland, pp. 577-588, 2016

15. Wang, Y. and Gelly, S.: Modifications of UCT and sequence-like simulations for Monte-Carlo Go, In Proc. IEEE Symp. Comput. Intell. Games, Honolulu, Hawaii, pp. 175-182, 2007

16. Zobrist, A.L.: Feature Extraction and Representation for Pattern Recognition and the Game of Go. Ph.D. thesis, University of Wisconsin, Madison, WI, USA, 1970

Ranya AL DARWICH[1], Krzysztof STRZECHA[2], Laurent BABOUT[3]

A Comparative Study to Extract Physical Properties from Digital Images

This paper evaluates the performance of three methods to detect the edges of a metallic droplet. The first method uses the features extraction and selection techniques to predict the edges of a droplet, the second use sobel filter, the third use active contour segmentation method. All the images that are studied are obtained from the THERMO-WET system. Various edge detection methods have been implemented in the MATLAB environment.

1. Introduction

Edge detection is a fundamental tool in image processing, and particularly in features extraction. Whereas, boundaries contain a rich of information in the image [4], it aims at distinguishing the points in an image at which the brightness is modified sharply or regularly. This paper presents a comparative study of three edge detection techniques i.e. features extraction and selection technique, sobel filter, and active contour segmentation method. The aim of this comparison study is detecting the boundary of a metallic droplet, and evaluate the best technique among the three. All the images that are studied are obtained from the THERMO-WET system.

The device, named THERMO-WET, as the first in the world, enabled automated measurement in a wide range of temperatures (up to 1080^0C) in a protective atmosphere [10]. These measurements are based on algorithms for image processing and analysis, and are conducted using the sessile drop technique [12]. Figure 1 shows an original image of a copper specimen obtained from THERMO-WET measurement system at a high temperature (1080^0C). The image shows good boundaries, with dark spot in the middle of the droplet, and there is a reflected light in the upper corner on the right. These two features are considered as artifacts. The segmentation techniques mentioned above are presented in more details in the following sections.

[1] Ph.D. student, Institute of Applied Computer Science, Lodz University of Technology, Poland, tel. +48426312750, fax: +420 541 146 276, e-mail: raldarwich@kis.p.lodz.pl
[2] Ph.D., Institute of Applied Computer Science, Lodz University of Technology, Institute of Applied Computer Science
[3] Ph.D., DSc., Institute of Applied Computer Science, Lodz University of Technology, Institute of Applied Computer Science

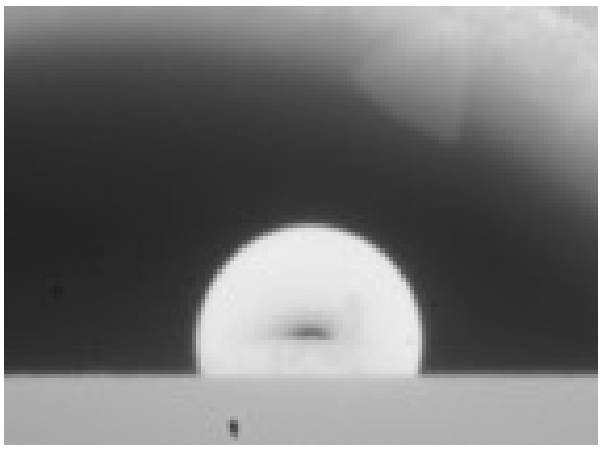


Figure. 1: Example image of copper specimen at 1080^0C.

2. Image Processing and Analysis Methods

2.1 Sessile Drop Method

The shape of the drops placed on a non-wettable surface results from two types of forces: the surface tension, which tries to give a droplet a spherical shape, and the gravity by which the drop is" flattened". In the case of the spherical drops, it is not possible to determine the surface tension. This becomes possible only when the effects of gravity are comparable with the influence of surface energy, and hence the larger droplets, the equatorial diameter of the metal droplets should be practically at least 0.5 cm. The distortion occurring in a spherical droplet allows one to determine the surface tension on the basis of the droplet dimensions and physical constants. Only symmetric drops could be taken into account. Among the currently used methods for calculating the surface tension three groups can be distinguished:

the method of the equatorial plane (measures of h and d (Figure 2)); the method of tangents (measures of Z and d);

the method of the total height (measures z0 or z) and the maximum diameter and the diameter of the base (respectively x0 or x).

The determination of the equatorial plane, according to the literature, is usually performed graphically. After the measurement of h and d, the ratio of d/2h is calculated. This ratio is essential for further calculations.

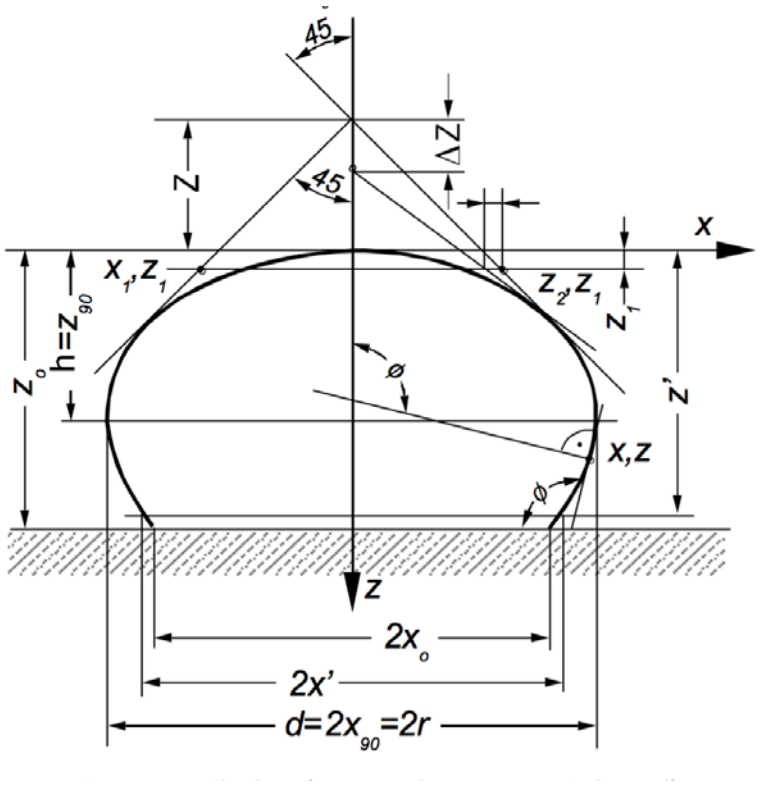


Figure. 2: Sessile drop (measured parameters indicated).

2.2 Segmentation methods

2.2.1 Image Feature Extraction and selection

Numeric images include up to several thousands of features, those features are either redundant or irrelevant. Features selection is a process used to remove redundant or noisy features from the ones extracted from images [7]. In order to improve the quality of the numeric images and to reduce the dimensionality of the data to be processed, features selection is needed here [14]. One of the most significant features of an image is color intensity in the case of grey scale images. There are many color features, including but not limited to color moments (CM), which is one of the simplest and effective color features. The color moments feature include the mean value, standard deviation, and skewness [13]. Texture is a commonly used feature in the segmentation and classification of images, it gives us information about the spatial arrangement of the color or the intensity of an image [2]. texture features have been calculated using a variety of statistical, structural, and spectral techniques especially co-occurrence matrices. Co- occurrence matrices are often used in texture analysis since they are able to capture the spatial dependence of gray-level values within an image [3]. The algorithm starts by dividing the image into blocks. Then it calculates the mean value, the standard deviation, and the skewness for each block. The algorithm also considers the grey-

level co-occurrence matrix from the image (GLCM), in the purpose of computing certain characteristics of the co-occurrence matrix. The most commonly used features that are calculated from the co- occurrence matrix are: energy, entropy, contrast, correlation, and homogeneity. In all cases, the co-occurrence matrix has been normalized by dividing all its elements by the total number of pairs of pixels considered [9]. Whereas, energy, returns the sum of squared elements in the GLCM. entropy, returns a measure of randomness of an image structure [11]. Contrast, returns the difference in the intensity of light. Correlation, returns a measure of how correlated is each pixel to its neighbor over the image. Homogeneity, returns a value that measures the nearest of the distribution of elements in the GLCM to the GLCM diagonal.

Figure 3 shows the standard deviation, the skewness, the energy, the entropy, the correlation, and the homogeneity, which are calculated in the case of blocks of size 3x3.

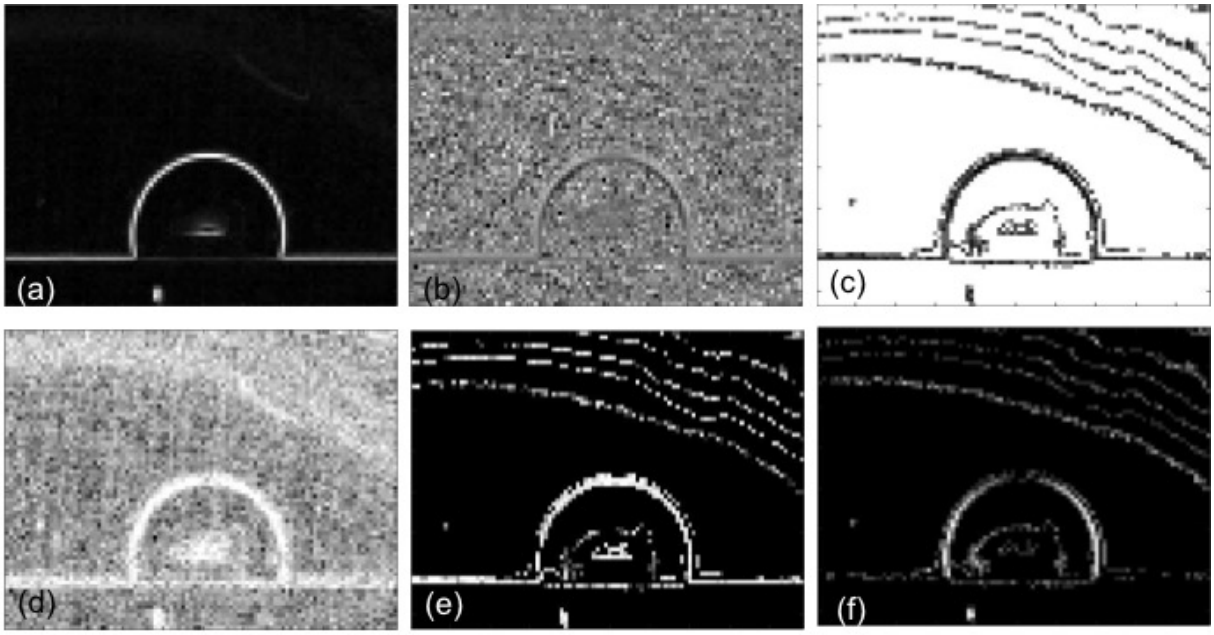


Figure. 3: (a) The standard deviation, (b) The skewness, (c) The energy of the image , (d) The entropy ,(e) The correlation, (f) The homogeneity.

2.2.2 Sobel Operator

Sobel operator is one of the most commonly used detection edges techniques. Generally, the method highlights edge in both horizontal and vertical directions and the magnitude of the gradient is usually displaced to several edges [8] Figure 4. Shows sobel horizontal edge-emphasizing filter, sobel vertical edge-emphasizing filter, and the magnitude of the gradient in the case of the image shown in figure 1.

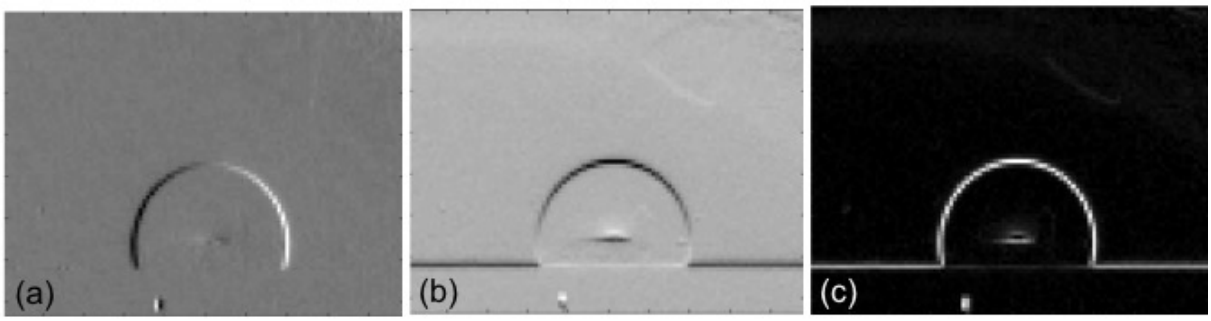


Figure. 4: (a) The sobel horizontal edge-emphasizing filter , (b) The sobel vertical edge- emphasizing filter ,(c) The magnitude.

2.2.3 Active Contour Method

Active contour, or snakes, is an image segmentation method that was proposed by Kass et al. [5]. Since that moment extensive research was done on active contour models for boundary detection. Caselles, Kimmel and Sapiro [1] and Kichenassamy, et al. [6] proposed new enhanced versions of the snake model called the geodesic/geometric active contour (GAC) model. The active contour consists of evolving a contour in images towards the boundaries of objects. It is based on strong mathematical properties and efficient numerical schemes based on the level set method [13]. The basic snake model is a deformed initial shape affected by external constraint

forces and image forces that pull it towards object contours, and internal forces that resist the deformation. The image forces serve to push the snake to the boundary of the objects, while the external constraint forces impose to put the snake close to the required local minimum [5].

Figure 5 shows an original image of a copper specimen obtained from THERMO-WET measurement system at a high temperature (1080C). Active contour method segments the 2D grayscale image, into a binary image where the background is black and the foreground is white using a mask, which is a binary image that gives an initial state of the active contour segmentation. The active contour method segments the image by evolving the contour for *n* number of iterations, which is set as a numeric scalar. The algorithm will stop the evolution if the contour position in the current iteration is the same as one of the contour positions from the most recent five iterations or if the maximum number of iterations is reached. The method used for segmentation is the Chan-Vese method [3] with a ContractionBias parameter equal to 0.3, which corresponds to the tendency of the contour to shrink inwards. Application of the active contour method results in the black and white image (Figure 5b). At this stage we use the Sobel edge detection method to find the edge of the binary image to get the contour of the droplet (Figure 5c).

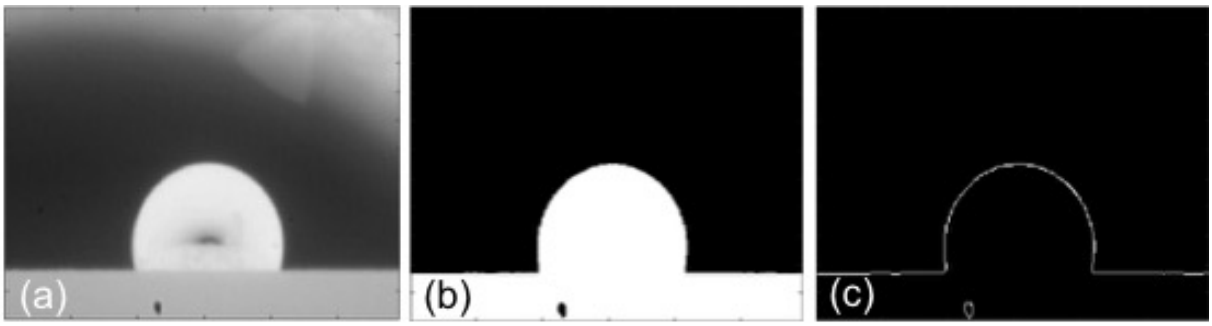


Figure. 5: Original image, copper, 1080C (a); segmented image using the active contour method (b); the contour of the image (c)

3. Results

Once the features are extracted as described in the previous section, we started selecting the most discriminating features. We found that the standard deviation is the most relevant feature

Table 1: The width and the height of the droplet for the three methods, to be compared with the true value estimated manually

among the other features since it shows clear boundaries compared with other features (see figure 3) By comparing this result with the image in Figure 4c after applying sobel operator, and the image in Figure 5c that relies on the active contour segmentation described in the previous section, one can estimate the most adapted method to extract as accurate as possible the geometric characteristics of the droplet.

One can obviously notice that the active contour gives identical width (d) and height (h) with the true value. Table 1 shows the width (d) and the height (h) of the droplet, for the three tested methods. The comparison is done as the evaluation of the droplet width (d) and the height (h) mentioned in section 2.1 and values are shown in table 1.

Table 1: The width and the height of the droplet for the three methods, to be compared with the true value estimated manually

	Sobel (magnitude)	Standard deviation	Active contour	The true value
Width (d)	117	114	113	113
Height (h)	60	60	59	59

4. Conclusion

This paper has presented a comparative study of three edge detection techniques to detect the boundary of a metallic droplet. The comparison shows that the active contour segmentation method is the most accurate method compared with the two other methods (i.e. features ex- traction and selection technique, and sobel filter). Note that values are compared with manual estimations that are considered as the true value. So far, the comparison has considered are one tested image of a droplet. More images should be tested in the near future to unambiguously conclude about the most appropriate segmentation methods to select for droplet characterization.

5. References

1. Casselles V., Kimmel R., Sapiro G.: *Geodesic Active Contours*, International Journal of Computer Vision, 1997, 22(1), 61-79.

2. CSE567.: *Computer Vision: chapter 7*, University of Washington, 2000,1-14.

3. Dettori L., Semler L.: *A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography*, Computers in biology and medicine, 2007,37(4), 486-498.

4. Junyan L., Qingju T., Yang W., Yumei L., Zhiping Z.: *Defects' geometric feature recognition based on infrared image edge detection*, 2014,67, 387-390.

5. Kass M., Witkin A., Terzopoulos D.: *Snakes: Active Contour Models*, International Journal of Computer Vision, 1998 1(4), 321-331.

6. Kichenassamy S., Kumar A., Olver P., Tannenbaum A., Yezzi A.: *Conformal Curvature Flows: from Phase Transitions to Active Vision*. Archive for Rational Mechanics and analysis, 1996, 134(3), 275-301.

7. Liang Y., Zhang M., Browne W.N.: *Image feature selection using genetic programming*, Engineering Applications of Artificial Intelligence, 2017, 62, 96-108.

8. Mathur N., Mathur S., Mathur D.: *A Novel Approach to Improve Sobel Edge Detector*, Procedia Computer Science, 2016,93, 431-438.

9. Petrou M., Garcia-Sevilla P.: *Image Processing: Dealing with Texture*, 2006, Vol. 1.

10. Sankowski D., Senkara J., Strzecha K., Jezewski S.: *Automatic Investigation of Surface Phenomena in High Temperature Solid and Liquid Contacts*, 18th IEEE Instrumentation And Measurement Technology Conference, Budapest, 2001,2, 1397-1400.

11. Shannon C.E.: *A Mathematical Theory of Communication*, Alcatel-Lucent Journal, 1948,27, 623-656

12. Strzecha K., Fabiajanska A., Koszmider T., Sankowski D.: *Selected Algorithms of Quantitative Image Analysis for Measurements of Properties Characterizing Interfacial Interactions at High Temperatures*, In Computer Vision in Robotics and Industrial Applications, World Scientific, 2014339-382.

13. Tian D.: *A Review on Image Feature Extraction and Representation Techniques*, International Journal of Multimedia and Ubiquitous Engineering, 2013,8(4), 385-396.

14. Zhang Y.,Li S.,Wang T.,Zhang Z.: *Divergence-based feature selection for separate classes*, Neurocomputing, 2013,101, 32-42.

Tomasz RYMARCZYK [12], Paweł TCHÓRZEWSKI[1]

Sensor Networks for Flood Embankment and Landfill in Tomography Imaging

This paper presents a nondestructive method to examine objects such as landfills and flood embankments. It will be innovative solution for evaluation study of the buildings and tanks, both in terms of the measuring method as the reconstruction algorithm. The application of modern tomographic techniques in conjunction with reconstruction algorithms will allow to perform a non-invasive and very accurate spatial assessment of the moisture. Electrical tomography includes many techniques for tomographic imaging of electrical parameters of an object placed in an examined area. Although many methods of evaluating dampness and damage already exist, there is no universally for use under various conditions of measurement.

1. Introduction

This paper presents a new method examining the flood embankment and the landfill dampness using the electrical impedance tomography (EIT) [1,2,8,13-14]. Numerical and optimization methods were based on gradient techniques [5,7-9,10,12]. The discussed technique can be applied to the solution of inverse problems in EIT. The 3D algorithms to identify unknown conductivities were implemented.

Flood is one of the most frequent taking place natural disaster, which cause many tragedies. The one way to enclose range of pouring areas near the river is raising flood embankments. By using them, it is possible to reach a higher level of freshet over main river-bed and crack down on flood. However, this fact has impact on rapid increasing amount of water in valley between embankment and the acceleration of the water flow and the flood wave. It causes erosion of upstream face of embankment (and landfill too) and it may destroy it [3,15]. Additionally, considering too little force of filtering capabilities of the embankment body, high water may lead to the grow of unsaved filtering ways which may lead to partial destruction of the embankment. For these reasons, designing embankments on the grounds of their location, shape and material is extremely complicated. Despite significant progress in the field of designing safe embankments there is a possibility of appearance of unexpected changes. Therefore, cheap and foolproof device is essential in monitoring the state of flood embankments or landfills before the visible effects of the destructive force of natural disaster.

2. Models and measurements

Electrical tomography (impedance and capacitance) provides a non-invasive technique for measuring different kinds of objects [4,5,11,16-19]. EIT determines the recovery of the conductivity of the interior of the examined object from a knowledge of currents and voltages applied to its surface. The data acquisition system collects the measured voltage from surface electrodes and then the data are processed. Conventional data acquisition systems require hardware to measure voltage, filter, demodulate and convert them into a digital form, and a signal processing unit to transfer the data to computer. The example of the flood embankment is given in Figure 1. Figure 2 shows the model of a landfill. The examples of different inserted electrodes into the embankments were presented in Figure 3. Figures 4 and 5 show special

[1)] Research and Development Center, Netrix S.A., Związkowa 26, 20-148 Lublin, POLAND

[2)] University of Economics and Innovation in Lublin, Projektowa 4, 20-209 Lublin, POLAND

multiply electrodes to measure inside the objects. The measurement system with the EIT device, transfer data to the cloud computing and image reconstruction was presented in Figure 6.

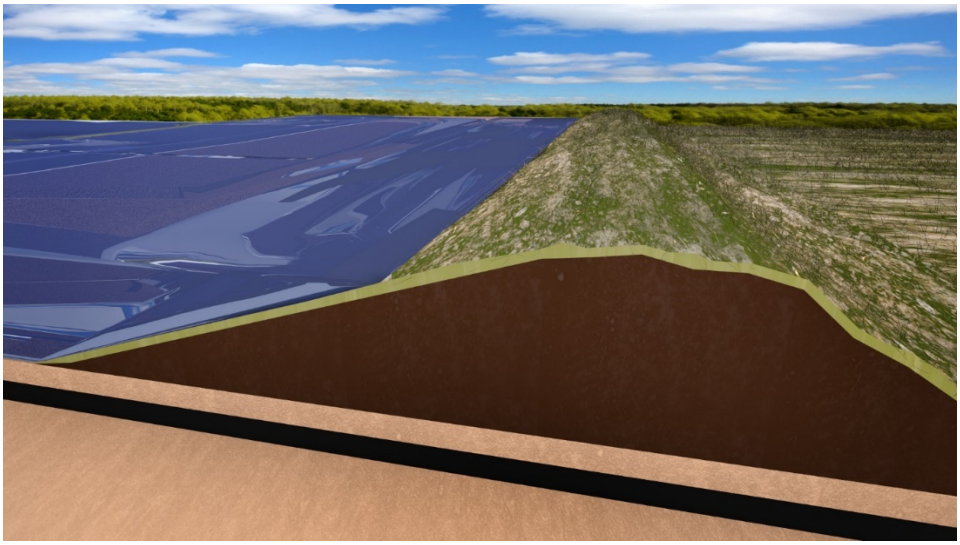


Fig.1. The example of the flood embankment.

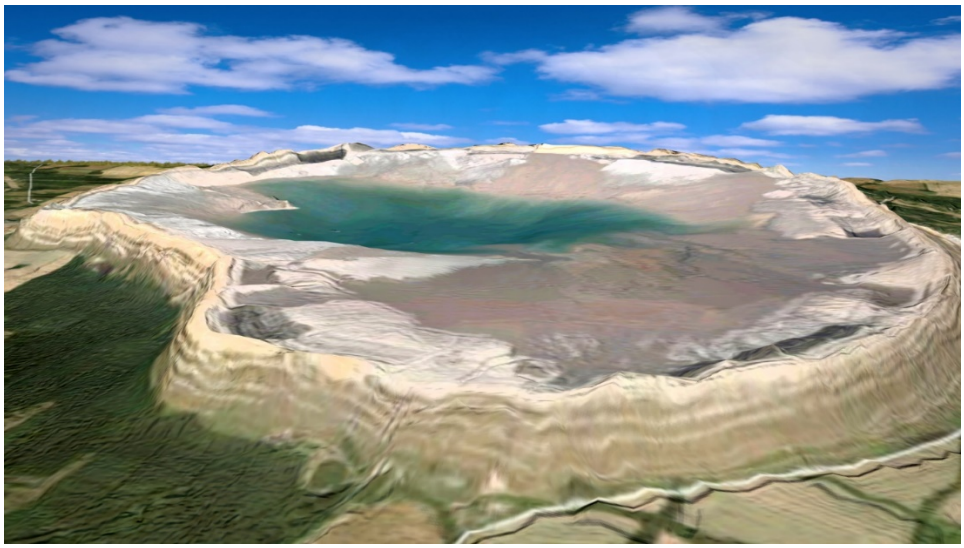


Fig.2. The example of the landfill.

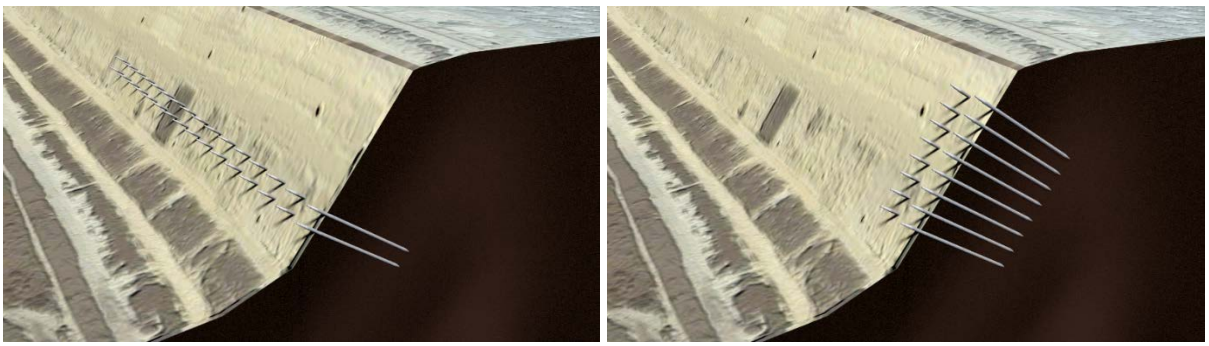


Fig.3. The examples of electrodes.

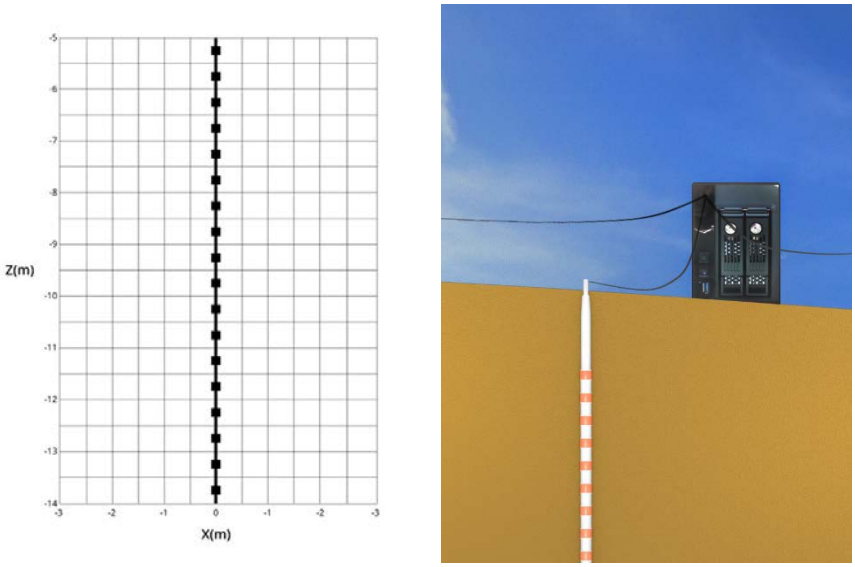


Fig.4. The example of multiply electrode.

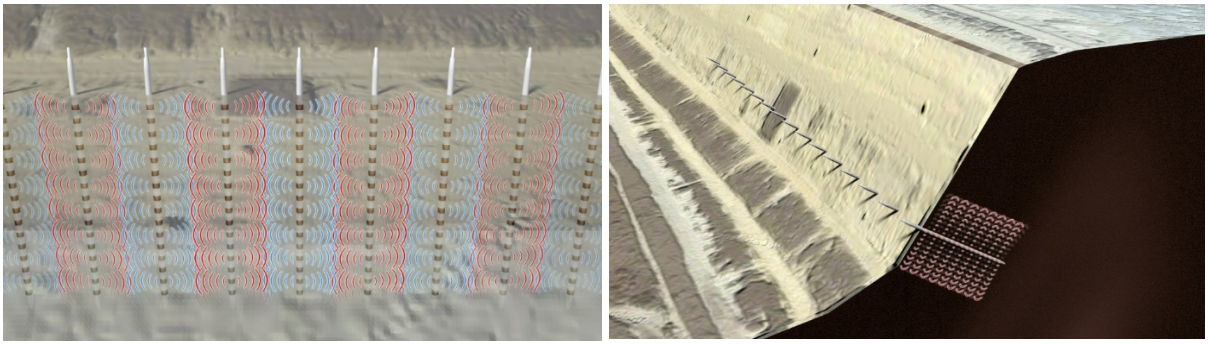


Fig.5. The example of electrodes.

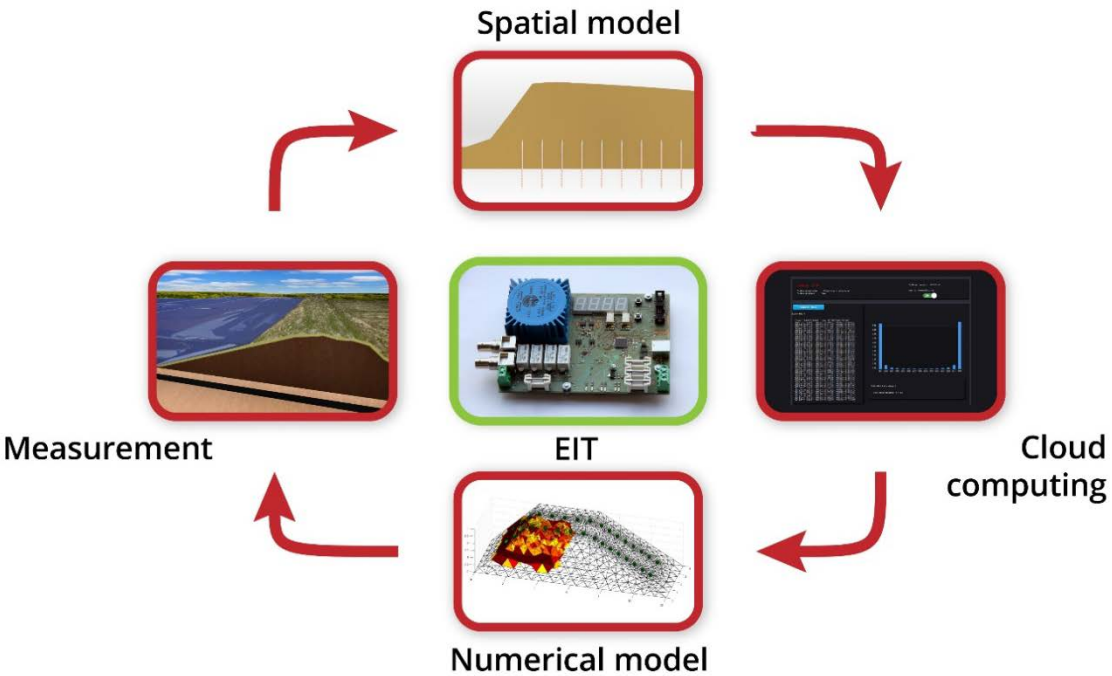


Fig.6. The measurement system.

3. Image reconstruction

In examples reported below, several EIT numerical measurement models of the flood embankment were presented. The conductivity of searched objects is unknown. The forward problem was solved by the finite element method. Figure 7 present the model with multiply electrodes. The embankment with 16 electrodes, simulated areas inside (geometrical mesh, original and reconstruction model) was presented in Figure 8. The Solution was obtained by Gauss-Newton method with Tikhonov and Laplace regularization. Figure 9 presents the different model of the embankment with 32 electrodes.

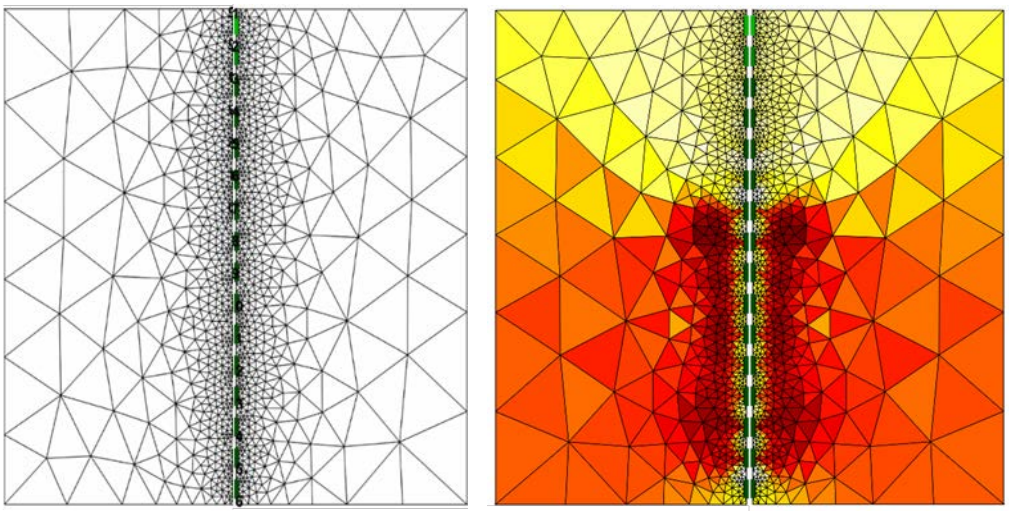


Fig.7. The example with multiply electrodes.

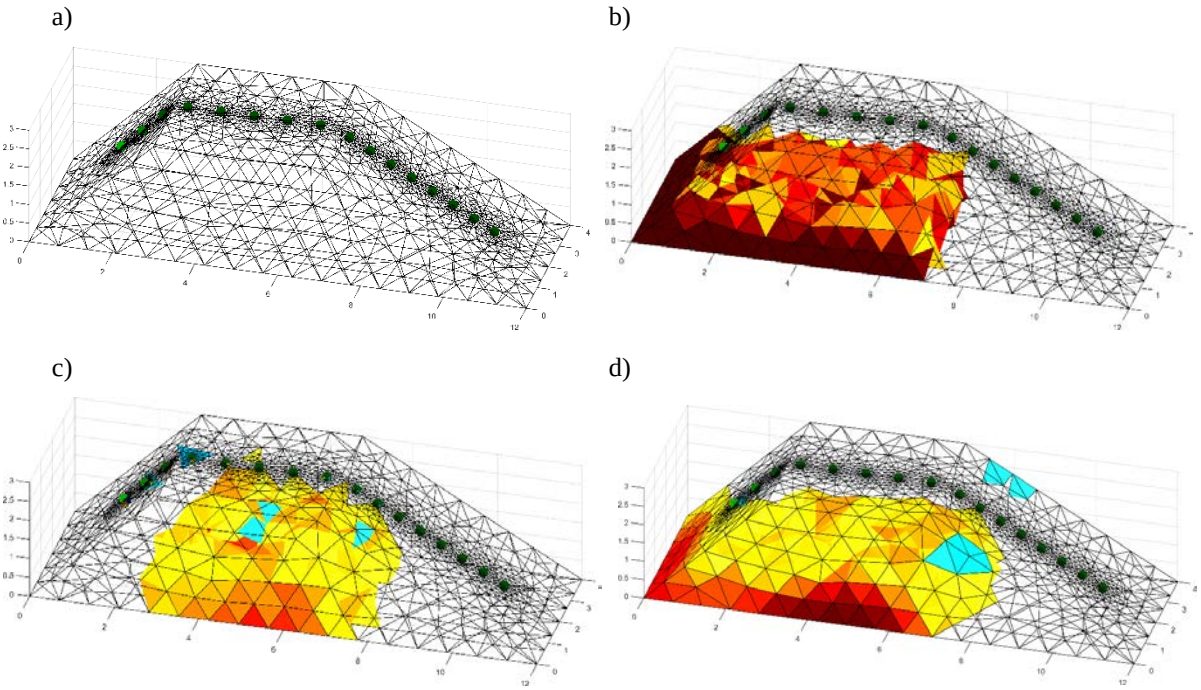


Fig. 8. The embankment: a) mesh model b) model of object, c) Gauss-Newton with Tikhonov regularization, d) Gauss-Newton with Laplace regularization.

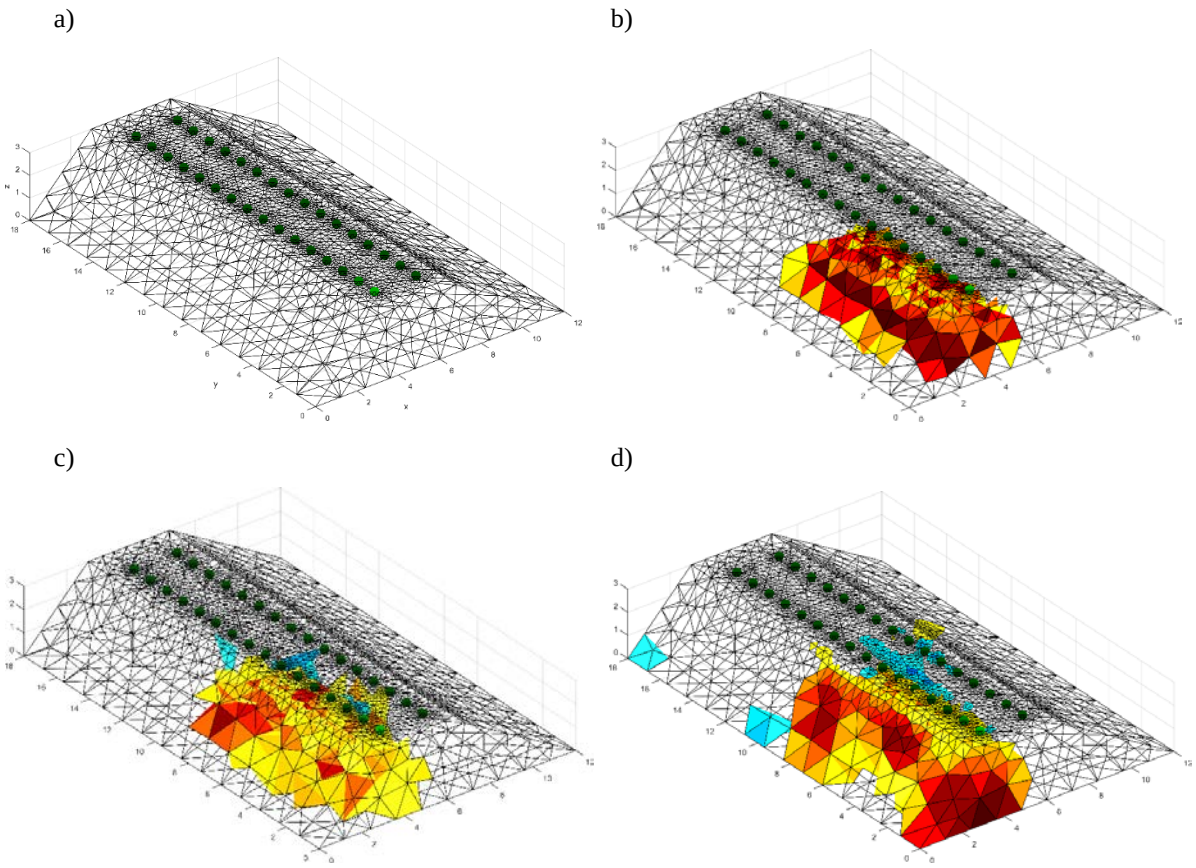


Fig. 9. The embankment: a) mesh model b) model of object, c) Gauss-Newton with Tikhonov regularization, d) Gauss-Newton with Laplace regularization.

4. Conclusion

The nondestructive method based on the electrical impedance tomography was presented to examine condition of flood embankments and landfills. The presented solution has been applied very successfully in such models. Even though this study reveals novel and promising results. This technique is very useful to solve the inverse problem in different kind of objects. There was described the method and verified its properties with simulations data. Different models were used to represent the numerical solution. Different type and number of electrodes were evenly distributed along and across the border of the grid estimation and numerical phantom. 16 and 32 sets of electrodes were evenly assigned to the experimental phantoms and their models. This study provides new and promising results.

References

1. Adler A., Lionheart W., *Uses and abuses of EIDORS: An extensible software base for EIT*, Phys. Meas., vol. 27, pp. 25–42, 2006.
2. Borcea L., *Electrical impedance tomography*, Inverse Problems, vol. 18, pp. 99–136, 2002.
3. Duda K., Adamkiewicz P., Rymarczyk T., *Nondestructive Method to Examine Brick Wall Dampness*, International Interdisciplinary Phd Workshop 2016, pp. 68-71, 2016.
4. Fiderek P., Jaworski T., Banasiak R., Nowakowski J., Kucharski J., Wajman R., *Intelligent system for the two-phase flows diagnosis and control on the basis of raw 3D 3D ECT data*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 7, Issue 1, pp. 17-23, 2017; DOI: 10.5604/01.3001.0010.4576.

5. Filipowicz S.F., Rymarczyk T., *Measurement Methods and Image Reconstruction in Electrical Impedance Tomography*, Przeglad Elektrotechniczny, vol. 88, Issue 6, pp.247-250. 2012.
6. Garbaa H., Jackowska-Strumiłło L., Grudzień K., Romanowski A., *Simulation of gravitational solids flow process and its parameters estimation by the use of electrical capacitance tomography and artificial neural networks*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 6, Issue 2, pp. 34–37, 2016.
7. Gola A., Świć A., *Computer-Aided Machine Tool Selection for Focused Flexibility Manufacturing Systems Using Economical Criteria*, Actual Problems of Economics, vol. 124, Issue 10, pp. 383-389, 2011.
8. Holder D. S., *Introduction to biomedical electrical impedance tomography Electrical Impedance Tomography Methods, History and Applications*, Bristol, Institute of Physics, 2005.
9. Kłosowski G, Kozłowski E, Gola A., *Integer linear programming in optimization of waste after cutting in the furniture manufacturing*. Advances in Intelligent Systems and Computing 2018; vol. 637, pp. 260-270, 2018.
10. Kryszyn J., Smolik W., *Toolbox for 3D Modelling and Image Reconstruction in Electrical Capacitance Tomography*, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska - Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, ISSN: 2083-0157, vol. 1, pp. 137-145, 2017.
11. Kryszyn J, Wróblewski P, Stosio M, Wanta D, Olszewski T, Smolik W, *Architecture of EVT4 data acquisition system for electrical capacitance tomography*, Measurement: Journal of the International Measurement Confederation, vol. 101, pp. 28-39, 2017.
12. Polakowski, K., Filipowicz, S.F., Sikora, J., Rymarczyk, T., *Quality of imaging in multipath tomography*, Przeglad Elektrotechniczny, vol. 85, Issue 12, pp. 134-136, 2009.
13. Rymarczyk T., *Characterization of the shape of unknown objects by inverse numerical methods*, Przeglad Elektrotechniczny, vol. 88, Issue 7B, pp. 138-140, 2012.
14. Rymarczyk T., Adamkiewicz P., Rymarczyk P. et al., *Monitoring Flood Embankment by Electrical Impedance Tomography*, International Interdisciplinary Phd Workshop 2016, pp. 63-67, 2016.
15. Rymarczyk T., *Using electrical impedance tomography to monitoring flood banks*, International Journal of Applied Electromagnetics and Mechanics, vol. 45, pp. 489–494, 2014.
16. Smolik W., Kryszyn J., Olszewski T., Szabatin R., *Methods of Small Capacitance Measurement in Electrical Capacitance Tomography*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, IAPGOS, ISSN: 2083-0157, vol. 1, pp. 105-110, 2017.
17. Smolik W., *Forward Problem Solver for Image Reconstruction by Nonlinear Optimization in Electrical Capacitance Tomography*, Flow Measurement and Instrumentation, vol. 21, pp. 70–77, 2010.
18. Soleimani M., Mitchell CN, Banasiak R., Wajman R., Adler A., *Four-dimensional electrical capacitance tomography imaging using experimental data*, Progress In Electromagnetics Research, vol. 90, pp. 171-186, 2009.
19. Wajman, R., Banasiak, R., *Tunnel-based method of sensitivity matrix calculation for 3D-ECT imaging*. Sensor Review, vol. 34, pp. 273–283, 2014. doi:10.1108/SR-06-2013-692

Karol DUDA[1], Tomasz RYMARCZYK, [12], Piotr BOŻEK[1], Paweł TCHÓRZEWSKI[1], Jan SIKORA[3,4]

EIT Hardware Solution for Data Acquisition

This article presents the design of the device to examine the objects by electrical impedance tomography. There were presented models of bricks with measurement system. There was discussed the method of determining the conductivity. The basic information about the built model system is given. Electrical impedance tomography, which is based on measuring potential difference, can be used to calculate conductivity.

1. Introduction

Electrical Tomography (ET) provides a novel non-invasive technique for measuring different patterns [2,4,7,8,13-16]. Aim to assess the feasibility and potential usefulness for detecting changes in patterns. New non-destructive measurements methods were tested. Electrical Impedance Tomography (EIT) is an imaging technique for detecting the internal conductivity distribution of an object by voltage measurements taken by an exterior electrode. Many different optimization techniques can be used to solve the problem [5,6,9]. In EIT, there was acquired data on the edge. There is calculated the potential distribution inside the region Ω under given boundary conditions and full information about region under consideration [10-12]. Most EIT works have used it as a dynamic imaging method, in which images of the impedance change compared to a baseline condition are obtained. EIT is the recovery of the conductivity of the interior of the examined object from a knowledge of currents and voltages applied to its surface. The data acquisition system collects the measured voltage from electrode and then the data is processed. Conventional data acquisition systems require hardware to measure voltage, to filter, demodulate and convert to digital form, and a signal processing unit to transfer the data to computer [1].

2. EIT device

There was created a new device after achieving good research results of measurements of the EIT. Its aim was to verify the repeatability of test results by eliminating laboratory equipment, and to validate the use of simple and cheap electronics to the structure of the EIT. Figure 1 presents the scheme of the block diagram operation of the measuring device. Measurements of the potential were performed at different angles of projection whereby the information needed to determine an approximate distribution of conductivity inside the object is obtained. The main goal of construction the EIT device was verification of measure ability device built from possibly simplest elements. A device should be small, low power consuming, should be compatible with external control digital signals and an external drop voltage meter. A device consists of voltage controlled current source and Analog keys. A current source has to force voltage drops at examine area boundaries, that it could be measured. A current source was built of two blocs: sine-wave generator and voltage controlled current source in Howland circuit. A sine-wave was generated by integrated circuit ICL8038.

1) Research and Development Center, Netrix S.A., Związkowa 26, 20-148 Lublin, POLAND

2) University of Economics and Innovation in Lublin, Projektowa 4, 20-209 Lublin, POLAND

3) Lublin University of Technology, Institute of Electronics and Information Technology, Nadbystrzycka 38d, 20-618, Lublin, Poland

4) Electrotechnical Institute, ul. Pożaryskiego 28, 04-703 Warszawa

It needs minimum amount of external passive elements: timing capacitor, level adjust resistor and resistive voltage divider that sets output signal frequency.

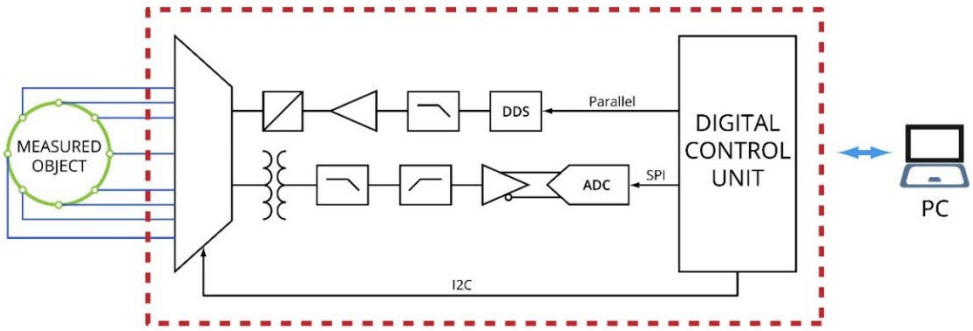


Fig. 1. Scheme of the block diagram operation of the measuring device.

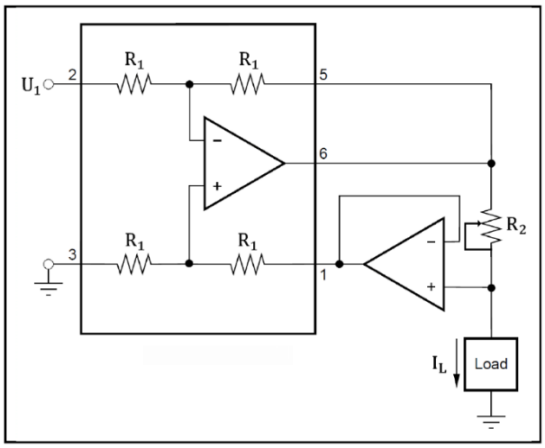


Fig. 2. Principle of operation of the current source.

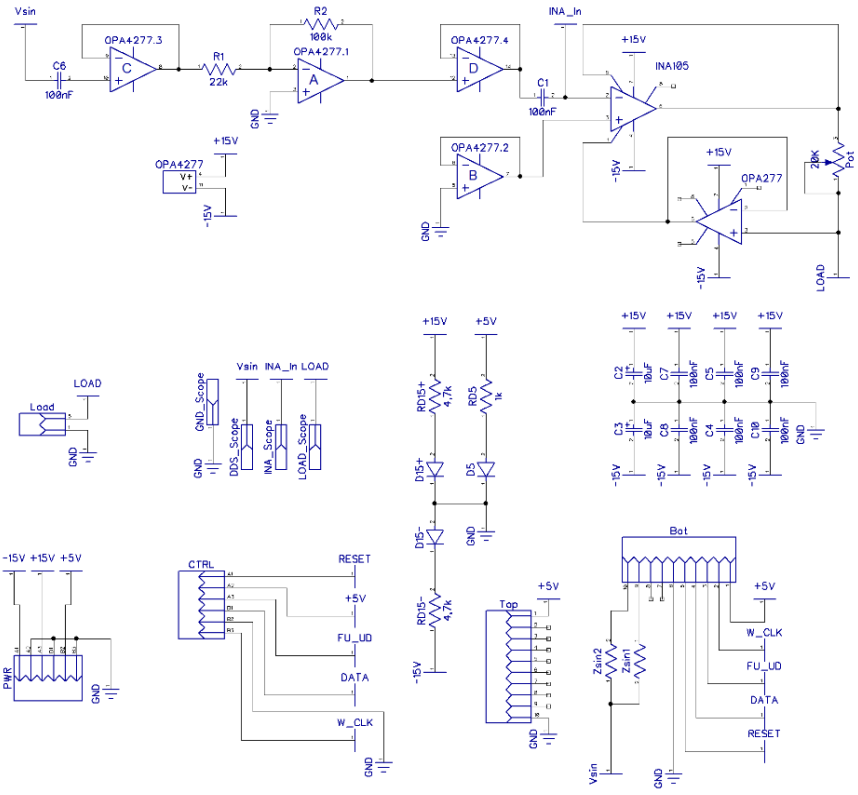


Fig. 3. Schematic diagram of the current source.

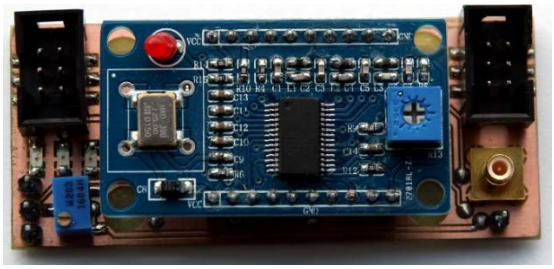


Fig. 4. Current source with injected generator DDS AD9850.

A circuit was design in way that could change output frequency between 1 kHz and 50 kHz by switching jumper. A sine-wave voltage generator controlled current source in Howland circuit. It was built from four operational amplifiers closed in one package LMC6484 and six precision potentiometers. First operational amplifier was configured in a way that it was possible to regulate current output amplitude. The second one, there was configured as voltage follower for reducing input impedance on the current source circuit. The current source presented below is based on the Howland layout, where a number of modifications have been made (Figures 2-4). A DDS generator module containing an AD9850 chip is required to operate the device, which is used to generate a sinusoidal signal that controls the load current. The module is plugged directly into the device. The frequency and amplitude of the DDS output is programmed via the serial interface or the parallel bus. In addition, the current source has a multi-turn potentiometer for manually adjusting the output current. The device has an output allowing communication with the built-in generator module. The whole system is powered from a symmetrical +/- 15VDC source.

3. Models, Measurements, Simulation and Reconstruction

In examples reported below, models and image reconstructions are presented. Figure 5 shows 2D and 3D mesh model of the brick. Ceramic bricks with connected 16 and 32 electrodes are presented in Figure 6. The EIT measuring strip contains 32 electrodes placed in, printed on a 3D printer, each frame (fig. 7). The frame design allows the elastic waist to be stretched through the waist so that the elements can be covered in different shapes and sizes with equal spacing between the electrodes. The electrodes are made of two-sided FR4 laminate, where the side adjacent to the object is connected to the SMB pin, which allows the electrode to be connected to the multiplexer while the other side is connected to the socket screen. The electrode is made in such a way that instead of holes on the SMB socket housing, only round recesses are made. The hole is made only on the socket pin. This causes the area adjacent to the area to be examined to have no disturbance, and the area measured is not connected to the screen. Figure 8 shows the image reconstruction of the brick (with hole and moisture) by Gauss-Newton. Figure 9 presents the image reconstruction by Total Variation and level set method (green line). Figure 10 and 11 show the image reconstruction by Gauss-Newton method and level set method (with different area of moisture). The image reconstruction by LSM gives the best result.

4. Conclusion

There was presented the construction of the EIT device, the model of the tested object and the image reconstruction by using the Gauss-Newton method, level sets and Total Variation. The presented methods have been applied very successfully in many areas of the scientific modelling. These approaches were based on sensitivity analysis. An efficient algorithm for solving the forward and inverse problems would also improve a lot of the numerical performances of the proposed methods. The presented methods gave good results in the

examples. The future work will be based on implementation statistical algorithms to solve the inverse problem and modifications of the measurement system.

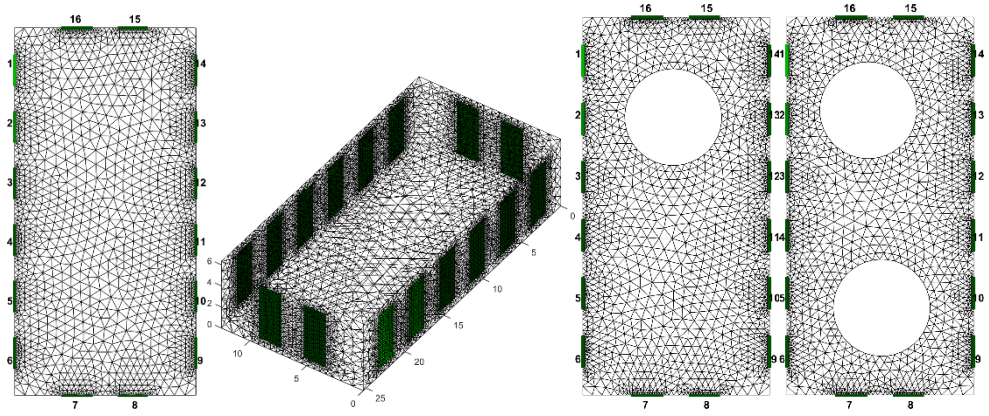


Fig. 5. 2D and 3D mesh model of the brick.

a) b) c)

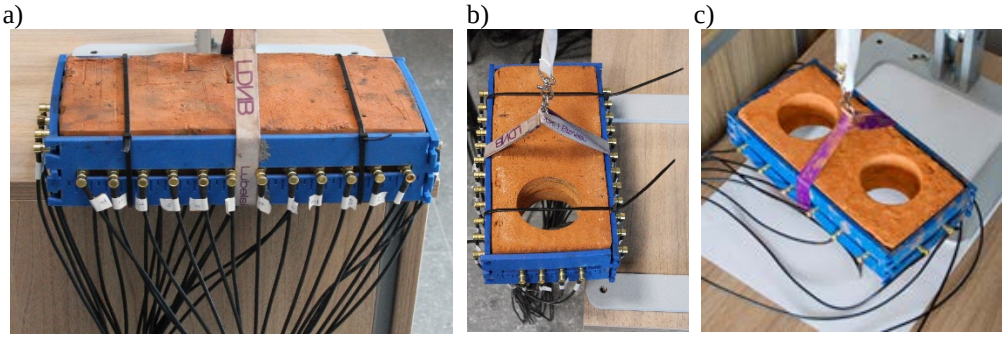


Fig. 6. A full ceramic brick with connected 32 electrodes.(a), (b) a one hole ceramic brick with connected 32 electrodes, (c) a two hole ceramic brick with connected 16 electrodes.

Fig. 7. Belt with 32 electrodes for EIT measurements.

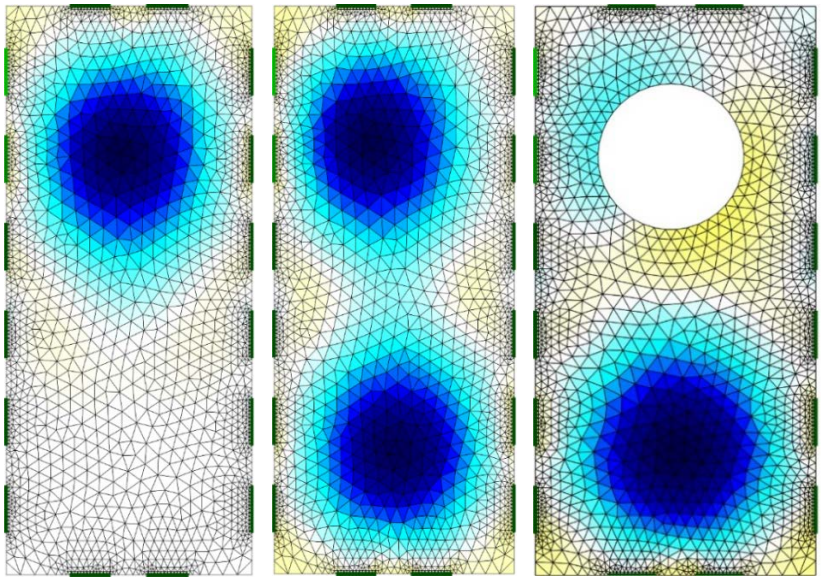


Fig. 8. The image reconstruction by Gauss-Newton method GNM (blue colour - image reconstruction).

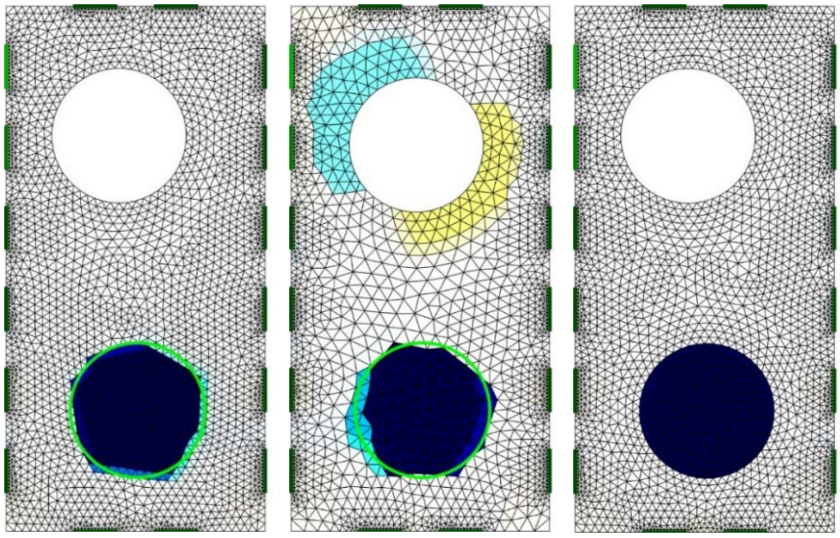


Fig. 9. The image reconstruction by Total Variation TV and level set method LSM (blue colour - image reconstruction by TV, green line – reconstruction by LSM).

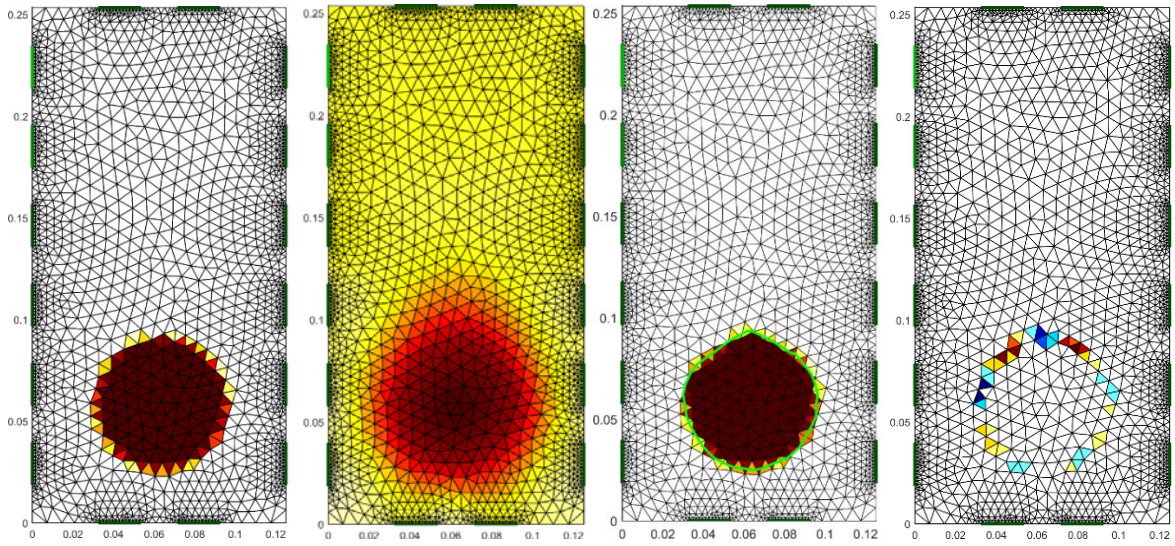


Fig. 10. The image reconstruction by Gauss-Newton method GNM and Level Set method LSM (brown colour - image reconstruction by GNM, green line – reconstruction by LSM).

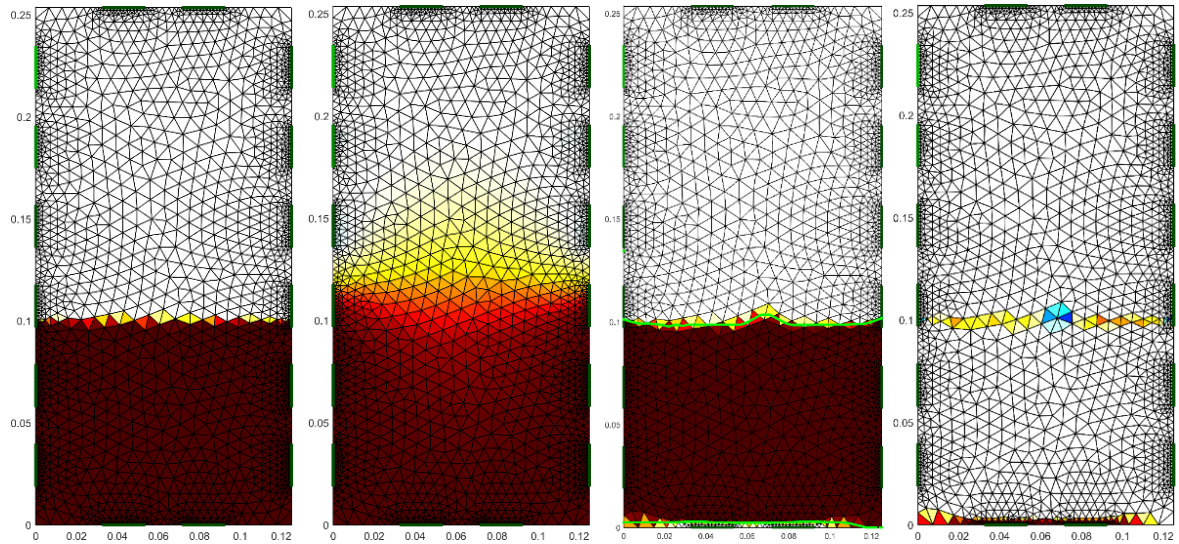


Fig.11. The image reconstruction by Gauss-Newton method GNM and Level Set method LSM (brown colour - image reconstruction by GNM, green line – reconstruction by LSM).

References

1. Duda K., Adamkiewicz P., Rymarczyk T., *Nondestructive Method to Examine Brick Wall Dampness*, International Interdisciplinary Phd Workshop 2016, pp. 68-71, 2016.
2. Fiderek P., Jaworski T., Banasiak R., Nowakowski J., Kucharski J., Wajman R., *Intelligent system for the two-phase flows diagnosis and control on the basis of raw 3D 3D ECT data*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 7, Issue 1, pp. 17-23, 2017; DOI: 10.5604/01.3001.0010.4576.
3. Filipowicz S.F., Rymarczyk T., *Measurement Methods and Image Reconstruction in Electrical Impedance Tomography*, Przeglad Elektrotechniczny, vol. 88, Issue 6, pp.247-250. 2012.
4. Garbaa H., Jackowska-Strumiłło L., Grudzień K., Romanowski A., *Simulation of gravitational solids flow process and its parameters estimation by the use of electrical capacitance tomography and artificial neural networks*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 6, Issue 2, pp. 34–37, 2016.
5. Gola A., Świć A., *Computer-Aided Machine Tool Selection for Focused Flexibility Manufacturing Systems Using Economical Criteria*, Actual Problems of Economics, vol. 124, Issue 10, pp. 383-389, 2011.
6. Kłosowski G, Kozłowski E, Gola A., *Integer linear programming in optimization of waste after cutting in the furniture manufacturing*. Advances in Intelligent Systems and Computing 2018; vol. 637, pp. 260-270, 2018.
7. Kryszyn J., Smolik W., *Toolbox for 3D Modelling and Image Reconstruction in Electrical Capacitance Tomography*, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska - Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, ISSN: 2083-0157, vol. 1, pp. 137-145, 2017.
8. Kryszyn J, Wróblewski P, Stosio M, Wanta D, Olszewski T, Smolik W, *Architecture of EVT4 data acquisition system for electrical capacitance tomography*, Measurement: Journal of the International Measurement Confederation, vol. 101, pp. 28-39, 2017.
9. Polakowski, K., Filipowicz, S.F., Sikora, J., Rymarczyk, T., *Quality of imaging in multipath tomography*, Przeglad Elektrotechniczny, vol. 85, Issue 12, pp. 134-136, 2009.
10. Rymarczyk T., *Characterization of the shape of unknown objects by inverse numerical methods*, Przeglad Elektrotechniczny, vol. 88, Issue 7B, pp. 138-140, 2012.
11. Rymarczyk T., Adamkiewicz P., Rymarczyk P. et al., *Monitoring Flood Embankment by Electrical Impedance Tomography*, International Interdisciplinary Phd Workshop 2016, pp. 63-67, 2016.
12. Rymarczyk T., *Using electrical impedance tomography to monitoring flood banks*, International Journal of Applied Electromagnetics and Mechanics, vol. 45, pp. 489–494, 2014.
13. Smolik W., Kryszyn J., Olszewski T., Szabatin R., *Methods of Small Capacitance Measurement in Electrical Capacitance Tomography*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, IAPGOS, ISSN: 2083-0157, vol. 1, pp. 105-110, 2017.
14. Smolik W., *Forward Problem Solver for Image Reconstruction by Nonlinear Optimization in Electrical Capacitance Tomography*, Flow Measurement and Instrumentation, vol. 21, pp. 70–77, 2010.
15. Soleimani M., Mitchell CN, Banasiak R., Wajman R., Adler A., *Four-dimensional electrical capacitance tomography imaging using experimental data*, Progress In Electromagnetics Research, vol. 90, pp. 171-186, 2009.
16. Wajman, R., Banasiak, R., *Tunnel-based method of sensitivity matrix calculation for 3D-ECT imaging*. Sensor Review, vol. 34, pp. 273–283, 2014. doi:10.1108/SR-06-2013-692

Tomasz RYMARCZYK, [12], Przemysław ADAMKIEWICZ[1], Krzysztof POLAKOWSKI[3], Jan SIKORA[4,5]

Algorithms for Imaging in Ultrasound and Radio Tomography

The presented results show that the obtained tomographic imaging can be a faithful image of the geometry of the examined area, which allows to determine its location in the analyzed space. It is important that the correct selection in the calculation of singular values because even small differences in these values can have a significant impact on the precision of the modeling of the modeled object.

1. Introduction

There are many tomographic techniques to exam objects [1,6,9,13-17]. This paper presents an attempt to briefly present selected mathematical methods for the reconstruction of tomographic images obtained with the ultrasonic / radio transmission tomography. Radio tomographic imaging (RTI) is a new technology that enables the device-free localization and tracking of people located within the deployment area of a wireless network. In radio tomography systems, the only sensor used is the radio, which measures the received signal strength (RSS) of signals it receives. Most importantly, these systems do not require people to carry any device, sensor, or tag to participate in the localization effort. The potential applications of these device-free localization systems are many, including ambient assisted living, security systems, rescue operations, and occupational safety systems in industrial areas, among others.

2. Transmission tomography and mathematical methods in image reconstruction

Most algorithms in the algebraic reconstruction techniques (ART) it is based on the methodology of approximation of functions by finite length series. In these methods it is assumed that the reconstructed image consists of a finite number of elements. The image is constructed by means of an algorithm that discretizes a given area into n squares of size $l \times 1$, whose geometric center corresponds to one pixel in the reconstructed image [2,3-5,8,10]. This means that to reconstruct the image, the same square grid should be applied to the object $f(x, y)$ and the image being created. The example grid is shown in Figure 1. The basics of algebraic reconstruction techniques were described by S. Kaczmarz [7] long before computer tomography and in its original version did not concern the reconstruction of the image from projection. It is assumed that each cell is described as a fixed function $f(x, y)$ and in the case of transmission of ultrasound tomography and radio tomography determines the desired ultrasound wave velocity or signal strength in a given cell.

[1)] Research and Development Center, Netrix S.A., Związkowa 26, 20-148 Lublin, POLAND

[2)] University of Economics and Innovation in Lublin, Projektowa 4, 20-209 Lublin, POLAND

[3)] Warsaw University of Technology, ul. Koszykowa 75, 00-662 Warsaw, POLAND

[4)] Lublin University of Technology, Institute of Electronics and Information Technology, Nadbystrzycka 38d, 20-618, Lublin, Poland

[5)] Electrotechnical Institute, ul. Pożaryskiego 28, 04-703 Warszawa

Let's define f_j as the constant value of each j cell, and n as the total number of pixels (square grid cells). For arithmetic reconstruction techniques, the radius s can be defined in two ways. It can be assume that the radius is a line with a width equal to the length of the side of the square. This means that we use the entire beam of parallel rays to consider (Ray-sum) beam passing through the area under consideration. It can be also assume that the radius has a pixel width (the point being the geometric center of the square cell of the grid). In the ultrasonic tomography based on the measurement of the transition times between the transmitters and receivers, the combined velocity values on the i paths of the ultrasound beams between the transmitters and receivers, determined by projections s_i. The relation between f_i and s_i is given:

$$\sum_{j-1}^{n} w_{ij} f_j = s_i, \qquad i = 1, 2, \dots, m, \tag{1}$$

Where m is the total number of rays (in all projections) and $w_i i_j$ is the weight factor determining the part of the value for the j cell relative to the entire measured value along the i beam.

a) b)

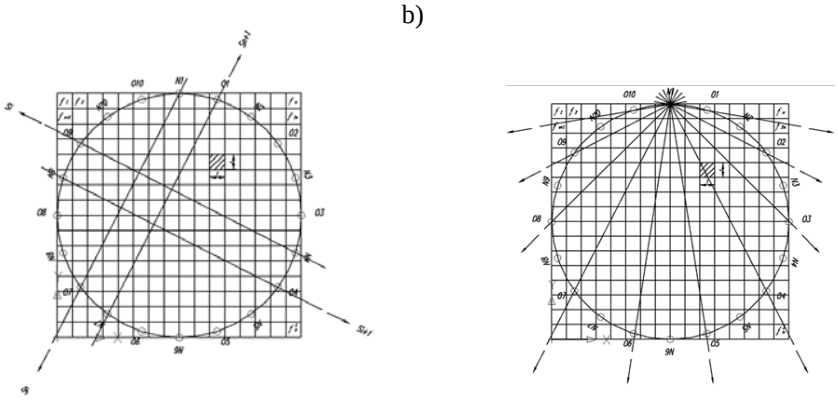


Fig.1. The Square grid applied to the object under investigation - the basis of all arithmetic methods of image reconstruction; a) rays in parallel geometry; b) divergent rays; N-transmitter, O- receiver,

The weight factor may be equal to the ratio of the square in the grid of the field by the radius of a given width to the value of the whole square of l^2 or, if we take a radius of pixel width, it may be equal to the ratio of the length of the radius of the square to the diagonal of the square. These dependencies are different from zero only for square grid through which the given beam passes. For all pixels whose beam does not intersect, the factor is zero (shown in Figure 2). If the number of rays m and the number of squares n of the grid are small then the conventional methods of matrix inversion can be used to solve the system described by given formula. With large sizes of m and n it is not possible to use direct matrix inversion.

To solve large systems of equations, very effective iterative numerical methods are used, based on the "projection method" proposed by S. Kaczmarz [7]. To explain the method of obtaining results, the previous formula should be presented in the form of a system of equations:

$$\begin{aligned} w_{11}f_1 + w_{12}f_2 + w_{13}f_3 + \cdots + w_{1n}f_n &= s_1, \\ w_{21}f_1 + w_{22}f_2 + w_{23}f_3 + \cdots + w_{2n}f_n &= s_2, \\ &\cdots \\ w_{m1}f_1 + w_{m2}f_2 + w_{m3}f_3 + \cdots + w_{mn}f_n &= s_m. \end{aligned} \tag{2}$$

Reconstructing an image using a square grid with n-pixels gives image with n degrees of freedom. Therefore, the image represented by [$f_1, f_2, \dots, f_n$] can be treated as a single point in n

dimensional space. This is due to the fact that in this space each of equations (2) is a hyperplane [5,8,10,11,12].

3. SVD method for solving overdetermined linear equations

The problem of solving an overlapping set of linear equations (3) is often complicated, especially in the case of the existence of measurement errors and it also can be a contradictory system:

$$\boldsymbol{Wf} = \boldsymbol{s}\,, \tag{3}$$

Where: $\boldsymbol{W}$ – matrix $\boldsymbol{m} \times \boldsymbol{n}$ and $\boldsymbol{m} > \boldsymbol{n}$, $s = [\boldsymbol{s_1}, \boldsymbol{s_2}, \ldots, \boldsymbol{s_m}]^{\boldsymbol{T}}$ –right side equation vector,

$\boldsymbol{f} = [fs_1, fs_2, \ldots, f_n]^T$ – solution.

One way to solve this problem is to find the vector $\boldsymbol{f}$ * with the lowest possible Euclidean norm, which for the given matrix $\boldsymbol{W}$ and vector $\boldsymbol{s}$ minimizes the Euclidean norm of the residual vector r = $\boldsymbol{s}$-**Wf**.

$$\|\boldsymbol{r}\|_2 = min\|\boldsymbol{s} - \boldsymbol{Wf}\|_2, \quad \|\boldsymbol{f}^*\|_2 = min\|\boldsymbol{f}\|_2, \tag{4}$$

This is the linear least squares. When defining an analytic form of a solution of a linear least squares task and examining its properties, we use the theorem on the distribution of any rectangular matrix to the product of orthogonal, diagonal and orthogonal matrix [8,10,11,12].

For any matrix $\boldsymbol{W} \in R_{m\times n}\ (m \geq n)$ there exist the orthogonal matrix $\boldsymbol{U} \in R_{m\times m}$, i $\boldsymbol{V} \in R_{n\times n}$ such that:

$$\boldsymbol{W} = \boldsymbol{UDV}^T\,, \tag{5}$$

$$\boldsymbol{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & d_{k-1} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & d_k & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \ddots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \vdots & d_n \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 \end{bmatrix} \in R_{mxn}\,, \tag{6}$$

Where: $d_1 \geq d_2 \geq \cdots \geq d_{k-1} \geq d_k > d_{k+1} = d_{k+2} = \cdots d_n = 0$, and k is a pseudo-array of matrix $\boldsymbol{W}$.

The values $\boldsymbol{d_i}$ are called special values of the $\boldsymbol{W}$ matrix, and the distribution (5) by the special values SVD. Knowing the distribution (5), we can easily determine all the solutions. Assume that the matrix $\boldsymbol{W}$has a defined distribution with respect to singular values where $\boldsymbol{Dd}$ is a diagonal matrix containing singular values. The difference in relation to matrix $\boldsymbol{D}$ is that there are zero rows and is a square matrix of $\boldsymbol{n\times n}$:

$$\boldsymbol{W} = \boldsymbol{U}\begin{bmatrix}\boldsymbol{Dd} \\ \boldsymbol{0}\end{bmatrix}\boldsymbol{V}^T. \tag{7}$$

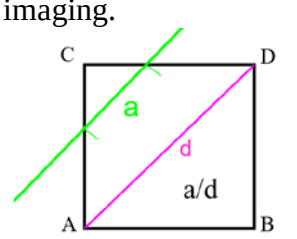
$$\boldsymbol{f}^* = \boldsymbol{W}^+\boldsymbol{s} = \boldsymbol{VD}^+\boldsymbol{g}, \tag{8}$$

Where the matrix $\boldsymbol{W}^+$ is called a pseudo-inverse matrix (or called inverse matrix in the sense of Moore-Penrose)

$$D^{+} = diag\left(\frac{1}{d_1}, \ldots, \frac{1}{d_k}, 0, \ldots, 0\right) \in R_{n \times m}, \tag{9}$$

4. Simulation and 2D reconstructions

The most common modification of the ART method is that in the appropriate places we calculate the ratio of the length of the segments: the radius inside the pixel and the diagonal length of the pixel. This method is costly because you need to examine the intersection points with each of the four sides of the pixel. A much more computationally computable method is when a pixel is represented as a circle inscribed in a square pixel (figure 3). The conversion of a square pixel into a circular pixel allows us to speed up an algorithm (over 12 times) in the case of 2D imaging.



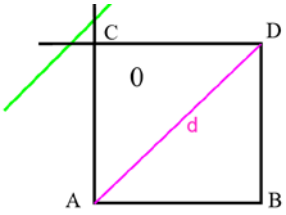


Fig.2. The principle of determining the elements of the coefficients matrix W.

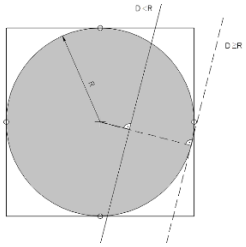


Fig.3. Modified wheel-shaped pixel.

Obtaining accurate tomographic images of the studied phenomena requires the use of a suitable mathematical methods. Proper selection of the reconstruction algorithm has a significant effect on the minimization of computational errors, which together with measurement errors determine the accuracy of the results obtained. We have to make a compromise choice of the solution that will provide the least possible residuum. Results showing the effect of noisy signal measurement are shown in Figure 4 and Figure 5. The noise level is relatively small and is 1%.

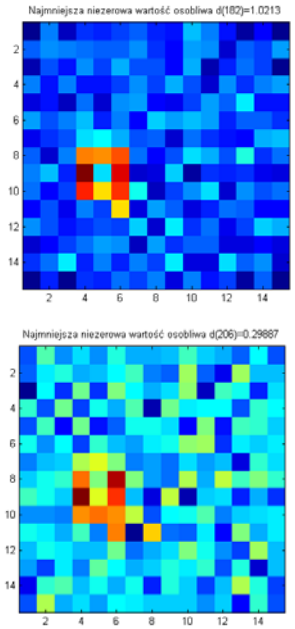


Fig.4. Results for 1% (noise level) distribution of material factor density - image of interior area for different number of peculiar values taken for a test solution.

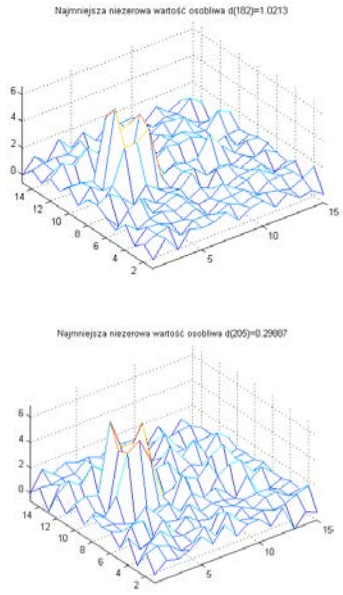


Fig.5. Results for 1% (noise level) distribution of material factor density - image of interior area for different number of peculiar values taken for a test solution..

In tomographic techniques, spatial resolution is always the most important issue, especially in bioengineering. This means, among other things, the ability to detect the smallest details. Let's examine the influence of the size of an object on its ability to identify. Let us consider the possibility of imaging a much smaller object whose side is 9.375% of the length of the whole object. For this purpose, it is necessary to increase discretization (reduce the size of pixels to maintain the overlap of the resulting system of equations. The results are shown in Figure 6.

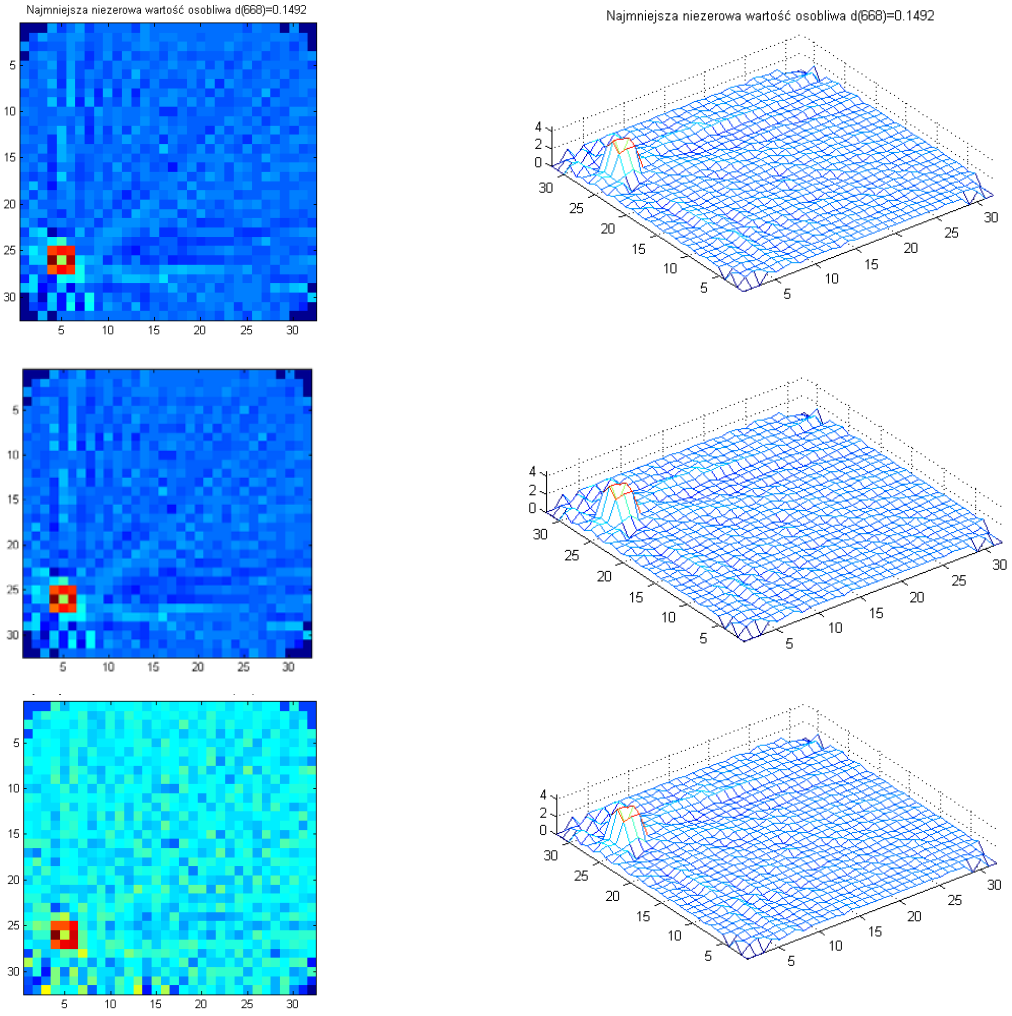


Fig.6. Images obtained for the next three trial solutions and the corresponding spatial distribution of the material factor density.

5. Conclusion

The primary advantage of tomography (both ultrasound and radio) is non-invasive measurement in the test environment, without any change in physic-chemical parameters that could interfere with the measurement results. Measurement of parameters such as signal transduction time, attenuation coefficient and its derivative after frequency allow (after appropriate reconfiguration transformations) the imaging of the internal structure of the investigated medium. Differences in local values of specific acoustic parameters (sound tomography) or signal strength (radio tomography) are the basis of this imaging. The image obtained by appropriate reconstructive methods shows that the distribution of local values of selected parameters obtained from the measurement of data by scanning technique from as many possible directions. This technique allows us to obtain quantitative images of an internal

structure in which the numerical values of each pixel describe the physical properties of the examined objects. The use of ART techniques has several advantages to other computational techniques. ART techniques also allow us the reconstruction of more contrasting tomographic images, which is particularly important in situations of observation of objects or in the course of rapid changes in their parameters. They also allow reconstructions of images for relatively small number of projections, and also allow reconstruction of projections with incomplete data.

References

1. Banasiak R, Wajman R, Sankowski D, Soleimani M., *Three-Dimensional Nonlinear Inversion of Electrical Capacitance Tomography Data Using a Complete Sensor Model*. Progress In Electromagnetics Research (PIER) vol. 100, pp. 219-234, 2010.
2. Cierniak R., Tomografia komputerowa, Budowa urządzeń CT, Algorytmy rekonstrukcyjne, Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2005.
3. Das Y., Boerner W.M.: *On Radar Target Shape Estimation Using Algorithms for Reconstruction from Projections*, „IEEE Transactions on Antennas and Propagation", AP–26(2), 1978.
4. Gudra T., Opieliński K.J.: *The multi–element probes for ultrasound transmission tomography*, „Journal de Physique" 4, 137, pp. 79–86, 2006.
5. Gudra T., Opieliński K.: Investigation of biological structures and their phantoms using ultrasound transmission tomography, „Molecular and Quantum Acoustics", 22, pp. 85–94, 2001.
6. Filipowicz S.F., Rymarczyk T., *The Shape Reconstruction of Unknown Objects for Inverse Problems*, Przeglad Elektrotechniczny, vol. 88, Issue 3A, pp. 55-57, 2012.
7. Kaczmarz S.: *Angenäherte Auflösung von Systemen Linearer Gleichungen*, „Bull. Acad. Polon. Sci. Lett. A", vol. 6–8A, pp. 355–357, 1937.
8. Kak A.C., Slaney M.: *Principles of Computerized Tomographic Imaging*, IEEE Press, New York, 1999.
9. Kryszyn J., Smolik W., Radzik B., Olszewski T., Szabatin R., *Switchless charge-discharge circuit for electrical capacitance tomography*, Measurement Science and Technology, vol. 25, Issue 11, pp. 1-9, 2014.
10. Lawson Ch.L., Hanson R.J.: *Solving Least Squares Problems*, „Classics in Applied Mathematics" 15, SIAM, 1995.
11. Opieliński K.J., Zastosowanie transmisji fal ultradźwiękowych do charakteryzowania i obrazowania struktury ośrodków biologicznych, Oficyna Wydawnicza Politechniki Wrocławskiej, 2011.
12. Polakowski, K., Filipowicz, S.F., Sikora, J., Rymarczyk, *Tomography technology application*, Przeglad Elektrotechniczny, vol. 84, Issue 12, pp. 227-229, 2008.
13. Rymarczyk T., *New Methods to Determine Moisture Areas by Electrical Impedance Tomography*, International Journal of Applied Electromagnetics and Mechanics, vol. 37, Issue 1-2, pp.79–87, 2016.
14. Rymarczyk T., Adamkiewicz P., Duda K., Szumowski J., Sikora J., *New Electrical Tomographic Method to Determine Dampness in Historical Buildings*, Archives of Electrical Engineering, vol. 65, Issue 2, pp. 273-283, 2016.
15. Rymarczyk T., Tchórzewski P., Adamkiewicz P., Duda K., Szumowski J., Sikora J., *Practical Implementation of Electrical Tomography in a Distributed System to Examine the Condition of Objects*, IEEE Sensors Journal, vol. 17, Issue:24, 2017, pp. 8166-8186, DOI: 10.1109/JSEN.2017.2746748.
16. Tchorzewski P., Rymarczyk T., Sikora J., *Using Topological Algorithms to Solve Inverse Problem in Electrical Impedance Tomography*, International Interdisciplinary Phd Workshop 2016, pp. 46-50, 2016.
17. Smolik W., Kryszyn J., Radzik B., Stosio M., Wróblewski P., Wanta D., Dańko L., Olszewski T., Szabatin R., *Single-shot high-voltage circuit for electrical capacitance tomography*, Measurement Science and Technology, vol. 28, Issue 2, p. 025902, 2017.

Karol DUDA[1], Tomasz RYMARCZYK, [12], Piotr BOŻEK[1], Przemysław ADAMKIEWICZ[1]

Prototype of ERT Device for Reconstruction of Resistive Distribution at Big Areas

There was presented a new device based on electrical resistance tomography. Its aim was to verify the repeatability of test results by eliminating laboratory equipment, and to validate the use of simple and cheap electronics to the structure of the ERT. Electrical resistance tomography, which is based on measuring potential difference, can be used to calculate conductivity.

1. Introduction

Electrical tomography (ET) is known that the inverse problem is nonlinear and highly ill-posed [1-3,6,11-13]. Many different numerical techniques can be used to solve the problem [3-8]. The problem is the low level of measured values which should be measured quite accurately and in a very short time. Electrical resistance tomography (ERT) involves placing electrodes on the examined object. In this work, there was created a new device after achieving good research results of measurements of the ERT. Its aim was to verify the repeatability of test results by eliminating laboratory equipment, and to validate the use of simple and cheap electronics to the structure of the ERT. The measurement in electrical resistance tomography involves placing electrodes on the examined object. This solution allows to obtain a conductivity distribution within the test object, and this consequently allows to find a distinction between materials of different conductivity [8-10].

2. Measurement model

ERT is a geophysical technique in which DC electrical current is injected into the ground between one pair of electrodes and the voltage is measured between another pair [1]. Electrical signals are then transmitted through selected electrode locations while electrical potential measurements are recorded at numerous other locations. This process is repeated systematically for many different source-receiver combinations and the resulting data-set enables the reconstruction of a cross-section through the survey area [10]. The cross-sectional image, or tomographic slice, depicts a spatial distribution of electrical resistivity, which is closely related to the internal structure of the object. Contrasts in the electrical properties of different geological materials enable earth scientists to set non-invasively map structures in the subsurface.

Table 1. Laboratory vs. real objects

Principal differences	
Point electrodes	Multisensor electrodes
Volatage to 5 [Vpp]	Voltage to 250 [Vpp]
Maximum current 10 [mA]	Maximum current 3 [A]
Measuring wire to 2 [m]	Measuring wire to 20 [m]

[1)] Research and Development Center, Netrix S.A., Związkowa 26, 20-148 Lublin, POLAND

[2)] University of Economics and Innovation in Lublin, Projektowa 4, 20-209 Lublin, POLAND

Corrosivity category: very small	Corrosivity category: very large

Table 1 presents differences between the laboratory model and real object in geophysical environment. Figure 1 shows the geophysical measurement model and Figure 2 presents model of ERT measurements.

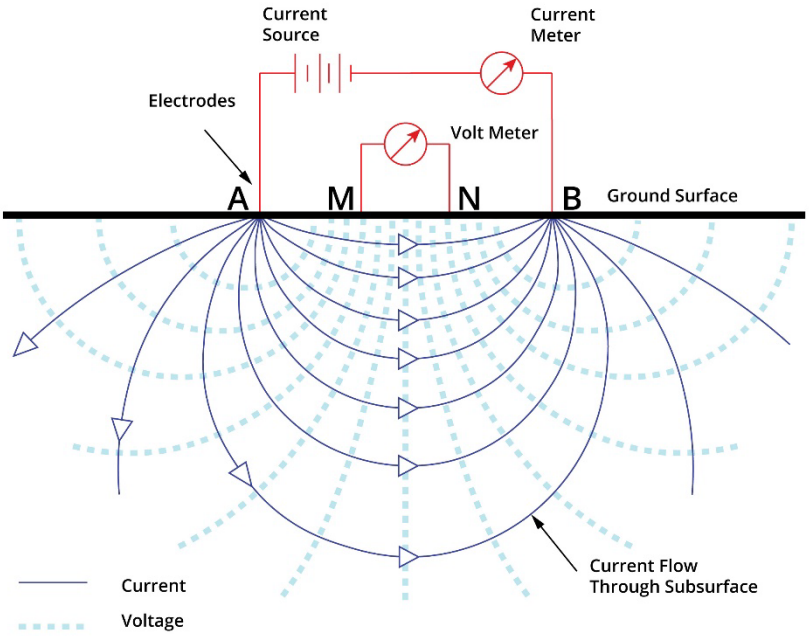


Fig. 1. Geophysical measurement model.

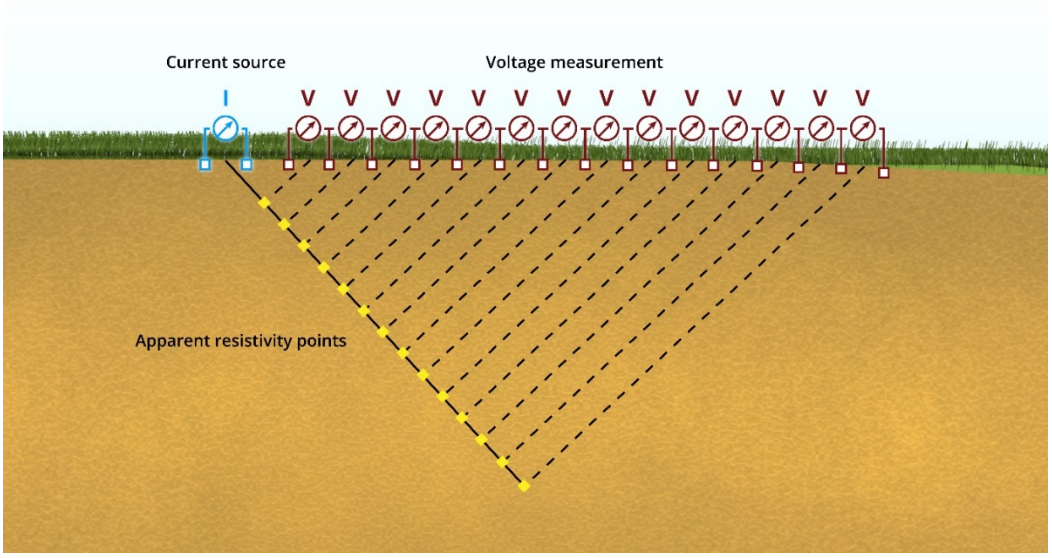


Fig. 2. ERT measurements in the geophysical environment.

3. ERT device

The ERT solution should make ERT measurements for reconstruction of resistive distribution at big areas (tens of meters) with of conducting media. The ERT equipment consists of some separated modules: current generator, measure block, multiplexer and controller. These modules should adapt to high voltages (about 250 V) and high currents

(about 3 A). There should be possibility to connect more than one multiplexer to this system. The designed modules should be supervised and served by a single controller. A schematic diagram of the system is presented in Figure 3.

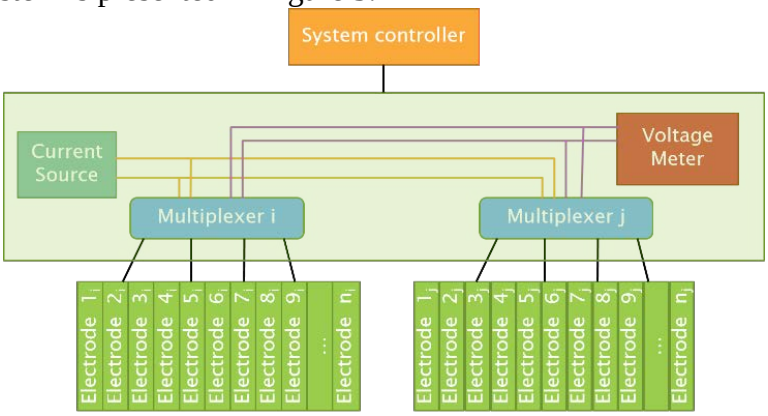


Fig. 3. The ERT measurement system schematic diagram.

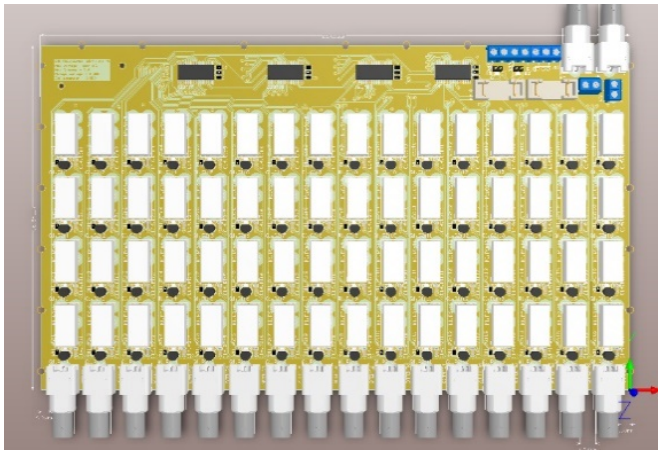


Fig. 4. The model of ERT multiplexer with mounted relays and BNC connectors.

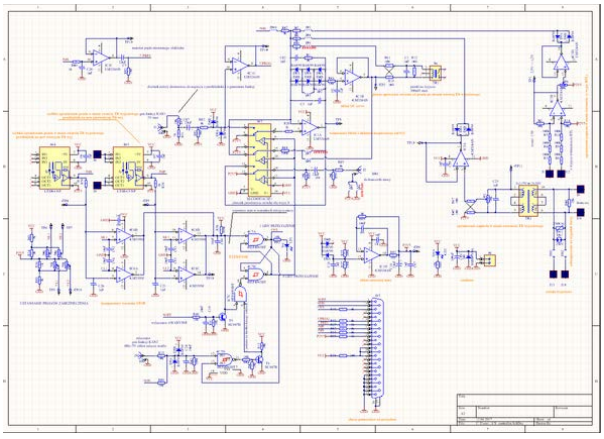


Fig. 5. Diagram of the amplifier circuit with protection.

The way of switching electrodes is the same as electrical impedance tomogrpahy device – multiplexer 4 to 16. To switching so high voltages and currents it is necessary to use electro-magnetic or solid-state relays. Because of device parameters and price electro-magnetic relays were chosen. To make multiplexer control easier, there are I2C bus expanders used. Thanks to this, it needs to use only two I2C wire – in place of eighteen. It makes possibility of expanding number of electrodes by connecting a few the same multiplexers.

The multiplexer line design has proper width of nets to deal with high currents, and clearances that deal with high voltages to avoid interferences or current arc. Electrodes will be connected to multiplexer by BCN connectors and coaxial cables. A PCB Layout model is presented in Fig. 4. Figure 5 shows the diagram of the amplifier circuit with protection. Measure block will be measure drop voltages between each electrode that multiplexer sets.

To making measures, 18-bit successive approximation ADC – AD7989 is used. A range of measured voltage this converter is ±5 V, and it is compatible with SPI interface. To improve the conditions of the input signal, before ADC two instrumental amplifiers AD623 are included and it is possible to set three gains: 2, 10, 20 V/V on them.

To match potentials between measuring signal and instrumental amplifier there is a toroidal transformer installed. It matches potentials and compensate signals to lower voltage. Changing ranges measured voltages are worked out by switching pairs of voltage divider resistors by relay. In order to generate a sufficiently large inducement signal, a direct digital synthesis (DDS) generator is constructed. DDS allows to generate a signal of any waveform and a wide frequency range. The generator will be digitally controlled. The current source will be based on a transformer and power transistors.

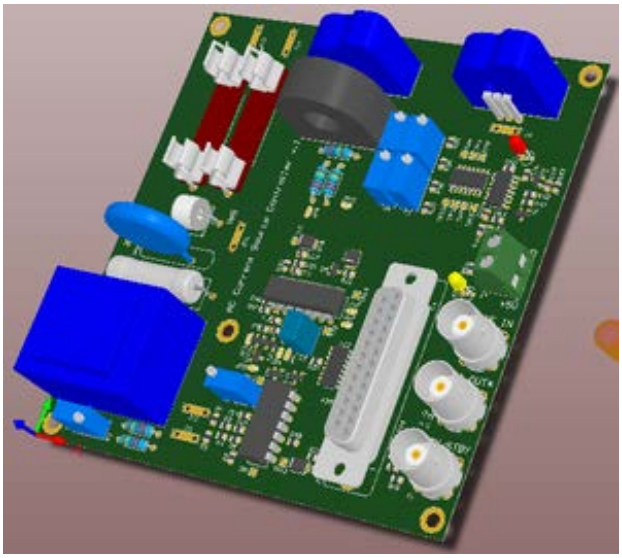


Fig. 6. PCB model of the circuit controlled by current source voltage.

Fig. 7. Multiplexer, VCCS and power amp in 2U.

The purpose of the circuit in Figure 6 is to compensate for an error on the output of the entire device, and to respond quickly to shorts of outputs, output openings, too high output currents, disconnection of the function generator by cutting off the current output. The output of the error amplifier circuit is fed to a power amplifier that amplifies the generated signal. The power amplifier output is connected to a transformer to reduce the output voltage and raise the

amplitude of the current. The output of the current generator can have parameters up to 230V and 3A at frequencies from 1 to 1kHz. All components, excluding the power amplifier, are housed in a 2U enclosure. The signal power amplifier as a finished device is also located in the 2U housing (Figure 7).

The purpose of the measuring module is to measure the potential difference on the measuring electrodes. Due to the varied parameters of the soil, which can vary according to atmospheric conditions, the range of measured voltages is from 1mV to 250V. The first step of the measuring module is a transformer that insulates galvanic signals from the entire circuit, lowers the voltage, and locates the signal with respect to the ground of the meter circuit. The next step is the pair of resistors that are switched by the relays. The purpose of this stage is to change the range of measured voltages. Behind the pair of resistors are operational amplifiers in the voltage repeater circuits, they serve to reduce the increase of the input impedance to the measurement path. The measuring path consists of two parallel measuring amplifiers. This solution eliminates the convergence of the measured signal, as well as the conversion of a simple to differential signal, which is necessary for the rapid measurement of many modern analogue-to-digital converters. That is precisely the Analog Devices AD7989, which was used in this module. This transmitter has an 18-bit resolution capability and can measure at a rate of 100k samples per second. The output of this transmitter is compatible with the SPI protocol. Measurement and range selection is monitored by the AtXmega processor, which can be sent to the monitoring unit for measurements or via USB (Figure 8).

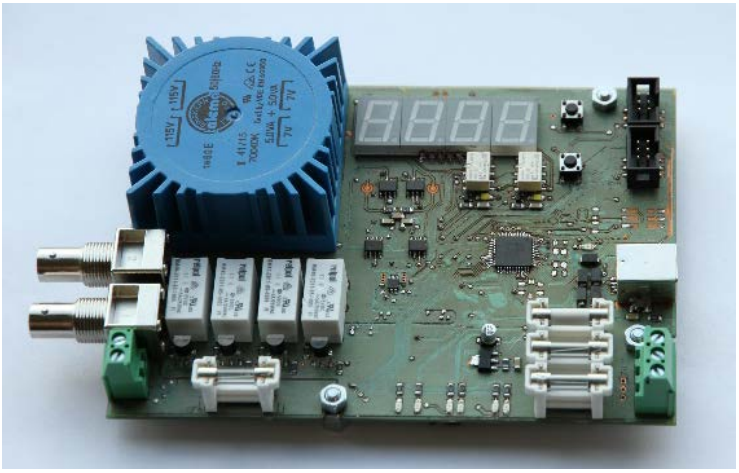


Fig. 8. PCB of voltage drop meter with a range of 1mV-250V (RMS).

Figure 9 shows the ERT field test bench with laboratory meters.

Fig. 9. The measuring ERT field of laboratory meters.

5. Conclusion

This article presents the design of the ERT device to measure in the real environment. The presented solution was based on electrical resistance tomography. The basic information about the built model system is given. There was described the concept of the measuring equipment for data acquisition. Electrical resistance tomography, which is based on measuring potential difference, can be used to calculate conductivity. Its aim was to verify the repeatability of test results by eliminating laboratory equipment, and to validate the use of simple and cheap electronics to the structure of the measurement system.

References

1. Jones G., Sentenac P., Zielinski M., *Desiccation cracking detection using 2-D and 3-D Electrical Resistivity Tomography: Validation on a flood embankment*, Journal of Applied Geophysics, vol. 106, July, 2014.
2. Duda K., Adamkiewicz P., Rymarczyk T., *Nondestructive Method to Examine Brick Wall Dampness*, International Interdisciplinary Phd Workshop 2016, pp. 68-71, 2016.
3. Filipowicz S.F., Rymarczyk T., *Measurement Methods and Image Reconstruction in Electrical Impedance Tomography*, Przeglad Elektrotechniczny, vol. 88, Issue 6, pp.247-250. 2012.
4. Gola A., Świć A., *Computer-Aided Machine Tool Selection for Focused Flexibility Manufacturing Systems Using Economical Criteria*, Actual Problems of Economics, vol. 124, Issue 10, pp. 383-389, 2011.
5. Kłosowski G, Kozłowski E, Gola A., *Integer linear programming in optimization of waste after cutting in the furniture manufacturing*. Advances in Intelligent Systems and Computing 2018; vol. 637, pp. 260-270, 2018.
6. Kryszyn J., Smolik W., *Toolbox for 3D Modelling and Image Reconstruction in Electrical Capacitance Tomography*, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska - Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, ISSN: 2083-0157, vol. 1, pp. 137-145, 2017.
7. Polakowski, K., Filipowicz, S.F., Sikora, J., Rymarczyk, T., *Quality of imaging in multipath tomography*, Przeglad Elektrotechniczny, vol. 85, Issue 12, pp. 134-136, 2009.
8. Rymarczyk T., *Characterization of the shape of unknown objects by inverse numerical methods*, Przeglad Elektrotechniczny, vol. 88, Issue 7B, pp. 138-140, 2012.
9. Rymarczyk T., Adamkiewicz P., Rymarczyk P. et al., *Monitoring Flood Embankment by Electrical Impedance Tomography*, International Interdisciplinary Phd Workshop 2016, pp. 63-67, 2016.
10. Rymarczyk T., *Using electrical impedance tomography to monitoring flood banks*, International Journal of Applied Electromagnetics and Mechanics, vol. 45, pp. 489–494, 2014.
11. Smolik W., *Forward Problem Solver for Image Reconstruction by Nonlinear Optimization in Electrical Capacitance Tomography*, Flow Measurement and Instrumentation, vol. 21, pp. 70–77, 2010.
12. Soleimani M., Mitchell CN, Banasiak R., Wajman R., Adler A., *Four-dimensional electrical capacitance tomography imaging using experimental data*, Progress In Electromagnetics Research, vol. 90, pp. 171-186, 2009.
13. Wajman, R., Banasiak, R., *Tunnel-based method of sensitivity matrix calculation for 3D-ECT imaging*. Sensor Review, vol. 34, pp. 273–283, 2014. doi:10.1108/SR-06-2013-692

Tomasz RYMARCZYK, [12]

Mathematical Analysis of Topological, Shape and Material Derivatives for Electrical Tomography

In this paper, there was presented mathematical analysis of topological, shape and material derivatives in imaging techniques. The proposed solution uses the level set method for topology optimization. The level set function is based on shape and topology optimization approach with piecewise constant conductivities. In this article, there was solved the inverse problem in electrical impedance tomography.

1. Topological methods

Level set methods have been applied very successfully in many areas of the scientific modelling, for example in propagating fronts and interfaces [9,11-13]. Therefore, they are used to study shape optimization problems [4,6,10]. Instead of using the physically driven velocity, the level set method typically moves the surfaces by the gradient flow of an energy functional. These approaches based on shape sensitivity include the boundary design of elastic. There are two features that make these methods suitable for the topology optimization. The structure is represented by an implicit function such that its zero-level set defines the boundary of the object. This function is often discretized on a regular grid that conveniently coincides with the finite element mesh used for structural analysis. Typical problem in the electrical tomography requires the identification of the unknown internal area from near-boundary measurements of the electrical potential. It is assumed that the value of the conductivity is known in sub-regions whose boundaries are unknown. Level set method is the numerical technique which can follow the evolution of interfaces (Fig. 1). These interfaces can develop sharp corners, break apart and merge together.
The motion is seen as the convection of values (levels) from the function ϕ with the velocity field $\vec{v}$. Such process is described by the Hamilton-Jacobi equation:

$$\frac{\partial \phi}{\partial t} + \vec{v} \cdot \nabla \phi = 0. \tag{1}$$

Reinitialization is necessary when flat or steep regions complicate the determination of the zero contour. This process is described by following partial differential equation:

$$\frac{\partial \phi}{\partial t} + S(\phi)(|\nabla \phi| - 1) = 0, \tag{2}$$

where: $S(\phi) = \phi \,/\sqrt{\phi^2 + \varepsilon^2}$, $|\varepsilon| \ll 1$. Equation (2) is solved until a steady state is achieved.

The topological method is based on the differentiability of solutions to variational inequalities with respect to the coefficients of the governing differential operator. It is required that the metric projection in the energy space. Such property is sufficient to obtain the directional differentiability of solutions to the variational inequality with respect to the boundary variations with respect to the changes in the topology. A useful concept for calculating derivatives for cost functional is the so-called material and shape derivative of states u. In the application of inverse problems, these states typically are the solutions of

[1)] Research and Development Center, Netrix S.A., Związkowa 26, 20-148 Lublin, POLAND

[42)] University of Economics and Innovation in Lublin, Projektowa 4, 20-209 Lublin, POLAND

partial differential equations which model the probing fields and which depend one way or another on the shape.

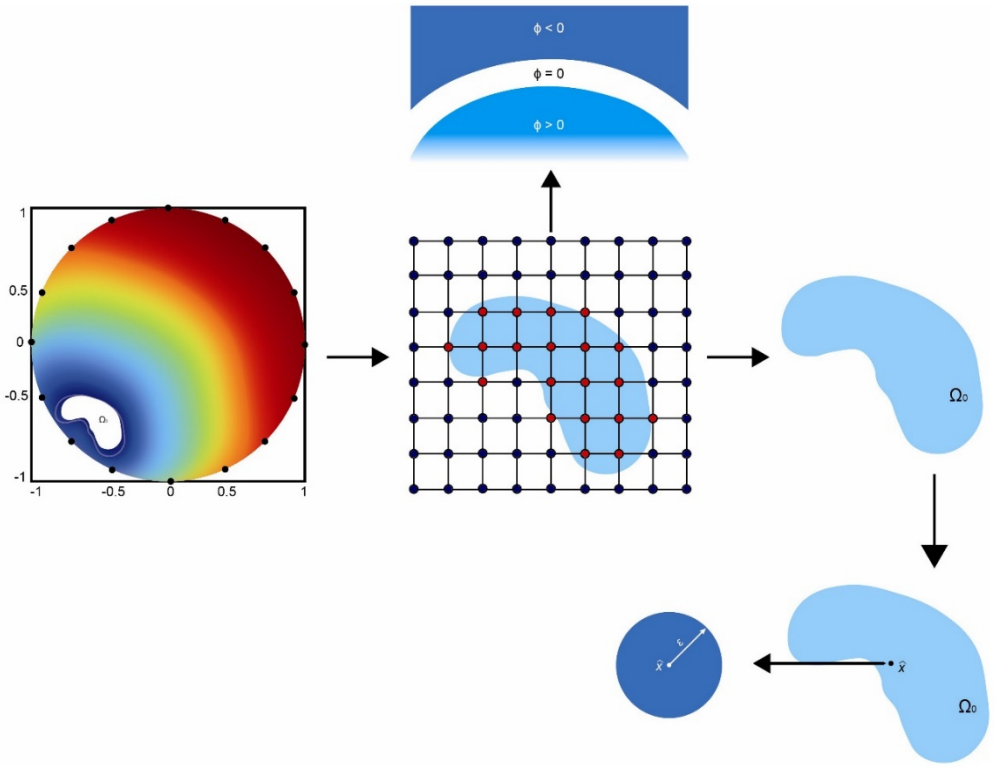


Fig.1. Idea of topological methods in the process optimization.

The material derivative concept is applied to the formulation of an inverse obstacle problem. This is the derivative with respect to the geometry for a moving interface. The material derivative of the primary variable is performed. It consists of the material derivative of the primary variable, the sensitivity with respect to the conductivity, and the sensitivity with respect to the deformation velocity. The boundary conditions of the adjoint problem are defined such that the desired material derivative of the primary variable is obtained.
The material derivative $\dot{u}(x)$ is given by:

$$\dot{u}(\vec{r}) \equiv \lim_{t\to 0} \frac{u_t(\vec{r} + t\vec{v}(\vec{r})) - u(\vec{r})}{t} \tag{3}$$

where $(x, y)\epsilon\Omega_t$.

This method is based on the fact that the definition of shape derivatives is presented to differentiate boundary and domain integrals with respect to a deformation of the contour. The shape derivative is introduced and the results for two simple shape functions (boundary and domain integrals) are given. In the following section, the shape derivative concept is extended by material derivatives. This is necessary for more complex shape functions where a spatial distributed variable depends on the geometry itself. At last, results of the shape derivative concept applied to the level set function are presented. Minimization using shape derivatives have been used for finding curves and surfaces. Shape and material derivatives are calculated easily by the adjoint variable method with few computational costs.
The shape derivative is following [13]:

$$u'(\vec{r}) \equiv \lim_{t\to 0} \frac{u_t(\vec{r}) - u(\vec{r})}{t} = \dot{u}(\vec{r}) - \vec{v}(\vec{r}) \cdot \nabla u(\vec{r}) \tag{4}$$

The topological derivative is defined as the first term of the asymptotic expansion of a given shape functional with respect to a small parameter that measures the size of singular domain perturbation. It represents the variation of the shape functional when the domain is perturbed by holes, inclusions, defects or cracks. Shape derivatives, material derivatives and topological derivatives have been incorporated into level set methods to investigate shape optimization problems [5,9,15,16]. The topological derivative evaluated for a given shape functional defined in a geometrical domain and dependent on a classical solution to elliptic boundary value problem is a principal tool in shape optimization. The shape optimization problem is usually defined as a minimization of a given shape functional. The shape functional can be written as an integral over the domain Ω or on the boundary $\partial\Omega$ of a function which depends on a solution $u(\Omega;x)$ of some boundary value problem.

2. Electrical tomography

There is well known that the inverse problem is nonlinear and highly ill-posed. The forward problem solution in EIT consists in determining potential distribution inside the region Ω under given boundary conditions and full information about region under consideration. For expediency, electrical impedance tomography (EIT) has historically been further subdivided into capacitance (ECT), for dielectric dominated systems, and resistance (ERT), for conducting processes [1-3,7,8,14,17,18]. This has arisen from an alternative approach to the hardware design for each sub-modality, however, the fundamental theory for both can be derived from Maxwell's equations. If a complex 'admittivity' term (γ) (also known as: complex conductivity or logically admittivity) is defined as:

$$\gamma = \sigma + i\omega\varepsilon \tag{5}$$

where σ is the electrical conductivity, ω is the angular frequency and ε is the permittivity. For an electric field strength (E) the free current density (J) in the area under investigation will be related by Ohm's law as follows:

$$\boldsymbol{J} = \gamma \boldsymbol{E} \tag{6}$$

It can also be shown that the gradient of the potential distribution (u) is related to electric field strength by:

$$\boldsymbol{E} = -\nabla \boldsymbol{u} \tag{7}$$

Assuming that there are no current sources or sinks within the region of interest then from Ampère's law:

$$\nabla \cdot \boldsymbol{J} = 0 \tag{8}$$

The potential distribution in the inhomogeneous, isotropic region:

$$\nabla \cdot (\gamma \nabla u) = 0, \tag{9}$$

where u – potential, or

$$\nabla \cdot (((\sigma + i\omega\varepsilon)\nabla u)) = 0, \tag{10}$$

According to the ratio of $\omega\varepsilon/\sigma$, when the resistance or the capacitance term is dominant, the governing equation is simplified further:

$$\nabla \cdot (\sigma \nabla u) = 0 \quad for \quad \frac{\omega\varepsilon}{\sigma} \ll 1 \quad \text{(ERT)} \tag{11}$$

$$\nabla \cdot (\varepsilon \nabla u) = 0 \quad for \quad \frac{\omega \varepsilon}{\sigma} \gg 1 \quad \text{(ECT)} \tag{12}$$

where electrical resistance tomography (ERT) means the resistance component is used for the tomography, while electrical capacitance tomography (ECT) used the capacitance component. Electrical tomography combined ERT and ECT is called electrical impedance tomography (EIT). It should be noted that frequently ERT is known as EIT. Measurements in Electrical Impedance Tomography performed by forcing alternating current and measure the voltage drops on the electrodes. As a result of the inverse problem solution is obtained distribution of conductivity in the study area.

a) b)

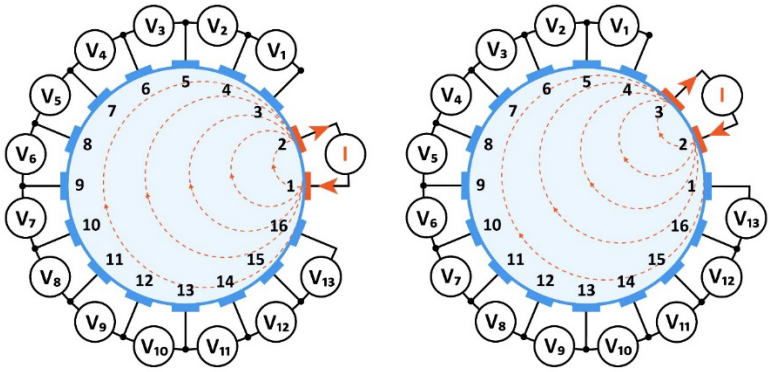


Fig. 2. Neighbouring measurement method in electrical impedance tomography.

Figure 2 presents the neighbouring (adjacent current method) of boundary potential data collection illustrated for a cylindrical volume conductor and 16 equally spaced electrodes: (a) projection-1, (b) projection-2.

Let p be the adjoint function satisfying:

$$\nabla \cdot (\gamma \nabla \mathrm{p}) = -(u - u_m), \tag{13}$$

where $u - u_m$ is difference between calculated and measurement voltages. The steepest descent direction $\vec{v}$ is given by [13]:

$$\vec{v} = -(\nabla \mathrm{u} \cdot \nabla \mathrm{p})\, \vec{n}, \tag{14}$$

where $\vec{n}$ is a normal vector. The normal velocity is evaluated by using weighted least squares interpolation to get:

$$\mathrm{v}_n^k = \nabla \mathrm{u}^k \cdot \nabla \mathrm{p}^k. \tag{15}$$

In next step the level set function is updated:

$$\phi^{k+1} = \phi^k - (\nabla \mathrm{u}^k \cdot \nabla \mathrm{p}^k)\, |\nabla \phi^k|\, \Delta t. \tag{16}$$

The level set method can be used to solve the minimization problem. Numerical methods of the shape and the topology optimization were based on the level set representation and the shape differentiation and were made possible topology changes during the optimization process.

Topological derivative using the level set function can solve the inverse problem in electrical impedance tomography. Velocity in Hamilton-Jacobi equation is defined following:

$$v_n = -[(\gamma - \gamma_i)(\nabla p_i \cdot \nabla u) + \beta |\gamma - \gamma_i| \mathcal{H}] \; on \; \Gamma_i \tag{17}$$

Laplace's equation is described as:

$$-\Delta u = 0 \; in \; \Omega^*, \; q\partial_n u = f \; on \; \Sigma, \; q\partial_n u = q_i \partial_n u_i \; on \; \Gamma_i, \tag{18}$$

$$-\Delta u_i = 0 \; in \; \Omega_i, \; u_i = u \; on \; \Gamma_i \tag{19}$$

The adjoint equation is following [5]:

$$-\Delta p = 0 \; in \; \Omega^*, \; q\partial_n p = 2(u - u_m) \; on \; \Sigma, \tag{20}$$

$$q\partial_n p = q_i \partial_n p_i \; on \; \Gamma_i, \; -\Delta p_i = 0 \; in \; \Omega_i, \; p_i = p \; on \; \Gamma_i \tag{21}$$

3. Example

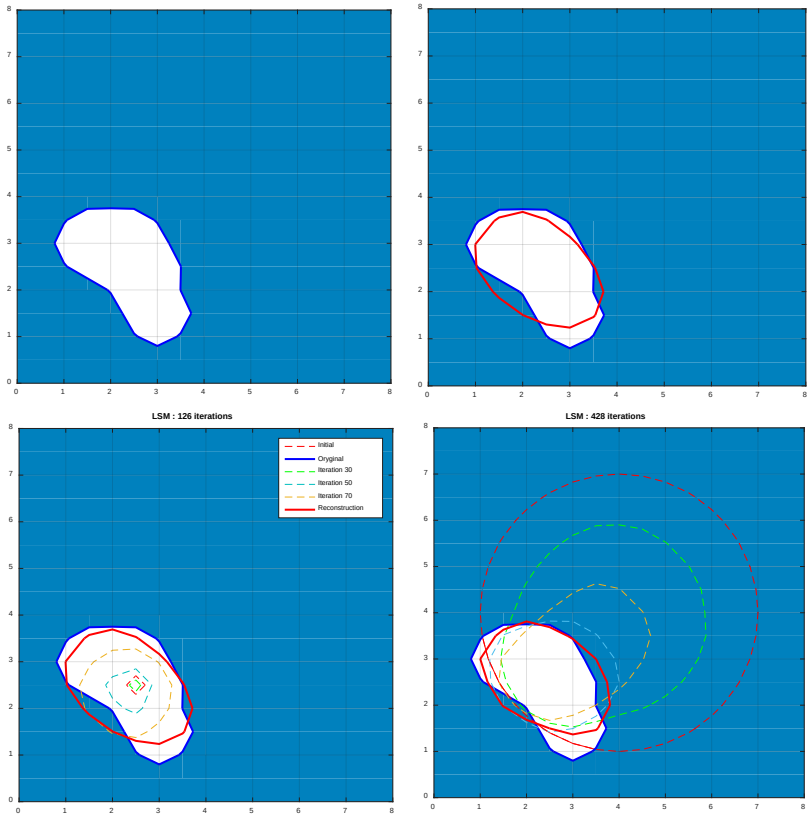


Fig. 3. Topological algorithm in process reconstruction in mesh 16x16: a) original object, b) final reconstruction, c) process reconstruction with initial small hole, d) process reconstruction in initial zero level set function.

The example of the image detection is shown in Fig. 3. Topological algorithm in process reconstruction was used in poor mesh 16x16. Individual images present: (a) original object, (b) final reconstruction, (c) process reconstruction with initial small hole, (d) process reconstruction in initial zero level set function.

4. Conclusion

Topological algorithms were implemented for obtaining the images in the process reconstruction by solving the inverse problem. There is shown that the results are dependent on initial values of design variables. Numerical simulations and measurement experiments show that the proposed algorithms give stability and higher accuracy of reconstruction results compared with deterministic methods. The results give a general overview that the proposed method can introduce an effective numerical solution for EIT. This work promises a new horizon to obtain better quality of the image reconstruction.

References

1. Duda K., Adamkiewicz P., Rymarczyk T., *Nondestructive Method to Examine Brick Wall Dampness*, International Interdisciplinary Phd Workshop 2016, pp. 68-71, 2016.
2. Fiderek P., Jaworski T., Banasiak R., Nowakowski J., Kucharski J., Wajman R., *Intelligent system for the two-phase flows diagnosis and control on the basis of raw 3D 3D ECT data*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 7, Issue 1, pp. 17-23, 2017; DOI: 10.5604/01.3001.0010.4576.
3. Filipowicz S.F., Rymarczyk T., *Measurement Methods and Image Reconstruction in Electrical Impedance Tomography*, Przeglad Elektrotechniczny, vol. 88, Issue 6, pp.247-250. 2012.
4. Gola A., Świć A., *Computer-Aided Machine Tool Selection for Focused Flexibility Manufacturing Systems Using Economical Criteria*, Actual Problems of Economics, vol. 124, Issue 10, pp. 383-389, 2011.
5. Hintermüller M., Laurain A., Novotny A., *Second-order topological expansion for electrical impedance tomography*, Advances in Computational Mathematics 36(2) 235–265, 2012.
6. Kłosowski G, Kozłowski E, Gola A., *Integer linear programming in optimization of waste after cutting in the furniture manufacturing*. Advances in Intelligent Systems and Computing 2018; vol. 637, pp. 260-270, 2018.
7. Kryszyn J., Smolik W., *Toolbox for 3D Modelling and Image Reconstruction in Electrical Capacitance Tomography*, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska - Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, ISSN: 2083-0157, vol. 1, pp. 137-145, 2017.
8. Kryszyn J, Wróblewski P, Stosio M, Wanta D, Olszewski T, Smolik W, *Architecture of EVT4 data acquisition system for electrical capacitance tomography*, Measurement: Journal of the International Measurement Confederation, vol. 101, pp. 28-39, 2017.
9. Osher S., Sethian J.A., *Fronts Propagating with Curvature Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations*, J. Comput. Phys. 79 12-49, 1988.
10. Polakowski, K., Filipowicz, S.F., Sikora, J., Rymarczyk, T., *Quality of imaging in multipath tomography*, Przeglad Elektrotechniczny, vol. 85, Issue 12, pp. 134-136, 2009.
11. Rymarczyk T., *Characterization of the shape of unknown objects by inverse numerical methods*, Przeglad Elektrotechniczny, vol. 88, Issue 7B, pp. 138-140, 2012.
12. Rymarczyk T., Adamkiewicz P., Rymarczyk P. et al., *Monitoring Flood Embankment by Electrical Impedance Tomography*, International Interdisciplinary Phd Workshop 2016, pp. 63-67, 2016.
13. Rymarczyk T., *Using electrical impedance tomography to monitoring flood banks*, International Journal of Applied Electromagnetics and Mechanics, vol. 45, pp. 489–494, 2014.
14. Smolik W., *Forward Problem Solver for Image Reconstruction by Nonlinear Optimization in Electrical Capacitance Tomography*, Flow Measurement and Instrumentation, vol. 21, pp. 70–77, 2010.
15. Sokołowski J. and Zolesio J.-P., *Introduction to shape optimization. Series in Computationnal Mathematics*, Springer Verlag 16, 1992.
16. Sokolowski J., Zochowski A., *On the topological derivative in shape optimization*, SIAM Journal on Control and Optimization 37 1251–1272, 1999.
17. Soleimani M., Mitchell CN, Banasiak R., Wajman R., Adler A., *Four-dimensional electrical capacitance tomography imaging using experimental data*, Progress In Electromagnetics Research, vol. 90, pp. 171-186, 2009.
18. Wajman, R., Banasiak, R., *Tunnel-based method of sensitivity matrix calculation for 3D-ECT imaging*. Sensor Review, vol. 34, pp. 273–283, 2014. doi:10.1108/SR-06-2013-692

Jakub SZUMOWSKI[1], Tomasz RYMARCZYK, [12], Paweł TCHÓRZEWSKI[1] Przemysław ADAMKIEWICZ[1], Jan SIKORA[3,4]

Prototype of Smart ECT Device for Data Acquisition in Quality Control Systems

This paper provides the device and the image reconstruction using the electrical capacitance tomography. The most characteristic features of presented project are compact size of data acquisition module and its low energy demand. Complexity and cost of electrical capacitance tomography systems implementation into manufacturing processes as well as substantial size of equipment are the main reasons of their absence in the industry. There was proposed the prototype solution equally efficient as most of the currently existing ones.

1. Introduction

Electrical capacitance tomographs are devices capable of performing analysis of pipeline fragment or a vessel filled with examined medium [1,3,14-17]. Obtained cross-section's image reconstruction can be analysed further for an automated quality control system's autonomous and correct decision. However, existing industrial solutions are still of significant size and structurally complex [2,11-15]. This fact greatly limits the number of actual deployments. Decision makers are afraid of electrical capacitance tomography implementation due to high costs of possible equipment failures and because of limited ability to support multiple elements in vast area of production lines. In response to the market demand and maintaining trend of complete automatic control, herein described system's project was created. It features some advantages of classic approach and meets the standards of modern industrial equipment at the same time. In this paper hardware issues, sample measurements and the image reconstruction are described.

Knowledge of the exact form of the technological processes is often crucial for achieving anticipated production quality or its proper efficiency. Control equipment providing man an essential information is, due to its role, mostly quite expensive and tightly integrated into production system. In order to monitor multiple elements there is a need for implementation of row of measuring devices. Furthermore, some complex phenomena have to be supervised by man due to technical problems with measurement. Such cases include research of multiphase flows, granular materials transport or inspecting processes of mixing substances using gases. In food, chemical or renewable energy industries those issues are significant and they are subjects of technological evaluation [4-10].

Figure 1 presents the idea of ECT model sensors (measuring electrodes, steel grounding screen, screening electrode, pipe model of dielectric material). Complex system to the data acquisition, the image reconstruction with cloud computing model was shown in Fig. 2.

[1)] Research and Development Center, Netrix S.A., Związkowa 26, 20-148 Lublin, POLAND

[2)] University of Economics and Innovation in Lublin, Projektowa 4, 20-209 Lublin, POLAND

[3)] Lublin University of Technology, Institute of Electronics and Information Technology, Nadbystrzycka 38d, 20-618, Lublin, Poland

[4)] Electrotechnical Institute, ul. Pożaryskiego 28, 04-703 Warszawa

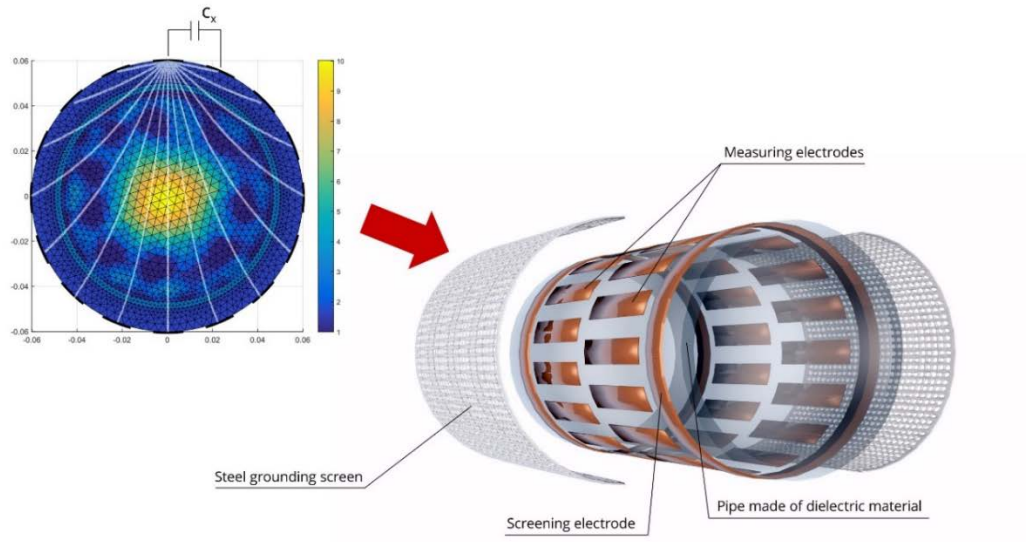


Fig. 1. Idea of ECT model sensors with the example of the image reconstruction (on the left side).

Fig. 2. Complex system to the data acquisition, the image reconstruction with cloud computing model.

2. Smart ECT

Electrical capacitance tomography scanners are devices capable of imaging interior of pipelines or vessels made of dielectric material. Obtained cross-section's reconstruction may be further analysed. However, existing solutions feature significant size and structural complexity. Project of herein presented low cost ECT system's prototype consists of 4 main elements: ECT sensor, data acquisition device, optional network infrastructure and a computer. Figure 3 shows the schematic model of Smart ECT measurement system. The data acquisition module is connected to ECT sensor and at the same time to a computer via network infrastructure. Direct low latency USB 2.0 connection is available as well. Computer processes measurements obtained from the device hence real-time image reconstruction may be achieved with proper software. Compact size of data acquisition module is a result of two significant factors. Firstly, the device's design consists of integrated circuits performing multiple functions simultaneously, e.g., Atmel ATmega32U4 merging classic microcontroller with USB controller or Texas Instruments CD4067BM integrated multiplexers instead of separate CMOS decoders and switches. Secondly, the number of electrodes has been limited to 16. These solutions greatly limit the number of tracks needed and thus overall PCB size. In electromagnetically stable environments optional Wi-Fi module with ESP-8266 chipset may be applied. Single PC workstation is able to receive and process transmissions from several

devices at the same time, provided they are all connected to the same network. The main part of the data acquisition module is Analog Devices AD7746 capacitance-to-digital converter. It is capable of measuring in range of ±4.096 pF with resolution up to 4 aF. Its high accuracy and precision is achieved at the expense of speed. Maximal output conversion rate is 90.1 Hz.

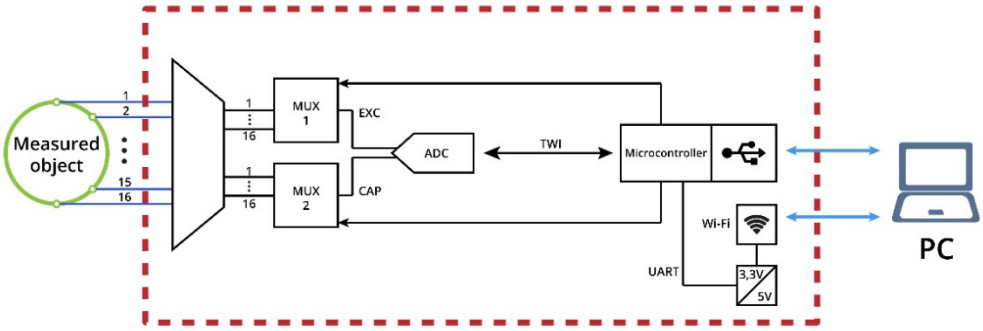


Fig. 3. Schematic model of Smart ECT measurement system.

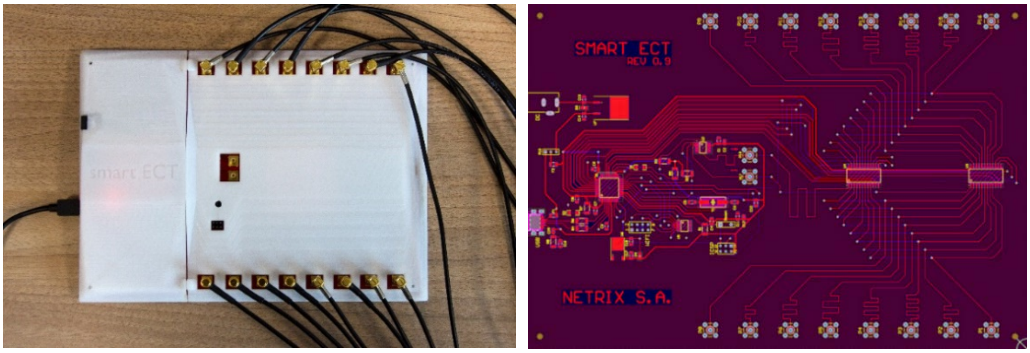


Fig. 4. Smart ECT (device and data acquisition module).

Smart ECT's efficiency has been tested with classic cylindrical ECT sensor built from a tube made of plexiglass with copper foil electrodes mounted externally. Internal diameter of the tube is 94 mm, while the external one is 100 mm. Data acquisition module was combined with ECT sensor via 16 coaxial cables RG174 type and 40 cm long. Due to its compact size data acquisition module may be mounted in close proximity to the ECT sensor. Figure 4 shows device's PCB mosaic and external overview respectively.

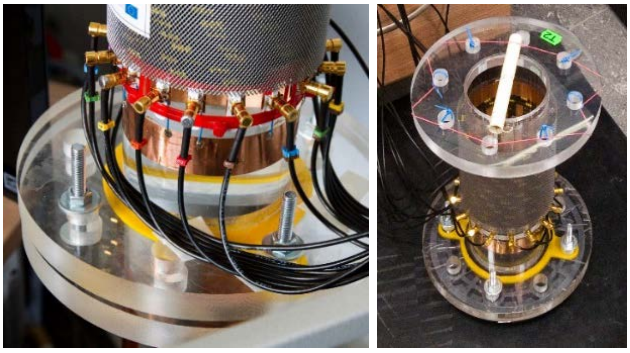


Fig. 5. The pipeline tomography model and the tube measurements.

The common approach to tomographic measurements data analysis is their processing in order to obtain a reconstructed image of the distribution of a required property in the examined cross section area. In the case of ECT, this property is the electric permittivity. The visualization of the inside of the investigated body is obtained on the basis of the recorded capacitance measurements. These permittivity distributions could be directly related to the material concentration present in the sensor (probe) space. In a typical measurement ECT

sensor use between 8 and 16 electrodes symmetrically spaced around cylindrical container (tube, pipe, etc.), as shown in the following Fig. 5. The tube measurements with rods are presented in Fig. 6.

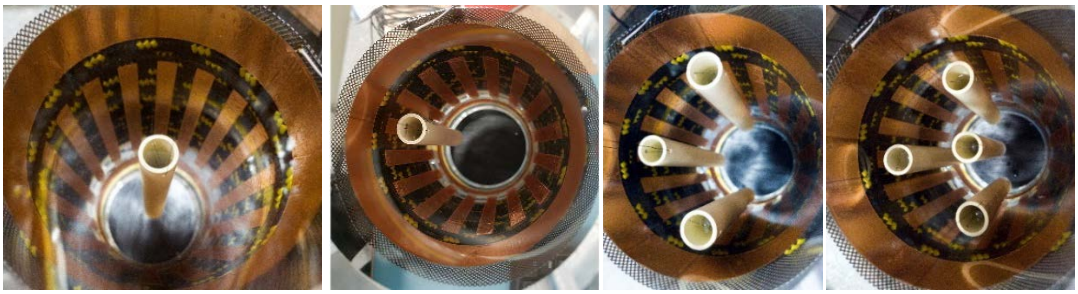


Fig. 6. Example of tube measurements with four different variants of rods.

3. Image Reconstruction

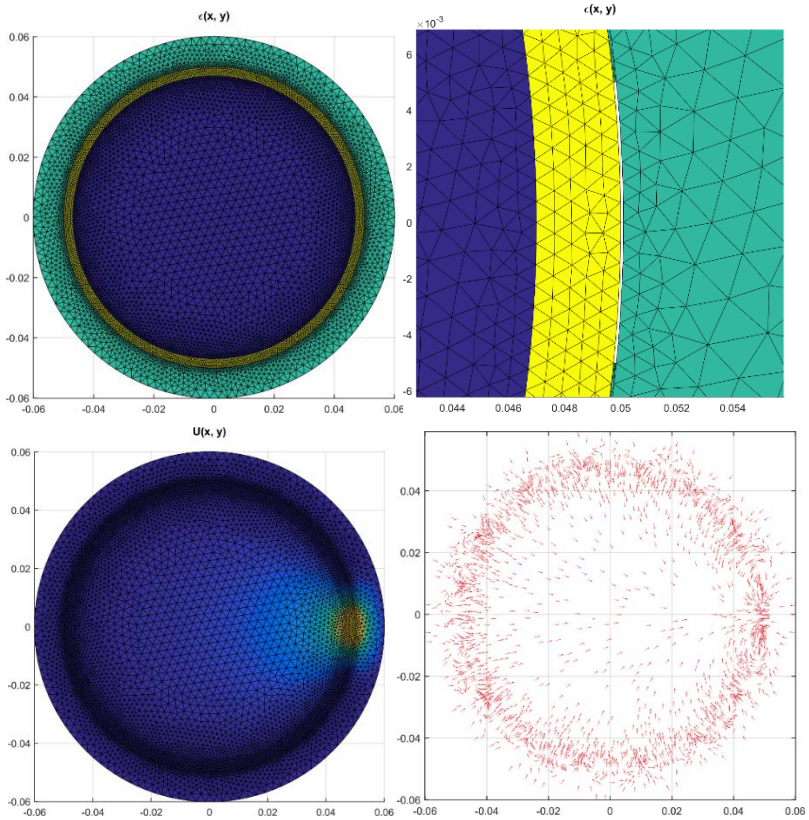


Fig. 7. Model of mesh and solution of the forward problem.

The purpose of the study was to obtain measurement results for various phantom configurations measured using alternative methods (chosen heuristically). The consequence of this action is both a practical verification of the quality of the individual solutions (also observing differences between them) and the same numerical package itself. The simple problem was solved using the finite element method. Grids used in calculations: rare 2 218 knots and 4 146 finite elements, dense 7 861 nodes and 15 146 finite elements (Fig.7). The examples of the image reconstruction were shown in Fig. 8 (simulation data) and in Fig. 9 (real measurement data). The Levenberg-Marquardt method was used to solve the inverse problem. The algorithm was developed by the authors, the experience of experimental confirmed its effectiveness better than other methods of gradient.

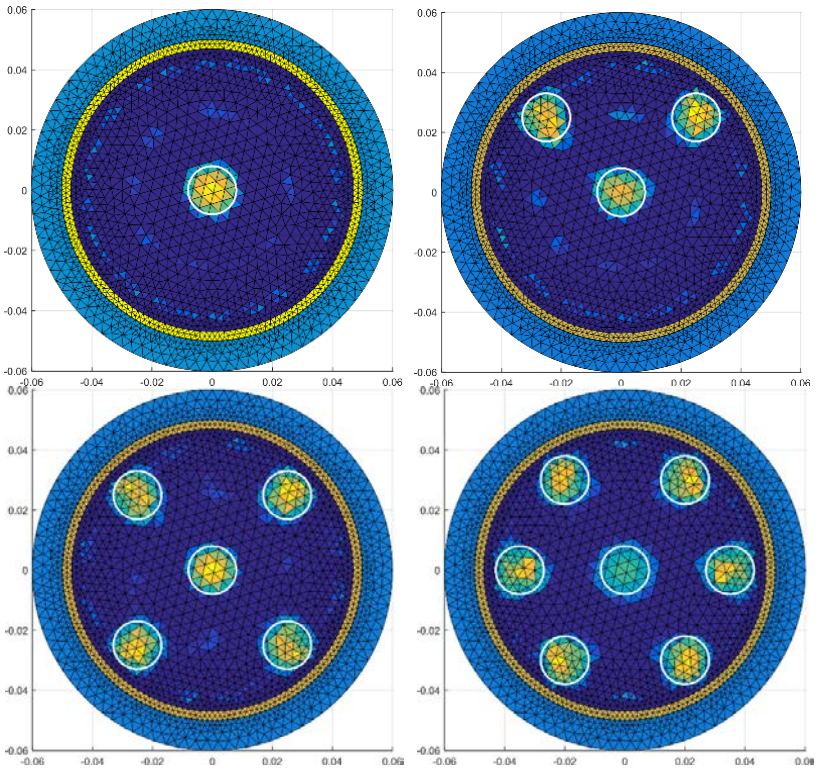


Fig. 8. The example of the image reconstruction.

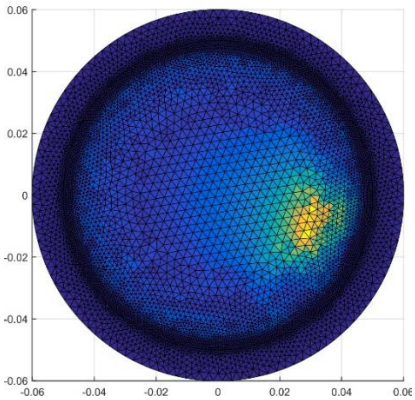


Fig. 9. The example of the image reconstruction by ECT - real data.

4. Conclusion

In this work, there was presented the solution using electrical capacitance tomography. There was designed Smart ECT device to image reconstruction. The development of new systems based on capacitive tomography technology can significantly improve the analysis of the properties of the tested object. The challenge is to create a single optimal and efficient measurement. It is also important to develop optimal algorithms for solving the inverse problem for accurate reconstruction. Using the algorithm prepared by the authors to solve the reverse problem using the Levenberg-Marquardt method proved to be effective. The algorithm was selected experimentally. The challenge is to create a single optimal and efficient measurement and low-cost system ensures easy application. It is also important to develop optimal algorithms for solving the inverse problem for accurate reproduction.

References

1. Beck M. S., Byars M., Dyakowski T., Waterfall R., He R., Wang S. J., Yang W. Q., *Principles and Industrial Applications of Electrical Capacitance Tomography*, Measurement and Control, September, Vol. 30, No. 7, 1997.
2. Duda K., Adamkiewicz P., Rymarczyk T., *Nondestructive Method to Examine Brick Wall Dampness*, International Interdisciplinary Phd Workshop 2016, pp. 68-71, 2016.
3. Fiderek P., Jaworski T., Banasiak R., Nowakowski J., Kucharski J., Wajman R., *Intelligent system for the two-phase flows diagnosis and control on the basis of raw 3D 3D ECT data*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 7, Issue 1, pp. 17-23, 2017; DOI: 10.5604/01.3001.0010.4576.
4. Filipowicz S.F., Rymarczyk T., *Measurement Methods and Image Reconstruction in Electrical Impedance Tomography*, Przeglad Elektrotechniczny, vol. 88, Issue 6, pp.247-250. 2012.
5. Gola A., Świć A., *Computer-Aided Machine Tool Selection for Focused Flexibility Manufacturing Systems Using Economical Criteria*, Actual Problems of Economics, vol. 124, Issue 10, pp. 383-389, 2011.
6. Holder D.S., *Electrical Impedance Tomography: Methods, History and Applications*, Series in Medical Physics and Biomedical Engineering, London, 2005
7. Kłosowski G, Kozłowski E, Gola A., *Integer linear programming in optimization of waste after cutting in the furniture manufacturing*. Advances in Intelligent Systems and Computing 2018; vol. 637, pp. 260-270, 2018.
8. Kryszyn J., Smolik W., *Toolbox for 3D Modelling and Image Reconstruction in Electrical Capacitance Tomography*, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska - Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, ISSN: 2083-0157, vol. 1, pp. 137-145, 2017.
9. Kryszyn J, Wróblewski P, Stosio M, Wanta D, Olszewski T, Smolik W, *Architecture of EVT4 data acquisition system for electrical capacitance tomography*, Measurement: Journal of the International Measurement Confederation, vol. 101, pp. 28-39, 2017.
10. Polakowski, K., Filipowicz, S.F., Sikora, J., Rymarczyk, T., *Quality of imaging in multipath tomography*, Przeglad Elektrotechniczny, vol. 85, Issue 12, pp. 134-136, 2009.
11. Rymarczyk T., *Characterization of the shape of unknown objects by inverse numerical methods*, Przeglad Elektrotechniczny, vol. 88, Issue 7B, pp. 138-140, 2012.
12. Rymarczyk T., Adamkiewicz P., Rymarczyk P. et al., *Monitoring Flood Embankment by Electrical Impedance Tomography*, International Interdisciplinary Phd Workshop 2016, pp. 63-67, 2016.
13. Rymarczyk T., *Using electrical impedance tomography to monitoring flood banks*, International Journal of Applied Electromagnetics and Mechanics, vol. 45, pp. 489–494, 2014.
14. Smolik W., *Forward Problem Solver for Image Reconstruction by Nonlinear Optimization in Electrical Capacitance Tomography*, Flow Measurement and Instrumentation, vol. 21, pp. 70–77, 2010.
15. Soleimani M., Mitchell CN, Banasiak R., Wajman R., Adler A., *Four-dimensional electrical capacitance tomography imaging using experimental data*, Progress In Electromagnetics Research, vol. 90, pp. 171-186, 2009.
16. Wajman, R., Banasiak, R., *Tunnel-based method of sensitivity matrix calculation for 3D-ECT imaging*. Sensor Review, vol. 34, pp. 273–283, 2014. doi:10.1108/SR-06-2013-692
17. Wang M., *Industrial Tomography: Systems and Applications*, Elsevier, 2015

Maciej PAŃCZYK[3], Tomasz RYMARCZYK, [12], Jan SIKORA[3,4]

Solving Inverse Problem Based on 2-Layer Capacitor as Reference Model of Damp Wall in Tomography Problem

In order to verify the model of the wall, the boundary element method was used to search for the distribution of dampness, mainly its height and shape of the boundary which 1conventionally separates the areas considered dry and damp. From the point of view of a simple task implemented on the basis of the boundary element method, the solution was enriched with the problem of functional heterogeneity by introducing (the modified Green function and infinite boundary elements). This approach allows for a faster and more direct calculation of the specific humidity limit in place of the distribution map.

1. Introduction

The geometry of a 2-dimensional model of a flat capacitor and a damp wall corresponds to a straight rectangular area. The damp area created by the capillary drainage of badly insulated walls and foundations (while drying the sun, wind and heating inside the building) resembles the parabolic shape (cross-sectional wall) [10,16]. The height of the damp point is usually the basic criterion in the dehumidification process. In order to verify the final wall model, the previously used methods of searching for the distribution of dampness have been tested well in the model of known analytical solution or the solution obtained by approximate numerical methods such as FEM or BEM [1,2,4,7,8,10,14,18-20].

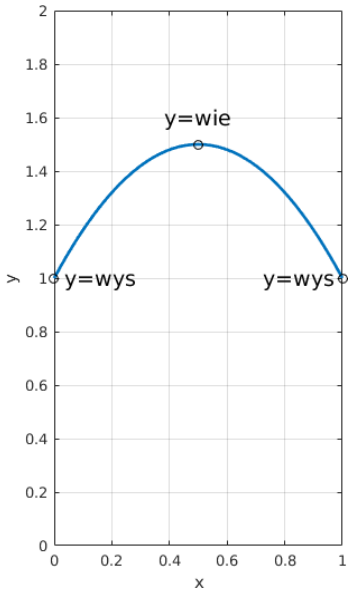


Fig. 1. Searching parameters "wys" and "wie" the parabola boundary of two areas ($y(x) = 4(wys - wie)x^2 - 4(wys - wie)x + wys$).

[1)] Research and Development Center, Netrix S.A., Związkowa 26, 20-148 Lublin, POLAND

[2)] University of Economics and Innovation in Lublin, Projektowa 4, 20-209 Lublin, POLAND

[3)] Lublin University of Technology, Nadbystrzycka 38d, 20-618, Lublin, Poland

[4)] Electrotechnical Institute, ul. Pożaryskiego 28, 04-703 Warszawa

The authors assumed that a 2-layer capacitor model would be the values of the quadratic equation describing the parabolic boundary of the centres (for testing the correctness, speed and obtaining relevant information about the distribution of material parameters). Due to the fact that in the wall model, the height of the parabolic vertex and its value at the edge of the wall (parabolic equation passing through 3 points) can be considered as the first step, to find the value of the coefficients „wys" and „wie" are shown in Figure 1. Figure 2 presents two-layer model of a flat capacitor with a parabolic boundary between the boundaries searched in the inverse problem.

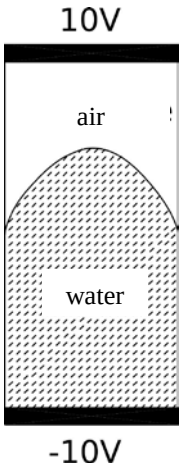


Fig. 2. Two-layer model of a flat capacitor with a parabolic boundary between the boundaries searched in the inverse problem.

2. Forward problem

The field studies might be split on two main branches. In that case are defined: the topology of the structure (interface boundary – outside and/or inside), boundary conditions, material coefficients (e.g. conductivity), the internal source or sources etc. Physical phenomena are described usually by sets of differential equations. Numerical techniques give us opportunity to find approximate solutions of differential equations which cannot be solved by means of analytical ones. Among various numerical tools let us concentrate our attention on the boundary element method (BEM). This method can be effectively employed on condition that partial differential equation can be transformed to integral form. Additionally, the Green's function has to be calculated. In the forward problem of calculating boundary values based on the assumed distribution of material parameters within a region, Laplace's equation was solved by the boundary element method. After transforming the differential equation into BIE (Boundary Integral Equation) and the discretization of boundary elements, the equation of the form [12,18,19] is solved:

$$C(\mathbf{r})\Phi(r) + \sum_{j=0}^{M-1} \Phi_j(\mathbf{r}') \int_{-1}^{+1} \frac{\partial G(|\mathbf{r}-\mathbf{r}'|)}{\partial n} J(\xi) d\xi = \sum_{j=0}^{M-1} \frac{\partial \Phi_j(\mathbf{r}')}{\partial n} \int_{-1}^{+1} G(|\mathbf{r}-\mathbf{r}'|) J(\xi) d\xi \tag{1}$$

where: r – is a fixed external point the so-called point of observation, in which the value of Φ was determined, r' – is a source point, coefficient C(r)=0,5, M is the total number of elements, Φ represents a potential function (electric potential for an electrostatic field), G is the fundamental solution of the Laplace equation, J – is Jacobian, which transforms with global coordinates (r) to local coordinates (ξ). The values Φ and ∂Φ/∂n (normal derivative Φ, n – a normal vector directed to the outside of the area), are treated as constants for everyone

element, these values are determined in half the length of each edge element (implemented zero-order). By writing functions from equation (1) as elements $A_{i,j}$ and $B_{i,j}$ we get an equation:

$$C(\mathbf{r})\Phi(\mathbf{r}) + \sum_{j=0}^{M-1} \Phi_j(\mathbf{r}')A_{i,j}(\mathbf{r},\mathbf{r}') = \\ = \sum_{j=0}^{M-1} \frac{\partial \Phi_j(\mathbf{r}')}{\partial n} B_{i,j}(\mathbf{r},\mathbf{r}') \tag{2}$$

Index *i* with equation (2) means the source points of each element. For each node (observation point) r The integration indicated in the equation is performed (1). Finally, this leads to the matrix writing of the equation (1) of the form:

$$\mathbf{A} \cdot \mathbf{\Phi} = \mathbf{B} \cdot \frac{\partial \mathbf{\Phi}}{\partial \mathrm{n}} \tag{3}$$

The [A] and [B] matrices contain functions as appropriate $A_{i,j}$ and $B_{i,j}$.
For every M element, the Dirichlet boundary condition or the Neumann boundary condition is known. Hence, the equation (3) must be ordered in order to enter the required boundary conditions and group them in vector **x** containing unknown and vector **b'**, containing data (known boundary conditions). After appropriate transformations, equation (3) will take the form:

$$\mathbf{Ax} = \mathbf{Bb'} \tag{4}$$

In equation (4), arrays A and B are formed by combining columns from both arrays depending on the boundary elements, depending on whether you know the values Φ or ∂Φ/∂n in the given element j. By multiplying matrix **B** and vector **b'**, the **rhs** vector is obtained and finally the equation takes the form:

$$\mathbf{Ax} = \mathbf{rhs} \tag{5}$$

In the inverse problem, which requires multiple solutions of the forward problem, such a distribution of material parameters is searched which results in well-known boundary values and complementary measurements of the potential distribution at the edge of the area.

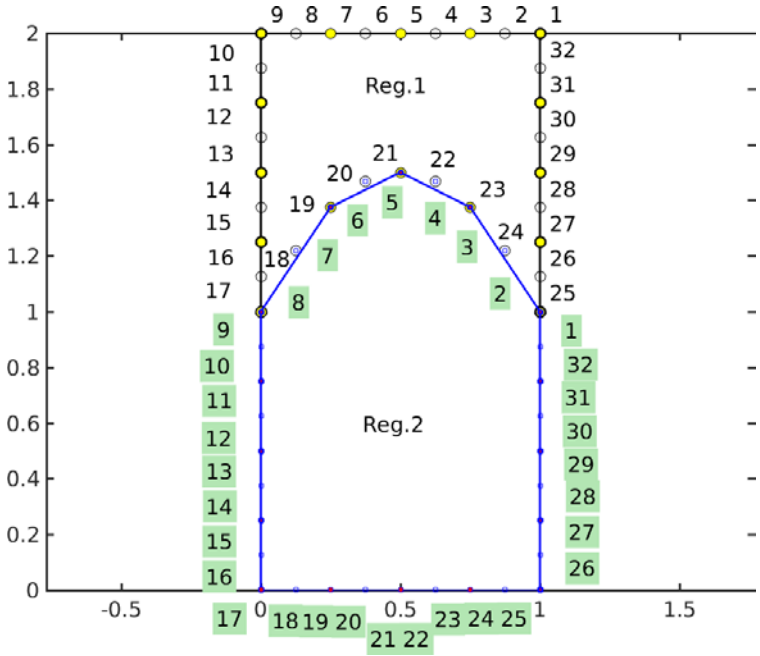


Fig. 3. An example numerical model of a two-layered flat capacitor realised by the boundary element method consisting of 32 elements of the zero order.

The numerical model of a two-layered parabolic flat-layer capacitor was made in a Matlab environment. A case in which each side is divided into 4 boundary elements is shown in Figure 3. The boundary conditions were: a Dirichlet condition enforcing a 10V potential on the first 4 elements of region 1 (nodes 1 to 9 of region 1 "upper") and -10V on 9-12 elements (nodes 17-25 of region 2 "lower") and Neumana on the rest sides of area. On the common border of the edges defined continuity of boundary conditions.

3. Calculations

In the inverse problem, parameters "wys" and "wie" were searched for the parabolic boundary between the two areas - Figure 3. The "measurement" values were derived from the case simulation, where the parameters wys = 1.0 and wie = 1.5. The starting point was equal to wys = 0.75 and wie = 1.75. The potential on the top cover of the capacitor was 10V and on the lower -10V. Solving the inverse problem in a Matlab environment required a 41-fold solution of a simple task by the boundary elements [2,4,7,8,10,12,18-19]. The total calculation time was 1.7 seconds, after which the desired parameters were found with the specified accuracy of 10-6.

The iterative search for the appropriate parabolic boundary between the areas is shown in Figure 4. The thin red lines were plotted with 4 elementary broken lines corresponding to the simplified up to 4 parabolic elements of subsequent iterations. Thick blue line with marked nodes reflects the solution you are looking for.

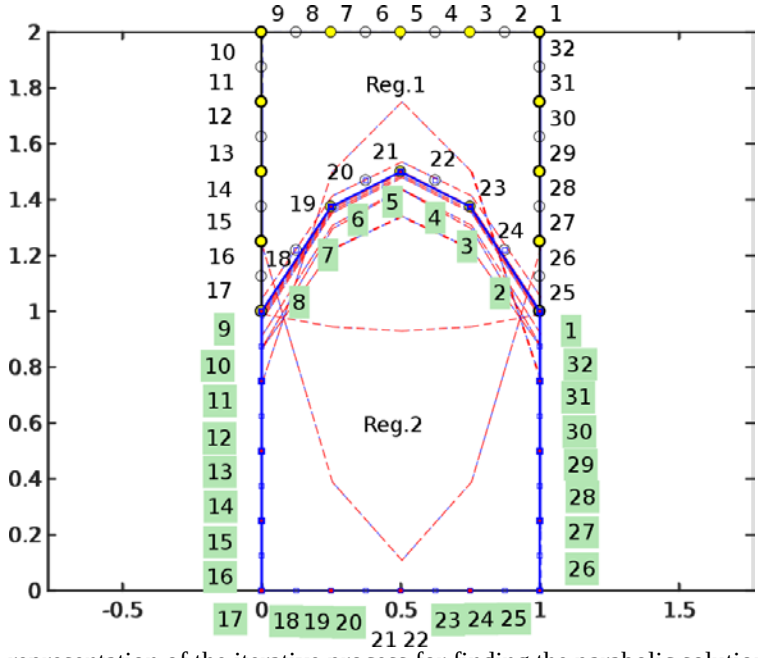


Fig. 4. Graphical representation of the iterative process for finding the parabolic solution of the boundary between areas in a two-layered flat capacitor model. The parabola represented in the 4 element broken line resulting from the assumed division of the side of the region into 4 boundary elements.

To solve the inverse problem the was used the Matlab function fmincon and the objective function defined as:

$$f_{objective} = \frac{(\Phi_{calculated} - \Phi_{measured})^2}{2} \tag{6}$$

The results of calculations in the Matlab environment and, in addition, the forward problem in the Comsol Multiphysics (for the obtained solution parameters and the high density of the component grid) in the form of the potential distribution map are shown in Figures 5a and 5b respectively.

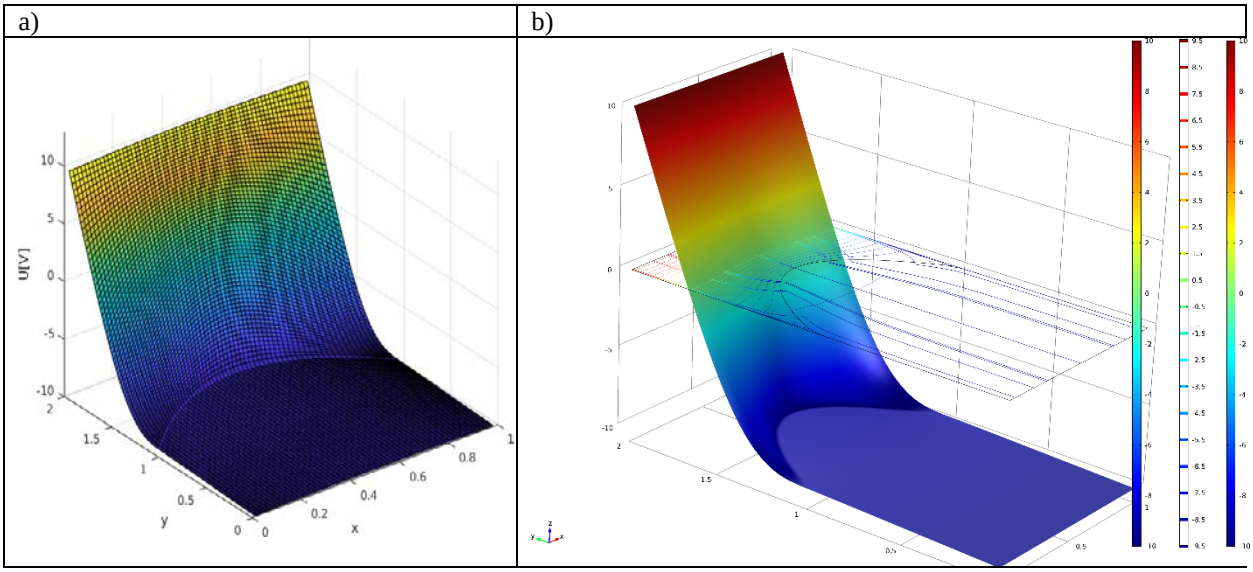


Fig. 5. Results of calculations: a) in Matlab environment, b) solution of forward problem in Comsol Multiphysics environment in the form of potential distribution map.

4. Conclusion

In this paper, there was presented the method to solve the moisture wall problem. The boundary element method was applied to solve this problem. The presented realization of the inverse problem does not entail special difficulties. Allows you to get the correct results and develop the issue with the next planned elements. In future works, there will be implemented level set method [3,15] and the topological derivative [17,23], which will allow the creation of a tomographic distribution of material parameters reflecting the number of regions with different parameters and their shape. Boundary element method will be enriched with the issues of functional heterogeneity [5,6,9,11-13,21,22], by introducing in the BEM the so-called the modified Green function [9,12] and infinite boundary elements [12]. It will allow to find the distribution of electrical permittivity (and in the target wall model of electrical conductivity associated with dampness). This approach will allow for faster and more direct (calculation of the specific humidity limit in place of the distribution map) where the wall does not contain defects or specific structural components, in most cases.

References

1. Banasiak R, Wajman R, Sankowski D, Soleimani M., *Three-Dimensional Nonlinear Inversion of Electrical Capacitance Tomography Data Using a Complete Sensor Model*. Progress In Electromagnetics Research (PIER) vol. 100, pp. 219-234, 2010.
2. Beer G., *Programming the Boundary Element Method. An Introduction for Engineers*, John Wiley & Sons Ltd., Chichester, 2001. M.H. Aliabadi. The Bouundary Element Method Volume 2. John Willey & Sons, Inbunden, 2005.
3. Filipowicz S.F., Rymarczyk T., *The Shape Reconstruction of Unknown Objects for Inverse Problems*, Przeglad Elektrotechniczny, vol. 88, Issue 3A, pp. 55-57, 2012.
4. Frijns J.H.M., de Snoo S. L., Schoonhoven R., *Improving the accuracy of the boundary element method by the use of second–order interpolation functions*, IEEE Transactions on Biomedical Engineering, 47(10):1336–1346, 2000.
5. Gray, L. J. and Kaplan, T. and Richardson, J. D. and Paulino, G. H. *Green's Functions and Bounday Integral Analysis for Exponentially Graded Materials: Heat Conduction*, Journal od Applied Mechanics Vol. 70 p.543-549 July 2003
6. Hoła J., Matkowski Z., Schabowicz K., *Impedance tomographic method of assessing*

the dampness of masonry, Z. 1-B/2007 Wydawnictwo Politechniki Krakowskiej
7. Jabłoński P., *Metoda elementów brzegowych w analizie pola elektromagnetycznego*. Wydawnictwo Politechniki Częstochowskiej, Częstochowa, 2003.
8. Jaworski., *Metoda Elementów Brzegowych: Zagadnienia potencjalne*, Oficyna Wydawnicza Politechniki Warszawskiej, 2000.
9. Kurgan, E., *Analiza pola magnetostatycznego w środowisku niejednorodnym metodą elementów brzegowych*, Uczelniane wydawnictwa naukowo-dydaktyczne AGH Kraków, 1999.
10. Majchrzak E., *Metoda Elementów Brzegowych w przepływie ciepła*, Wydawnictwo Politechniki Częstochowskiej, Częstochowa, 2001.
11. Paulino, G. H. and Sutradhar, A., *The simple boundary element method for multiple cracks in functionally graded media governed by potential theory: a three-dimensional Galerkin approach*, International Journal for Numerical Methods in Engineering No. 65 p. 2007-2034, 2005.
12. Pańczyk M., Sikora J., *Brzegowe elementy nieskończone: podstawy analizy obszarów nieograniczonych metodą elementów brzegowych*, Wydawnictwo IEL. Warszawa 2014.
13. Paulino, G. H. and Sutradhar, A. and Gray, L. J., *Boundary element Methods for Functionally Graded Materials. International Association for Boundary Element Methods*, IABEM 28-30 May 2002 Austin.
14. Polakowski, K., Filipowicz, S.F., Sikora, J., Rymarczyk, *Tomography technology application*, Przeglad Elektrotechniczny, vol. 84, Issue 12, pp. 227-229, 2008.
15. Rymarczyk T., *New Methods to Determine Moisture Areas by Electrical Impedance Tomography*, International Journal of Applied Electromagnetics and Mechanics, vol. 37, Issue 1-2, pp.79–87, 2016.
16. Rymarczyk T., Adamkiewicz P., Duda K., Szumowski J., Sikora J., *New Electrical Tomographic Method to Determine Dampness in Historical Buildings*, Archives of Electrical Engineering, vol. 65, Issue 2, pp. 273-283, 2016.
17. Rymarczyk T., Tchórzewski P., Adamkiewicz P., Duda K., Szumowski J., Sikora J., *Practical Implementation of Electrical Tomography in a Distributed System to Examine the Condition of Objects*, IEEE Sensors Journal, vol. 17, Issue:24, 2017, pp. 8166-8186, DOI: 10.1109/JSEN.2017.2746748.
18. Sikora J, *Numeryczne metody rozwiązywania zagadnień brzegowych. Podstawy metody elementów skończonych i metody elementów brzegowych*, Wydawnictwa uczelniane Politechniki Lubelskiej. Lublin 2009.
19. Sikora J., *Boundary Element Method for Impedance and Optical Tomography*, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2007
20. Smolik W., Kryszyn J., Radzik B., Stosio M., Wróblewski P., Wanta D., Dańko L., Olszewski T., Szabatin R., *Single-shot high-voltage circuit for electrical capacitance tomography*, Measurement Science and Technology, vol. 28, Issue 2, p. 025902, 2017.
21. Sutradhar, A. and Paulino, G. H. and Gray, L. J., *On hypersingular surface integrals in the symmetric Galerkin boundary element method: application to heat conduction in exponentially graded materials*, International Journal for Numerical Methods in Engineering No. 62 p.122-157, 2005.
22. Sutradhar, A. and Paulino, G. H., *A simple boundary element method for problems of potential in non-homogeneous media*, International Journal for Numerical Methods in Engineering No, 60 p. 2203-2230, 2004.
23. Tchorzewski P., Rymarczyk T., Sikora J., *Using Topological Algorithms to Solve Inverse Problem in Electrical Impedance Tomography*, International Interdisciplinary Phd Workshop 2016, pp. 46-50, 2016.

Jakub SZUMOWSKI[1], Tomasz RYMARCZYK, [12], Przemysław ADAMKIEWICZ[1], Tomasz CIEPLAK[3], Jan SIKORA[4,5]

Prototype of acquisition data module for Hybrid Tomograph Systems

In this paper, there was presented the project of hybrid electrical tomography device, which is currently under development. It assumes making use of two measuring methods: electrical capacitance tomography and electrical impedance tomography and allows to perform up to 32-channel measurements. The tomographic system consists of a central processing unit and a set of electrodes and a software solution in the form of a computing cloud and a Big Data cluster for processing, visualizing and analysing data.

1. Introduction

The Hybrid Tomograph combining electrical capacitive tomography (ECT) [1,3,5,9,10,15-18] and electrical impedance tomography (EIT) [2,4,7,12-14] technology is designed to provide a non-invasive way to test the spatial distribution of moisture. The tomographic system consists of a central processing unit and a set of electrodes and a software solution in the form of a computing cloud and a Big Data cluster for processing, visualizing, writing and analysing data [6-8,11]. Both solutions use MQTT protocol to communicate with each other to ensure speed and scalability.

2. Measurement system

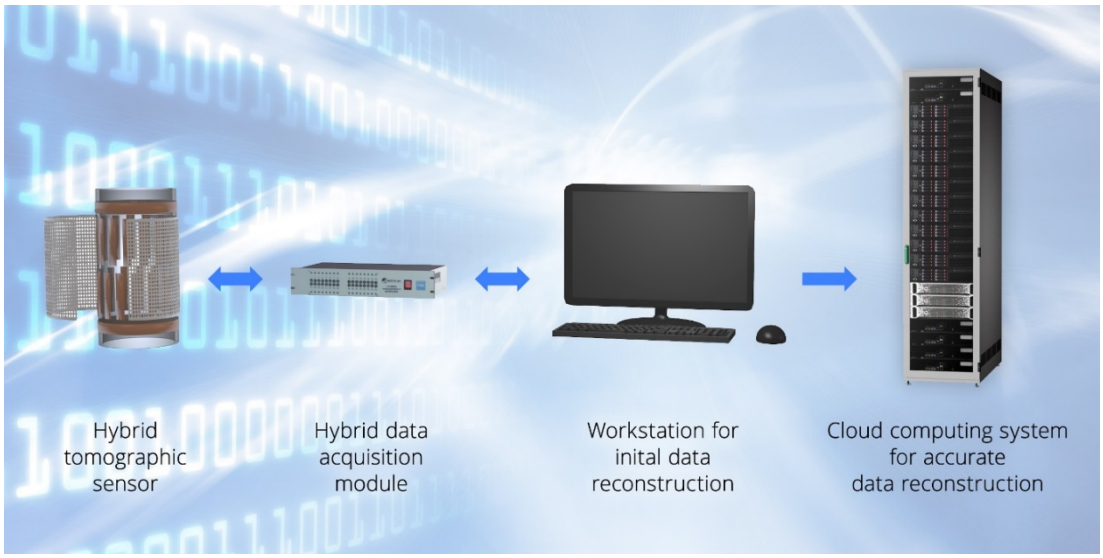


Fig 1. Model of the measurement system.

[1)] Research and Development Center, Netrix S.A., Związkowa 26, 20-148 Lublin, POLAND

[2)] University of Economics and Innovation in Lublin, Projektowa 4, 20-209 Lublin, POLAND

[3] Lublin University of Technology, Nadbystrzycka 38, 20-618 Lublin, POLAND

[4)] Lublin University of Technology, Institute of Electronics and Information Technology, Nadbystrzycka 38d, 20-618, Lublin, Poland

[5)] Electrotechnical Institute, ul. Pożaryskiego 28, 04-703 Warszawa

The project of hybrid electrical tomography device is currently under development. It assumes making use of two measuring methods: electrical capacitance tomography and electrical impedance tomography and allows to perform up to 32-channel measurements. It consists of several key parts closed in a single 2U Schroff chassis (Fig. 9). Idea of the measurement system was presented in Figure 1 and 2.

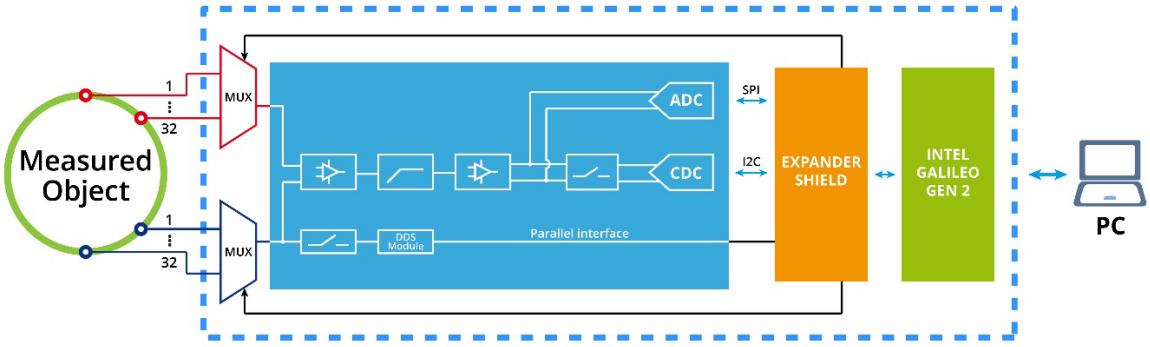


Fig 2. The measurement system.

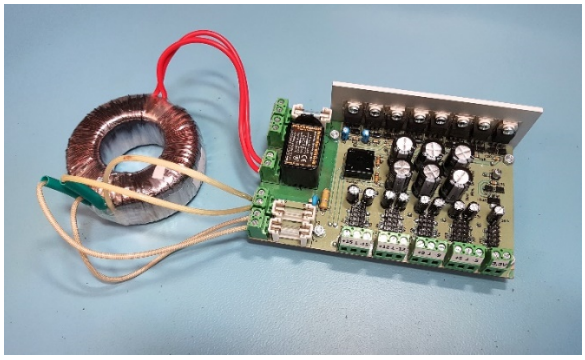


Fig. 3. The block diagram of the data acquisition module.

Stabilized power supply is equipped with symmetrical voltage outputs of ±15 V, ±12 V, ±9 V, ±5 V, ±3.3 V and may provide current up to 1 A. Due to filters applied frequency-response characteristic of output signal is flat regardless of a load level. Power supply board is presented in Figure 3. Another crucial part of hybrid tomography scanner is based on CMOS logic 64-channel binary driven multiplexer. It is mounted vertically inside the chassis so specific design requirements needed to be met. The component is divided into two printed circuit boards connected together with 96 pins, both parts being mounted directly to the front panel. They work independently but share common 32 SMB sockets. Joined parts form a 4- to 32-channel Analog multiplexer with low capacitance offset. Single channel is equipped with a set of two SPST CMOS Vishay Electronics DG413 Analog switches. Every 16 channels are controlled with a single high speed CMOS 4- to 16-line Texas Instruments CD74HC154M decoder. Solution presented in Figure 29 is much more stable than integrated multiplexers used in Smart ECT. It provides less noise, similar capacitance offset on every channel and it is compatible with any control board working in 5V logic and fitted with 20 control pins. Schematics of a single channel and a set of decoders driven by logic components are shown in Figure 4.

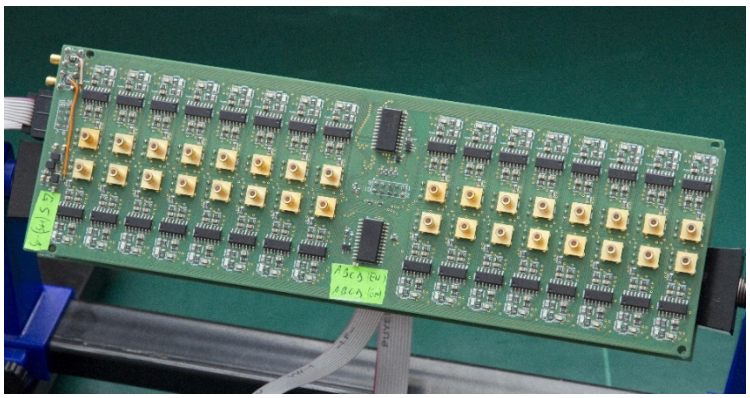


Fig. 4. Hybrid Tomograph Multiplexer - view from front panel.

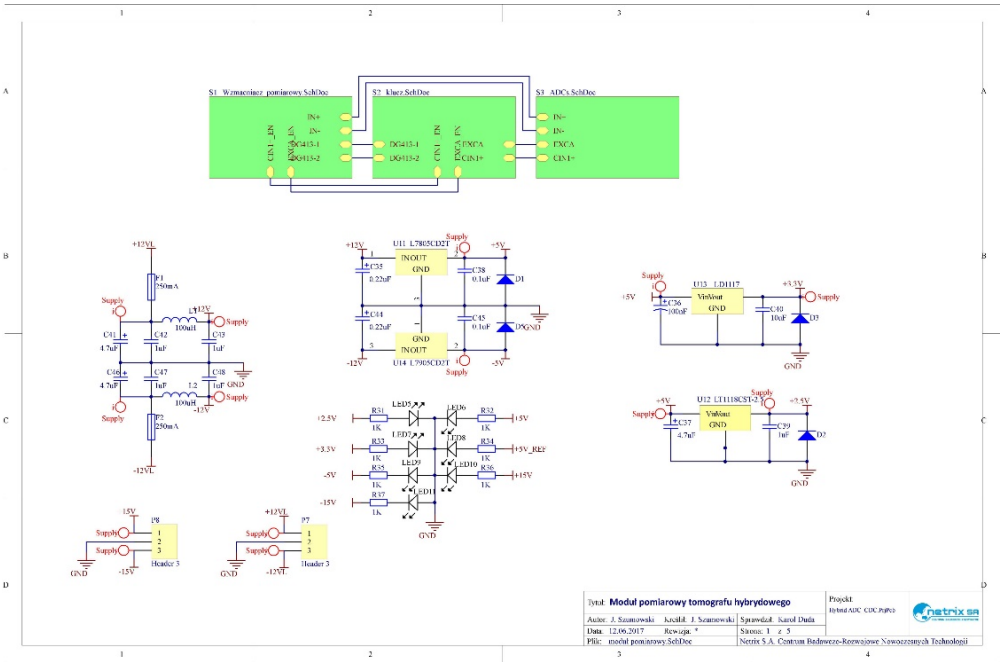


Fig. 5. Block diagram of the measuring module.

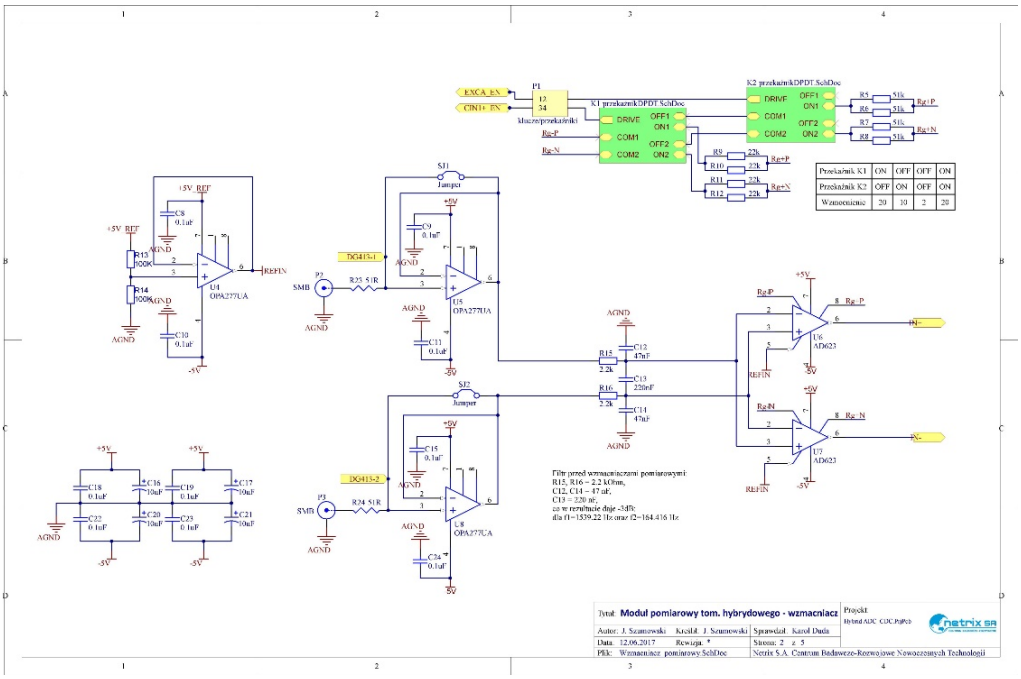


Fig. 6. AD7989-1 transducer input amplifier circuit.

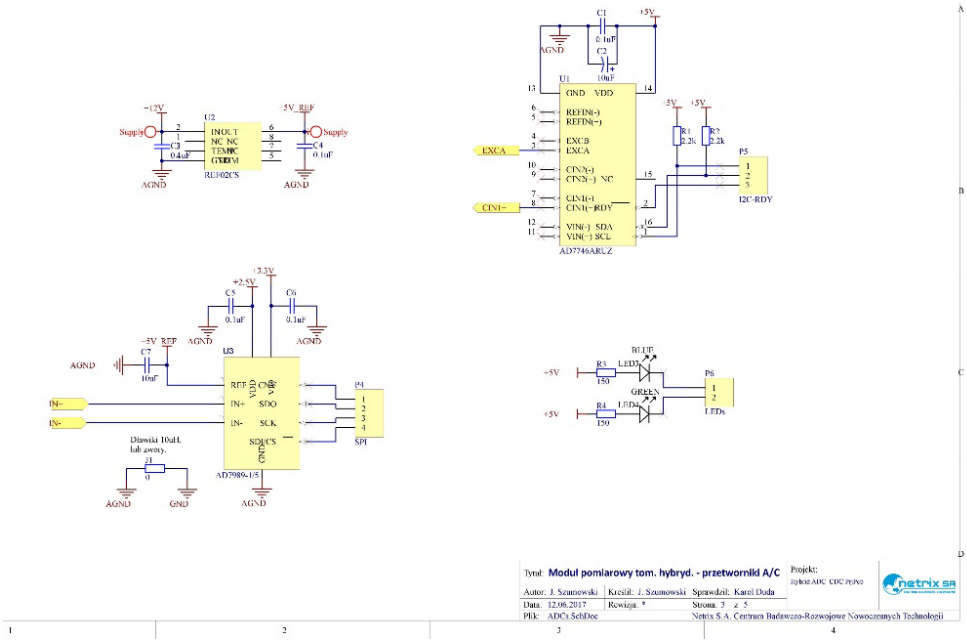


Fig. 7. Analog Devices AD7746 Analog-to-Digital Converters and Analog Devices AD7989-1.

The measuring module, despite its small physical dimensions (80x60 mm), contains two analogue-to-digital converters, whose measurement paths are separated by CMOS analog keys, the adjustable gain signal section, and its own voltage stabilization section ± 5 V, 3.3 V and 2, 5 V. The block diagram of the connections between transducers and stabilizers is presented in Figure 5, the amplifier circuit in Figure 6 and analog-to-digital converters in Figure 7. The separation keys on the signal lines are implemented analogously to the multiplexer keys. The transducers used are Analog Devices AD7989-1 for voltage measurement (EIT) and Analog Devices AD7746 for capacitance measurement (ECT). The prototype of the measuring module is shown in Figure 8.

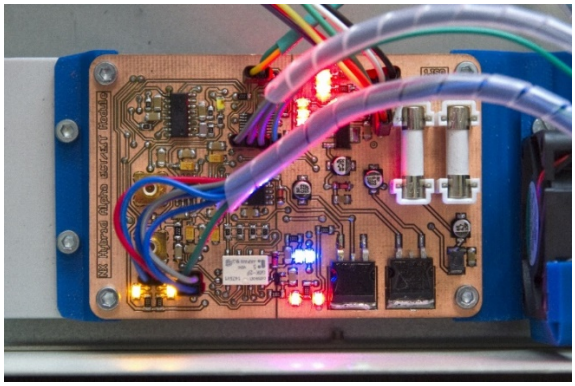


Fig. 8. Measuring module.

Two last parts of hybrid tomography scanner are control board and analog converters board. They may be merged into one PCB - pros and cons of such design decisions are currently being discussed. For testing purposes of formerly described parts Intel Galileo 2 has been chosen as a control board for both analog converters and multiplexer due to its flexibility and versatility. Its insufficient number of general purpose inputs/outputs was extended with a custom-built Arduino-compatible expander shield with two Microchip MCP23017 ICs onboard. MQTT communication protocol was implemented allowing for remote control over the Internet. The control circuit is engaged in acquiring information from analog converters but also in preprocessing and sending it further for image reconstruction. The secondary aim of the project is to enable Big Data systems to gather, store and calculate such information in

real time. Thus, every frame of information has to be formatted into JSON and sent to a specified server with MQTT protocol.

Data acquisition part will be equipped with two independent Analog to digital converters: Analog Devices AD7989-1 18-bit 100kSPS SAR ADC and AD7745 24-bit CDC being single-channel variation of AD7746 used in Smart ECT. They will enable hybrid measurements and user will have the opportunity of precisely tuning their parameters and behavior. It will be possible to choose only one tomographic technique or both of them, run sequentially.

Fig.9. Hybrid Tomograph.

5. Conclusion

This paper's aim is to present a project of the hybrid tomograph. The device is combining electrical capacitive tomography and electrical impedance tomography to provide a non-invasive way to test the spatial distribution of moisture. The system consists additional a software solution in the form of a computing cloud and a Big Data cluster for processing, visualizing, writing and analysing data. Future works there will concentrate on modification the hybrid measuring module, preparing a prototype's tests with ECT and EIT measurements and correction of irregularities and design errors.

References

1. Beck M. S., Byars M., Dyakowski T., Waterfall R., He R., Wang S. J., Yang W. Q., *Principles and Industrial Applications of Electrical Capacitance Tomography*, Measurement and Control, September, Vol. 30, No. 7, 1997.
2. Duda K., Adamkiewicz P., Rymarczyk T., *Nondestructive Method to Examine Brick Wall Dampness*, International Interdisciplinary Phd Workshop 2016, pp. 68-71, 2016.
3. Fiderek P., Jaworski T., Banasiak R., Nowakowski J., Kucharski J., Wajman R., *Intelligent system for the two-phase flows diagnosis and control on the basis of raw 3D*

3D ECT data, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 7, Issue 1, pp. 17-23, 2017; DOI: 10.5604/01.3001.0010.4576.
4. Filipowicz S.F., Rymarczyk T., *Measurement Methods and Image Reconstruction in Electrical Impedance Tomography*, Przeglad Elektrotechniczny, vol. 88, Issue 6, pp.247-250. 2012.
5. Garbaa H., Jackowska-Strumiłło L., Grudzień K., Romanowski A., *Simulation of gravitational solids flow process and its parameters estimation by the use of electrical capacitance tomography and artificial neural networks*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 6, Issue 2, pp. 34–37, 2016.
6. Gola A., Świć A., *Computer-Aided Machine Tool Selection for Focused Flexibility Manufacturing Systems Using Economical Criteria*, Actual Problems of Economics, vol. 124, Issue 10, pp. 383-389, 2011.
7. Holder D.S., *Electrical Impedance Tomography: Methods, History and Applications*, Series in Medical Physics and Biomedical Engineering, London, 2005
8. Kłosowski G, Kozłowski E, Gola A., *Integer linear programming in optimization of waste after cutting in the furniture manufacturing*. Advances in Intelligent Systems and Computing 2018; vol. 637, pp. 260-270, 2018.
9. Kryszyn J., Smolik W., *Toolbox for 3D Modelling and Image Reconstruction in Electrical Capacitance Tomography*, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska - Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, ISSN: 2083-0157, vol. 1, pp. 137-145, 2017.
10. Kryszyn J, Wróblewski P, Stosio M, Wanta D, Olszewski T, Smolik W, *Architecture of EVT4 data acquisition system for electrical capacitance tomography*, Measurement: Journal of the International Measurement Confederation, vol. 101, pp. 28-39, 2017.
11. Polakowski, K., Filipowicz, S.F., Sikora, J., Rymarczyk, T., *Quality of imaging in multipath tomography*, Przeglad Elektrotechniczny, vol. 85, Issue 12, pp. 134-136, 2009.
12. Rymarczyk T., *Characterization of the shape of unknown objects by inverse numerical methods*, Przeglad Elektrotechniczny, vol. 88, Issue 7B, pp. 138-140, 2012.
13. Rymarczyk T., Adamkiewicz P., Rymarczyk P. et al., *Monitoring Flood Embankment by Electrical Impedance Tomography*, International Interdisciplinary Phd Workshop 2016, pp. 63-67, 2016.
14. Rymarczyk T., *Using electrical impedance tomography to monitoring flood banks*, International Journal of Applied Electromagnetics and Mechanics, vol. 45, pp. 489–494, 2014.
15. Smolik W., Kryszyn J., Olszewski T., Szabatin R., *Methods of Small Capacitance Measurement in Electrical Capacitance Tomography*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, IAPGOS, ISSN: 2083-0157, vol. 1, pp. 105-110, 2017.
16. Smolik W., *Forward Problem Solver for Image Reconstruction by Nonlinear Optimization in Electrical Capacitance Tomography*, Flow Measurement and Instrumentation, vol. 21, pp. 70–77, 2010.
17. Soleimani M., Mitchell CN, Banasiak R., Wajman R., Adler A., *Four-dimensional electrical capacitance tomography imaging using experimental data*, Progress In Electromagnetics Research, vol. 90, pp. 171-186, 2009.
18. Wajman, R., Banasiak, R., *Tunnel-based method of sensitivity matrix calculation for 3D-ECT imaging*. Sensor Review, vol. 34, pp. 273–283, 2014. doi:10.1108/SR-06-2013-692

Edward KOZŁOWSKI[1], Tomasz RYMARCZYK [23], Tomasz CIEPLAK[1]

Application Elastic Net to Identification of Objects in EIT

The multicollinearity problem in linear models (the predictors of the linear model are highly correlated with each other) does not allow to determine exactly influence of each predictor on the response variable. This can lead to the poor prediction. The addition of penalty depending on quantities of parameters to the least square criterion allows us to determine the biased estimators but also to reduce the variance of estimators. To image reconstruction in electrical impedance tomography the elastic net was used, which is connection of two methods: ridge regression and LASSO.

1. Introduction

Electrical impedance tomography (EIT) image reconstruction problem is an ill-posed inverse problem. In EIT, electrical voltages are injected into an object using a set of electrodes attached on the surface of the object and the potentials are measured. The conductivity of the object is reconstructed based on the known voltages and measured potentials. Electrical impedance tomography reconstruction requires accurate modelling. EIT is an imaging modality in which the conductivity distribution of an examined object is estimated from measurements of electrical voltages and electrode potentials at the boundary. To achieve quantitative information of the conductivity change, it would be preferable to use a non-linear model in the solution of the difference imaging [1-2,6-17]. There are many optimization methods [4-11]. The presented algorithm based on Elastic Net is new proposition in electrical impedance tomography.

2. Elastic net

Let the linear system be described by the state equation

$$Y = X\beta + \varepsilon, \tag{1}$$

where $Y \in R^n$, $X \in R^{n\times(k+1)}$ denote the observation matrices of response and input variables respectively, $\beta \in R^{k+1}$ denotes the vector of unknown parameters. When the linear model (1) contains the intercept, then the first column of matrix X is a column of ones. The object $\varepsilon \in R^n$ in linear system (1) presents a sequence of disturbances, which is usually defined as a vector of independent identically distributed random variables with normal distribution $N\left(\tilde{0},\sigma^2 I\right)$, which, $\tilde{0} \in R^n$ is a zeros vector but $I \in R^{n\times n}$ is an identity matrix. The classical Least Square Method depends on identification of unknown parameters $\beta = \left(\beta_0, \beta_1, ..., \beta_k\right)$ in (1) by solution the task

[1] Lublin University of Technology, Nadbystrzycka 38, 20-618 Lublin, POLAND

[2)] Research and Development Center, Netrix S.A., Związkowa 26, 20-148 Lublin, POLAND

[3)] University of Economics and Innovation in Lublin, Projektowa 4, 20-209 Lublin, POLAND

$$\min_{\beta \in R^{k+1}} \left\| Y - X\beta \right\|^2 \tag{2}$$

If $\det(X^T X) \neq 0$, then the best linear estimator of unknown parameters β is

$$\hat{\beta} = \left(X^T X \right)^{-1} XY$$

The problem is often when $X^T X$ is singular.

The main goal of tomography is to make an image reconstruction. During measurements we can see, that the measured values from some electrodes are strongly correlated (due to the way of measurement). In this case we have a multicollinearity problem. When the independent variables (predictors) are correlated (collinear), then the matrix $X^T X$ tends to singular matrix. Using the least square method we obtain the large absolute values of some estimators of unknown parameters. The predictions based on this model is unstable. The most common approach depends on reduction of set of the input variables (deleting the same predictors which involve in multicollinearity). Then we have a problem of deciding which predictor variables would be included to regression model. For example, by comparing AIC (Akaike Information Criterion) value for linear models with different set of predictors we may choice a best model.

An alternative way to solve this problem depends on decomposition $X = TP$ into orthogonal scores T and loadings P (see e.g. [19], [20]). Next, we determine the linear regression Y not on X but on first $m-$column of the scores T. The number m is determined by analyzing the root mean square error of prediction. In Principal Component Regression the scores and loadings are chosen to explain the variance of X and come from singular value decomposition. The scores are defined as the left singular vectors of X multiplied with the corresponding singular values, but the loadings are defined as the right singular vectors of X. In Partial Least Square Regression the scores and loadings are chosen to explain the covariance between X and Y.

Another way to determine the linear regression when the input variables are collinear depends on solution the task

$$\min_{(\beta_0, \beta') \in R^{k+1}} \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \beta_0 - x_i \beta' \right)^2 + \lambda P_\alpha(\beta'), \tag{3}$$

where $x_i = (x_{i1}, ..., x_{ik})$, $\beta' = (\beta_1, ..., \beta_k)$ for $1 \leq i \leq n$ and P_α is a elastic net penalty given by

$$P_\alpha(\beta') = (1-\alpha)\frac{1}{2}\left\| \beta' \right\|_{L_2} + \alpha \left\| \beta' \right\|_{L_1} = \sum_{j=1}^{k} \left(\frac{1-\alpha}{2} \beta_j^2 + \alpha \left| \beta_j \right| \right) \tag{4}$$

We see that the penalty is a linear combination of norms L_1 and L_2 of unknown parameters β'. The introduction the penalty function dependent from parameters to the objective function allows to shrink the estimators of unknown parameters.

The parameter λ in task (3) denotes the coefficient of penalty, but the parameter $0 \leq \alpha \leq 1$ creates the compromise between LASSO (Least Absolute Shrinkage and Selection Operator) and ridge regression. The ridge regression ($\alpha = 0$) is called Tikhonov regularization (see e.g [22]) and is one of the most commonly used for regularization of linear models. LASSO ($\alpha = 1$) was introduced by Roberta Tibshirani (see e.g [18], [22], [3]). This method performs the variable selection and regularization in linear statistical models. For the ridge regression the penalty is calculated in norm L_1 but for LASSO in L_2. Difference between ridge regression and LASSO is symbolic, only the norms are changed. The ridge regression shrinks

coefficients for correlated predictors towards each other. When the correlated predictors depend on any latent factor, then ridge regression allows to uniformly distribute the strength of latent factor on these predictors. Whereas LASSO is indifferent to correlated predictors. This method allows to determine the preferred predictor and to ignore the rest. By applying LASSO method we obtain a model, where the many coefficients to be close to zero, and as a result we receive a sparse model. The elastic net is a connection of ridge regression and LASSO (see e.g. [21]). Choosing the appropriate α we may create the compromise between ridge regression and LASSO.

By solution the task (3) for fixed λ and α we estimate the unknown parameters of linear system (1) with correlated predictors. Then the prediction based on model (1) is

$$\hat{Y} = X\hat{\beta}, \tag{5}$$

where $\hat{\beta} = \left(\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k\right)$ is a vector of estimators of unknown parameters.

3. Numerical examples

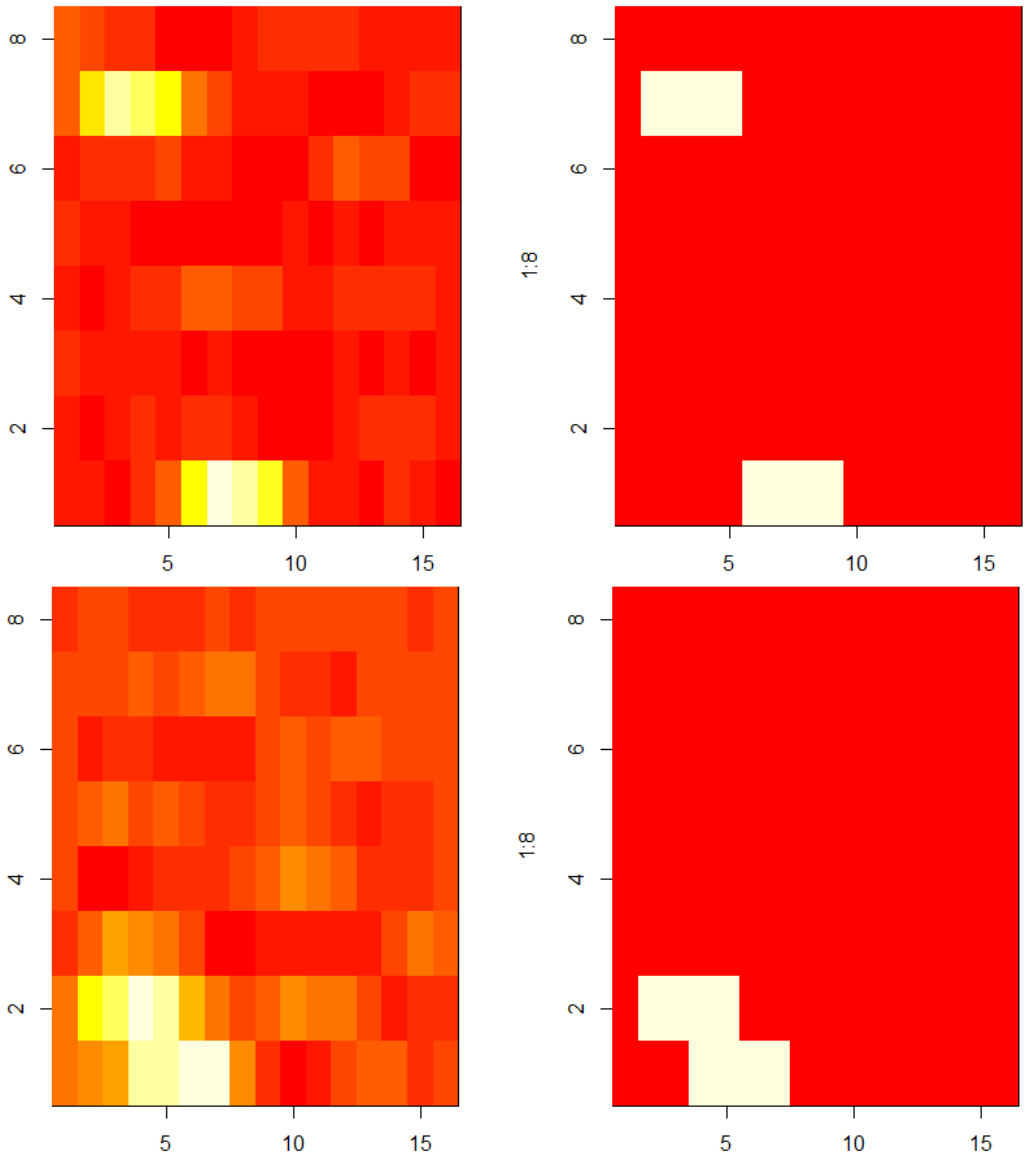


Fig. 1: The example reconstruction of 2 objects (left side – reconstruction, right side – real image).

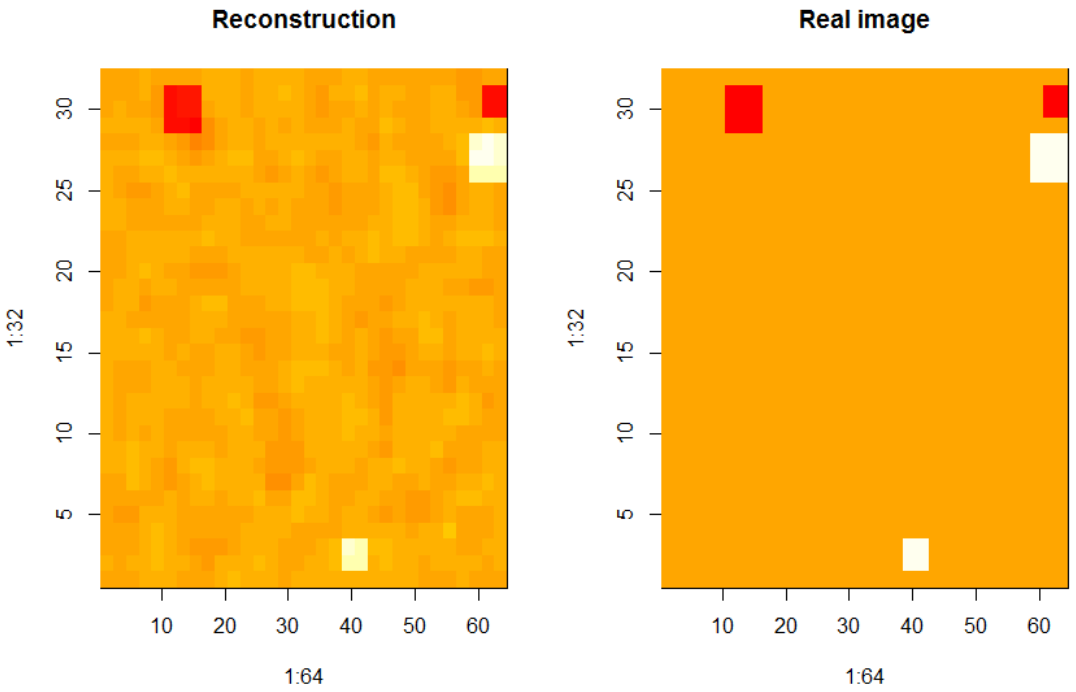


Fig. 2: The example reconstruction of 4 objects.

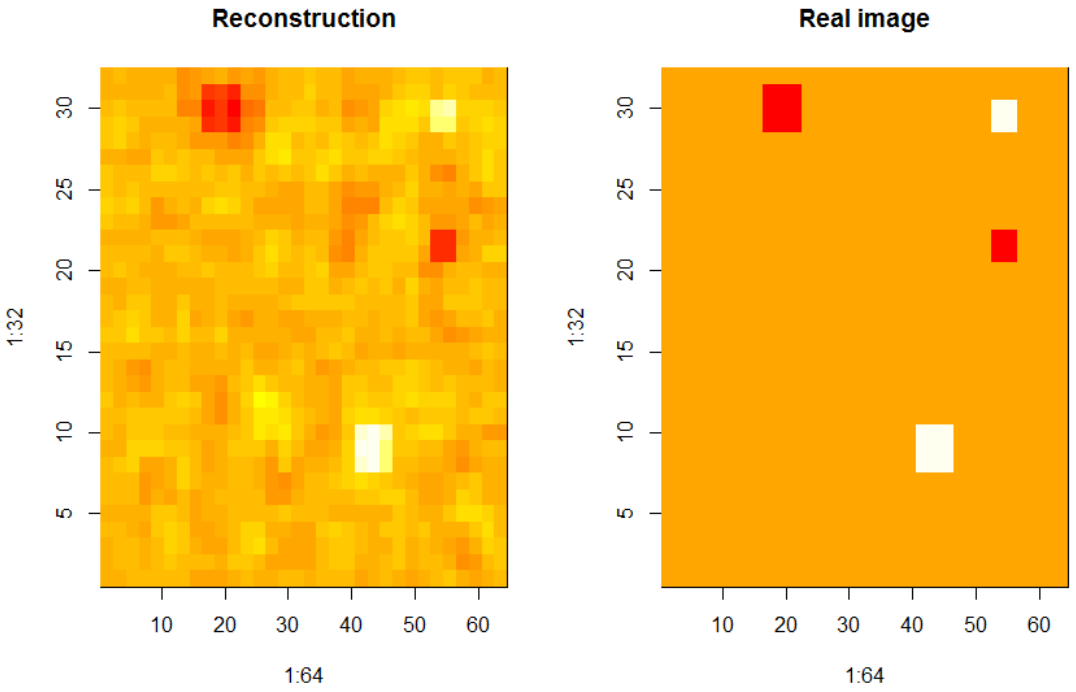


Fig. 3: The example reconstruction of 4 objects.

The measuring elements of EIT has 16 electrodes, so we make 208 measurements (in the linear model we have 208 predictors). Reconstruction looks like a 64 * 32 = 2048 image (for each pixel we construct a linear model with 208 correlated predictors: 16 projection angles times 13 voltages). Figure 1,2 and 3 show the examples of the image reconstruction with different kind of objects. Images were reconstructed with high precision, better than gradient methods.

3. Conclusion

In this paper, there were proposed algorithms based on statistical methods for the purpose of more accurate and stable reconstruction results in solving EIT inverse problem. To image reconstruction in electrical impedance tomography the elastic net was used, which is connection of two methods: ridge regression and LASSO. The combination of the proposed algorithms is effective in the simulation. This work gives promise results as a new horizon to solve practical problems.

References

1. Banasiak R, Wajman R, Sankowski D, Soleimani M., *Three-Dimensional Nonlinear Inversion of Electrical Capacitance Tomography Data Using a Complete Sensor Model*. Progress In Electromagnetics Research (PIER) vol. 100, pp. 219-234, 2010.
2. Filipowicz S.F., Rymarczyk T., *The Shape Reconstruction of Unknown Objects for Inverse Problems*, Przeglad Elektrotechniczny, vol. 88, Issue 3A, pp. 55-57, 2012.
3. Friedman J., Tibshirani R., Hastie T., *Regularization paths for generalized linear models via coordinate descent*, Journal of Statistical Software, Vol. 33, nr 1, 2010.
4. Fonseca T., Goliatt L., Campos L., Bastos F., Barra L., Santos R., *Machine Learning Approaches to Estimate Simulated Cardiac Ejection Fraction from Electrical Impedance Tomography*, Advances in Artificial Intelligence - IBERAMIA 2016, pp 235-246, 2016.
5. Kłosowski G., Gola A., *Risk-based estimation of manufacturing order costs with artificial intelligence* [in:] Ganzha M., Maciaszek L., Paprzycki M. (eds.), Proceedings of the 2016 Federated Conference on Computer Science and Information Systems (FEDCSIS), IEEE, pp. 729-732, 2016, DOI: 10.15439/2016F323.
6. Kryszyn J., Smolik W., *2D modelling of a sensor for electrical capacitance tomography in ECTSIM toolbox* , Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 1, pp. 146-149, 2017.
7. Kryszyn J., Smolik W., Radzik B., Olszewski T., Szabatin R., *Switchless charge-discharge circuit for electrical capacitance tomography*, Measurement Science and Technology, vol. 25, Issue 11, pp. 1-9, 2014.
8. Majchrowicz M., Kapusta P., Jackowska-Strumiłło L., Sankowski D., *Acceleration of image reconstruction proces in electrical capacitance tomography 3D in heterogeneous, multi-GPU system*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 7, Issue 1, pp. 37-41, 2017. DOI: 10.5604/01.3001.0010.4579
9. Polakowski, K., Filipowicz, S.F., Sikora, J., Rymarczyk, *Tomography technology application*, Przeglad Elektrotechniczny, vol. 84, Issue 12, pp. 227-229, 2008.
10. Rybak G., Chaniecki Z., Grudzień K., Romanowski A., Sankowski D., *Non–invasive methods of industrial process control*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 4, Issue 3, pp. 41–45, 2014.

11. Rymarczyk T., *New Methods to Determine Moisture Areas by Electrical Impedance Tomography*, International Journal of Applied Electromagnetics and Mechanics, vol. 37, Issue 1-2, pp.79–87, 2016.
12. Rymarczyk T., Adamkiewicz P., Duda K., Szumowski J., Sikora J., *New Electrical Tomographic Method to Determine Dampness in Historical Buildings*, Archives of Electrical Engineering, vol. 65, Issue 2, pp. 273-283, 2016.
13. Rymarczyk T., Tchórzewski P., Adamkiewicz P., Duda K., Szumowski J., Sikora J., *Practical Implementation of Electrical Tomography in a Distributed System to Examine the Condition of Objects*, IEEE Sensors Journal, vol. 17, Issue:24, 2017, pp. 8166-8186, DOI: 10.1109/JSEN.2017.2746748.
14. Tchorzewski P., Rymarczyk T., Sikora J., *Using Topological Algorithms to Solve Inverse Problem in Electrical Impedance Tomography*, International Interdisciplinary Phd Workshop 2016, pp. 46-50, 2016.
15. Wajman R., Fiderek P., Fidos H., Jaworski T., Nowakowski J., Sankowski D., Banasiak R., *Metrological evaluation of a 3D electrical capacitance tomography measurement system for two-phase flow fraction determination*, Measurement Science and Technology, vol. 24, 065302, 2013.
16. Smolik W., Kryszyn J., *Linea rover ranges iterative algorithms for image reconstruction in electrical capacitance tomoraphy*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 1, pp. 115-120, 2017.
17. Smolik W., Kryszyn J., Radzik B., Stosio M., Wróblewski P., Wanta D., Dańko L., Olszewski T., Szabatin R., *Single-shot high-voltage circuit for electrical capacitance tomography*, Measurement Science and Technology, vol. 28, Issue 2, p. 025902, 2017.
18. Tibshirani R., *Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society*, Series B, Vol. 58, nr. 1, pp. 267–288, 1996.
19. Wehrens R., Mevik B.H., *The pls package: Principal Component and Partial Least Square Regressian in R, Journal of Statistical Software*, Vol. 18, nr 2, 2007.
20. Wehrens R., *Chemometrics with R. Multivariate Data Analysis in the Natural Science and Life Sciences*, Springer 2011.
21. Zou H., Hastie T., *Regularization and variable selection via the elastic net*, Journal of the Royal Statistical Society, Series B, Vol. 67, nr 2, pp. 301–320, 2005.
22. Xin Yan, Xiao Gang Su, *Linear regression analysis: Theory and computing*, World Sciencific, 2009.

Edward KOZŁOWSKI[1], Tomasz RYMARCZYK[23], Przemysław ADAMKIEWICZ[2]

Applying PLSR to Detection of Objects in Electrical Impedance Tomography

The highly correlated predictors with each other's in linear models do not allow to determine the precisely influences of these predictors on the output variable. Directly application the least square method to estimate the unknown parameters may lead to a poor prediction. To solve this problem the correlated predictors can be decomposed into orthogonal components, which is called a Latent Variables. In Principal Component Regression, the latent variables are chosen to explain as much as possible the variance of predictors, but the Partial Least Square Regression concentrates on the covariance between predictors and response variables. The partial least square regression will be used in electrical impedance tomography.

1. Introduction

This paper presents the solution for the identification unknown object. Many optimization algorithms can be used to solve this problem [1-12,19]. Numerical methods were based on the Partial Least Square Regression. Discussed technique can be applied to the solution of inverse problems in the electrical impedance tomography. There were implemented the new algorithms to identify unknown conductivities. The Electrical Impedance Tomography (EIT) is a technique of imaging the distribution of conductivity inside the tested object from measurements of the distribution of potentials on the object's surface. The EIT deals with an inverse problem in which given the measured voltages on electrodes it estimates the conductivity distribution by using an image reconstruction algorithm [13-17].

2. Problem Formulation

Let the linear system be described by the state equation

$$Y = X\beta + \varepsilon, \tag{1}$$

where $Y \in R^n$, $X \in R^{n\times(k+1)}$ denote the observation matrices of response and input variables respectively, $\beta \in R^{k+1}$ denotes the vector of unknown parameters. When the linear model (1) contains the intercept, then the first column of matrix X is a column of ones. The object $\varepsilon \in R^n$ in linear system (1) presents a sequence of disturbances, which is usually defined as a vector of independent identically distributed random variables with normal distribution $N\left(\tilde{0},\sigma^2 I\right)$, which, $\tilde{0} \in R^n$ is a zeros vector but $I \in R^{n\times n}$ is an identity matrix. The classical Least Square method depends on identification of unknown parameters $\beta = \left(\beta_0, \beta_1, ..., \beta_k\right)$ in (1) by solution the task

$$\min_{\beta\in R^{k+1}} \left\| Y - X\beta \right\|^2 \tag{2}$$

[1] Lublin University of Technology, Nadbystrzycka 38, 20-618 Lublin, POLAND

[2)] Research and Development Center, Netrix S.A., Związkowa 26, 20-148 Lublin, POLAND

[3)] University of Economics and Innovation in Lublin, Projektowa 4, 20-209 Lublin, POLAND

If $\det(X^T X) \neq 0$, then the best unbiased linear estimator of unknown parameters β is

$$\hat{\beta} = \left(X^T X\right)^{-1} XY \tag{3}$$

The problem appears when the matrix $X^T X$ is singular or tends to singular.

The main aim of tomography is to make a carefully image reconstruction of object. During measurements we can see, that the measured values from some electrodes are strongly correlated. In this cases we have a multicollinearity problem. When the reading values from electrodes are correlated (collinear), then the matrix $X^T X$ tends to singular matrix. When we use directly the least square method, then we may obtain the unstable prediction based on linear model (1). The possible ways of solution of multicollinearity problem depend on application the stepwise regression (see e.g. [22]), the elastic net (see e.g. [18], [23], [3]) or decomposition of correlated predictors into orthogonal components (see e.g. [20], [21])

$$X = TP^T, \tag{4}$$

where T is a matrix of scores and P is a matrix of loadings.

3. Principal Component Regression

From the singular value decomposition (SVD) the data matrix X can be presented as

$$X = UDV^T$$

where U and V are the unitary matrices, D is a diagonal matrix with non-negative real numbers on the diagonal. The matrix of scores is defined as the matrix of left singular vectors of X multiplied with diagonal matrix of singular values $T = UD$, but the matrix of loadings is a matrix of the right singular vectors and $P = V$. In PCR we use the scores to explain the linear regression (1), because the scores are orthogonal, so there are no multicollinearity problem. Analyzing the Root Mean Square Error of prediction we usually approximate the data matrix X by a - number of principal components

$$X \approx \left(U_{(a)} D_{(a)}\right) V_{(a)}^T = T_{(a)} {P_{(a)}}^T. \tag{5}$$

Next we regress the response variable Y on the scores $T_{(a)}$ instead on X. From (3) and unitarily the unknown parameters in linear model (1) are estimated as

$$\hat{\beta} = P_{(a)} \left({T_{(a)}}^T T_{(a)}\right)^{-1} {T_{(a)}}^T Y = P_{(a)} \left(D_{(a)} {U_{(a)}}^T U_{(a)} D_{(a)}\right)^{-1} D_{(a)} {U_{(a)}}^T Y = V_{(a)} {D_{(a)}}^{-1} {U_{(a)}}^T Y \tag{6}$$

In practice PCR is performed on the mean centered data matrix X, see e.g. [20], [21] and the intercept is calculated afterwards. The prediction based on model (1) is

$$\hat{Y} = X\hat{\beta}, \tag{7}$$

where $\hat{\beta}$ is a vector of estimators of unknown parameters β is given by formula (6).

4. Partial Least Square Regression

Just like PCR, the Partial Least Square Regression defines the latent variables to explain the relations in linear model (1). In PCR the principal components are chosen to explain the variance of data matrix X and come from singular value decomposition. In PLSR the components (called the latent variables) are chosen to capture as much as possible the covariance between X and Y. Like PCR, PLSR is performed on the mean centered data matrix X and the intercept is calculated afterwards. The latent variables are calculated

iteratively. More about PLSR the readers can be found in e.g. [20], [21], [3]. The PLSR algorithm will be presented below.

1. First we put $E_1 = X$, $F_1 = Y$ and $i = 1$.
2. Let w_i is a first left singular vector of SVD of the crossproduct matrix $E_i^T F_i$
3. Determine the normalized score

$$t_i = \tilde{t}_i \Big/ \sqrt{\tilde{t}_i^T \tilde{t}_i}$$

where $\tilde{t}_i = E_i w_i$.

4. The i-th loadings of X and Y are obtained by regressing E_i and F_i against the score t_i

$$p_i = E_i^T t_i \text{ and } q_i = F_i^T t_i .$$

5. The information related to the latent variable based on score t_i is subtracted from data matrices E_i and F_i

$$E_{i+1} = E_i - t_i p_i^T \text{ and } F_{i+1} = F_i - t_i q_i^T .$$

6. Put $i = i + 1$ and go to step 2.

After every iteration the vectors w_i, t_i, p_i, q_i are saved as a i-th columns in matrices W, T, P and Q respectively. The number of latent variables is determined by analyzing the root mean square error of prediction. The matrix

$$R = W\left(P^T W\right)^{-1} \tag{8}$$

is generalized inverse of P^T and the matrix of X scores satisfies the equation $T = XR$. From above we have $X = TP^T$ and $Y = TQ^T$. Thus we have this same situation as in PCR. Next we regress the response variable Y on the scores T instead on X and consider the equation

$$Y = X\beta + \varepsilon = T\left(P^T \beta\right) + \varepsilon .$$

From (3) we have

$$P^T \hat{\beta} = \left(T^T T\right)^{-1} T^T Y .$$

Finally from (8) we obtain the estimators of unknown parameters

$$\hat{\beta} = R\left(T^T T\right)^{-1} T^T Y = R\left(T^T T\right)^{-1} T^T T Q^T = RQ^T . \tag{9}$$

5. Numerical example

The measuring system has 16 electrodes, so we make 208 measurements (in the linear model we have 208 predictors). Reconstruction looks like a 64 * 32 = 2048 image (for each pixel we construct a linear model with 208 correlated predictors).

Identification of XYR dependencies from voltages.

Measurement data was grouped to determine linear relationships between dependent variables X and independent. For the analysis of polynomial regression, there was used the classical method of the smallest squares. Unfortunately, many of these parameters were statistically insignificantly different from zero (they added more noise to the model than explained the

dependencies, which was due to a large correlation between some of the independent variables. Thus, for each model, explanatory variables have been selected that are strongly correlated with the explanatory variable and at the same time are poorly correlated to one another. Figure 1 shows the example reconstruction of 1 object.

In the example in Figure 2, the number of principal components is 30. An attempt to reconstruct an image when there are 4 objects in the model being tested (model parameters for each pixel are estimated for up to 3 objects only, and we search 4 objects).

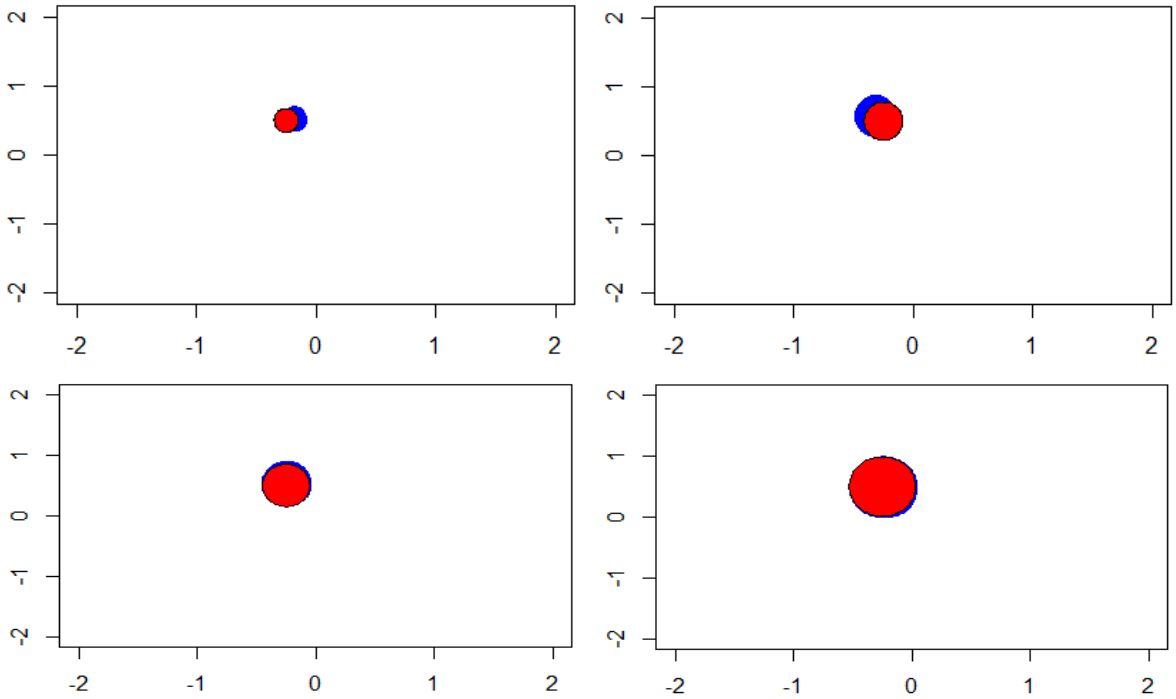
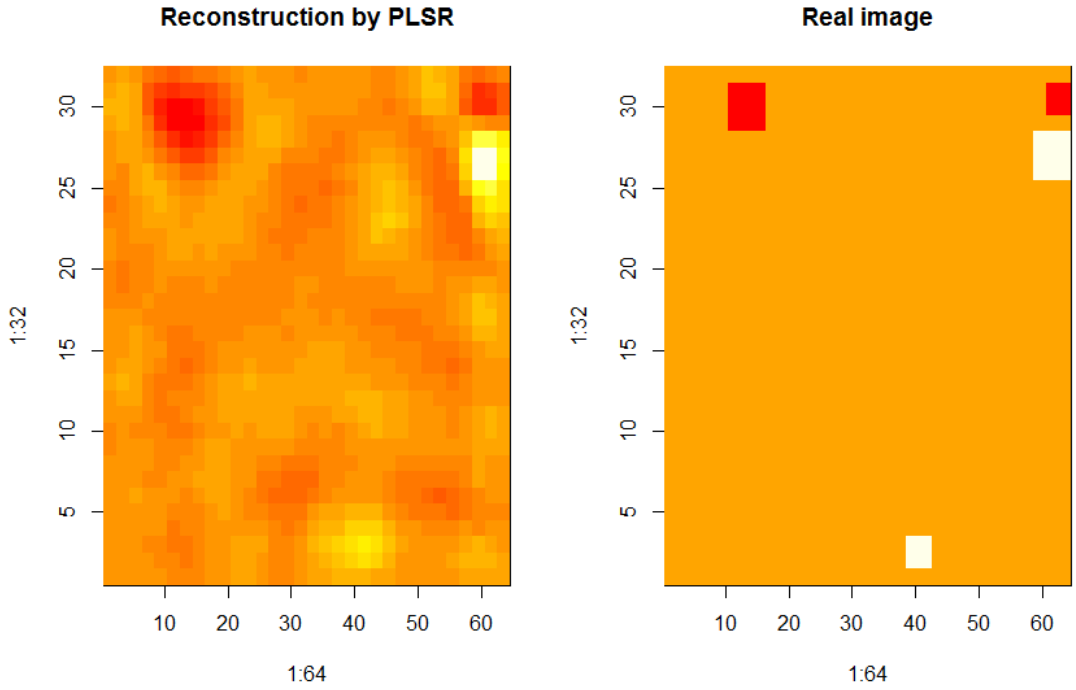


Fig. 1: The example reconstruction of 1 object a) radius=0.1, b) radius=0.16, c) radius=0.2, d) radius=0.28



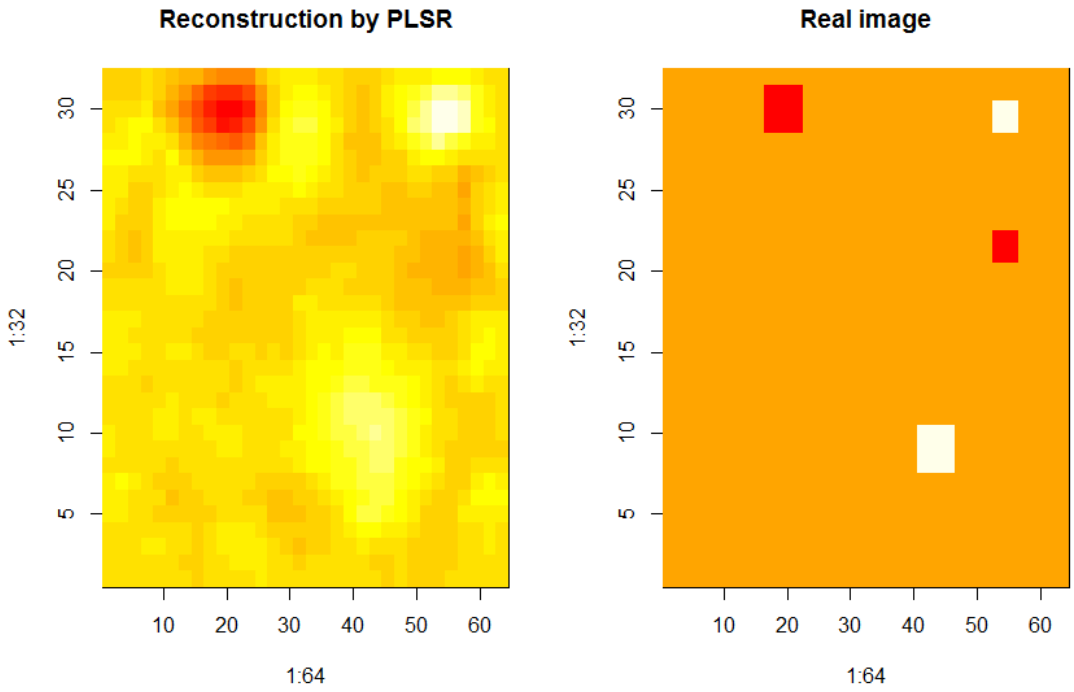


Fig. 2: The example reconstruction of 4 objects.

6. Conclusion

In this paper, there was presented the solution for the identification unknown object. Numerical methods were based on the Partial Least Square Regression. In Principal Component Regression, the latent variables are chosen to explain as much as possible the variance of predictors. An efficient algorithm for solving the inverse problems would also improve a lot of the numerical performances of the proposed method. The process of reconstruction is enough good, because the region borders are located nearly the searched object edges. The presented method has been applied very successfully in many areas of the scientific modelling.

References

1. Banasiak R, Wajman R, Sankowski D, Soleimani M., *Three-Dimensional Nonlinear Inversion of Electrical Capacitance Tomography Data Using a Complete Sensor Model*. Progress In Electromagnetics Research (PIER) vol. 100, pp. 219-234, 2010.
2. Filipowicz S.F., Rymarczyk T., *The Shape Reconstruction of Unknown Objects for Inverse Problems*, Przeglad Elektrotechniczny, vol. 88, Issue 3A, pp. 55-57, 2012.
3. Friedman J., Tibshirani R., Hastie T., *Regularization paths for generalized linear models via coordinate descent*, Journal of Statistical Software, Vol. 33, nr 1,2010.
4. Kłosowski G., Gola A., *Risk-based estimation of manufacturing order costs with artificial intelligence* [in:] Ganzha M., Maciaszek L., Paprzycki M. (eds.), Proceedings of the 2016 Federated Conference on Computer Science and Information Systems (FEDCSIS), IEEE, pp. 729-732, 2016, DOI: 10.15439/2016F323.
5. Kłosowski G., Gola A., Świć A., *Application of Fuzzy Logic Controller for Machine Load Balancing in Discrete Manufacturing System*, Lecture Notes in Computer Science, vol. 9375, pp. 256-263, 2016, DOI: 10.1007/978-3-319-24834-9_31
6. Kryszyn J., Smolik W., *2D modelling of a sensor for electrical capacitance tomography in ECTSIM toolbox* , Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 1, pp. 146-149, 2017.

7. Kryszyn J., Smolik W., Radzik B., Olszewski T., Szabatin R., *Switchless charge-discharge circuit for electrical capacitance tomography*, Measurement Science and Technology, vol. 25, Issue 11, pp. 1-9, 2014.
8. Majchrowicz M., Kapusta P., Jackowska-Strumiłło L., Sankowski D., *Acceleration of image reconstruction proces in electrical capacitance tomography 3D in heterogeneous, multi-GPU system*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 7, Issue 1, pp. 37-41, 2017. DOI: 10.5604/01.3001.0010.4579
9. Martens H., *Reliable and relevant modellingof real world data: A personal account of the development of PLS regression, Chemometrics and intelligent laboratory systems*, Vol. 58, nr 2, pp. 85-95, 2001.
10. Polakowski, K., Filipowicz, S.F., Sikora, J., Rymarczyk, *Tomography technology application*, Przeglad Elektrotechniczny, vol. 84, Issue 12, pp. 227-229, 2008.
11. Rybak G., Chaniecki Z., Grudzień K., Romanowski A., Sankowski D., *Non–invasive methods of industrial process control*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 4, Issue 3, pp. 41–45, 2014.
12. Rymarczyk T., *New Methods to Determine Moisture Areas by Electrical Impedance Tomography*, International Journal of Applied Electromagnetics and Mechanics, vol. 37, Issue 1-2, pp.79–87, 2016.
13. Rymarczyk T., Adamkiewicz P., Duda K., Szumowski J., Sikora J., *New Electrical Tomographic Method to Determine Dampness in Historical Buildings*, Archives of Electrical Engineering, vol. 65, Issue 2, pp. 273-283, 2016.
14. Rymarczyk T., Tchórzewski P., Adamkiewicz P., Duda K., Szumowski J., Sikora J., *Practical Implementation of Electrical Tomography in a Distributed System to Examine the Condition of Objects*, IEEE Sensors Journal, vol. 17, Issue:24, 2017, pp. 8166-8186, DOI: 10.1109/JSEN.2017.2746748.
15. Smolik W., Kryszyn J., *Linea rover ranges iterative algorithms for image reconstruction in electrical capacitance tomoraphy*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 1, pp. 115-120, 2017.
16. Smolik W., Kryszyn J., Radzik B., Stosio M., Wróblewski P., Wanta D., Dańko L., Olszewski T., Szabatin R., *Single-shot high-voltage circuit for electrical capacitance tomography*, Measurement Science and Technology, vol. 28, Issue 2, p. 025902, 2017.
17. Tchorzewski P., Rymarczyk T., Sikora J., *Using Topological Algorithms to Solve Inverse Problem in Electrical Impedance Tomography*, International Interdisciplinary Phd Workshop 2016, pp. 46-50, 2016.
18. Tibshirani R., *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society, Series B, Vol. 58, nr. 1, pp. 267–288, 1996.
19. Wajman R., Fiderek P., Fidos H., Jaworski T., Nowakowski J., Sankowski D., Banasiak R., *Metrological evaluation of a 3D electrical capacitance tomography measurement system for two-phase flow fraction determination,* Measurement Science and Technology, vol. 24, 065302, 2013.
20. Wehrens R., Mevik B.H., *The pls package: Principal Component and Partial Least Square Regressian in R, Journal of Statistical Software*, Vol. 18, nr 2, 2007.
21. Wehrens R., *Chemometrics with R. Multivariate Data Analysis in the Natural Science and Life Sciences*, Springer 2011.
22. Xin Yan, Xiao Gang Su, *Linear regression analysis: Theory and computing*, World Sciencific, 2009.
23. Zou H., Hastie T., *Regularization and variable selection via the elastic net*, Journal of the Royal Statistical Society, Series B, Vol. 67, nr 2, pp. 301–320, 2005.

Konrad KANIA[1], Tomasz RYMARCZYK [12]

Quality Control in Production Process - Algorithms for Image Analysis

Modern production processes are increasingly complicated while customers require higher quality products at the lowest prices. This situation increases the importance of process optimization. Article contains description of the quality control system based on optical detection algorithms.

1. Introduction

Modern production processes are increasingly complex, while customers are demanding higher quality products at the lowest price. This situation increases the importance of process optimization. Products and requirements are changing faster and faster, technologists have less and less time to learn the process and to optimize it solely on the basis of their own experience. On the other hand, very often the processes are automated and metered in detail, so we have plenty of data describing it. In such a situation, we cannot give up the chance of decision support. Analysis of the data may be used for process optimization in different ways and it can affect many aspects of process management. We can detect disturbances in the process, find the causes affecting the problems with quality, and choose the optimal settings for the process, comparing different preparation procedures and many others. It is worth noting that in the processes, there are many people involved in with different tasks and permissions. Therefore, data analysis tool should provide access control. The research project assumes creation of two measurement platforms: a set of multiphase flow system and a mini production line. In the first case, we will be analysing two-phase flows of liquid (water) and air [4,8,9,11]. That type of flow is commonly used in chemical reactors where air is mixing substances. In the second case, we will be analysing moving objects on the mini production line using image detection techniques. The system will be also using biometric reader's witches will provide access control. [1,3,7].

2. Line production

The main part of the measuring system is closed (oval) leaf production line of variable speed travel.

Fig. 1. Sketch of the mini production line system

[1)] Research and Development Center, Netrix S.A., Związkowa 26, 20-148 Lublin, POLAND

[2)] University of Economics and Innovation in Lublin, Projektowa 4, 20-209 Lublin, POLAND

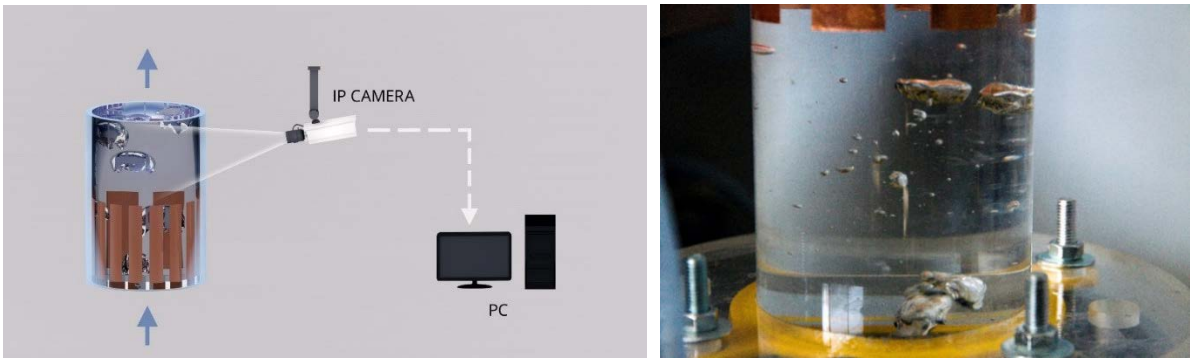


Fig. 2. Schematic of the measurement system to study the multiphase flows.

Optical detection system as in the previous system is based on the IP cameras (Fig.1 and 2). The program will be analysing images from the cameras in real time. The analysed objects will recognize by their characteristic features such as: shape, size, colour, etc. At the same time, the image will be saved on the disk array. Such a system will provide not only accurate modelling in micro scale production facility, but will also be a base for the development of fast vision analysis algorithms.

3. Image segmentation

The purpose of segmentation is to separate the objects of the image to be recognized. The first step is usually to scale the image according to the accepted reference values. This can be achieved in a variety of ways, such as histogram stretching, histogram shift, or rescaling with one or two reference values. The ability to effectively apply a particular scaling method depends on the properties of scaled objects (for example, a large variety of colour objects of the same type). Developed algorithms in the field of imaging by skilfully separating individual objects and their fragments. Separating parts of an image with common features would allow you to more precisely define boundaries between objects and detect anomalies within them. To analyse the image, you will move from pixel to object level. The proposed solutions will segment the image objects obtained by the indicated techniques and isolate the sought shapes in these images [5,6,10]. Algorithms for segmentation multiphase flows, analysis and image reconstruction were presented in Figure 3 and 4

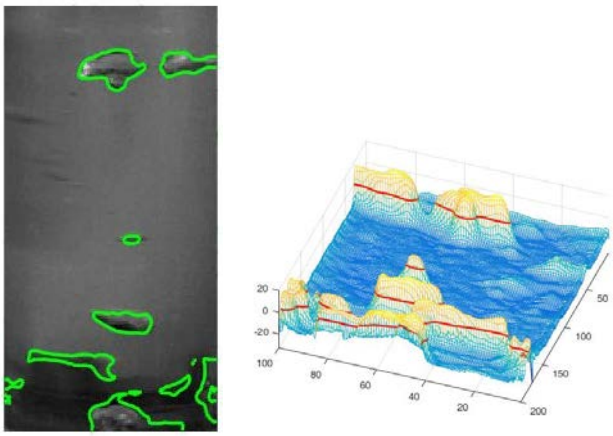


Fig. 3. Image segmentation of gas in the water by the level set method.

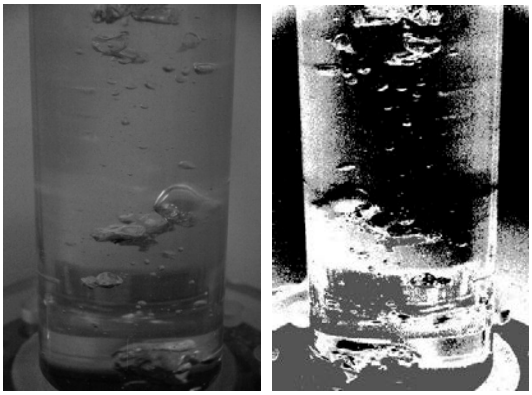


Fig. 4. Genetic algorithm: 5 classes, populations 8000, a filter for noise removal pepper and salt.

4. Object detection

This chapter is a summary of a recent work focused on processing images from industrial cameras in order to detect an object on a conveyor belt and examine its quality parameters. As testing objects bottle nuts were used. First, we introduce methods of separation and recognition of a type of nuts based on colour thresholding. Next, we present an algorithm of examination whether a nut is on a bottle or not. The last part of the paper is about methods based on the code book algorithm to separate multiple objects in one picture.

Determination of type of a bottle nut

The first application of image processing is about locating a nut on a conveyor belt and determining its type by the colour qualification. In the testing setting there is a camera located over the conveyor belt. Axis of the camera view is perpendicular to the plane of the belt. There are three types of nuts from bottles of sparkling water in three different levels of carbonation encoded in different colour of nuts. The algorithm uses several types of colour thresholding. First of them extracts pixels of a nut from pixels of a background (the belt). Then extracted pixels are counted to determine size of the object to decide whether there is a nut in the picture. If the nut is found, next steps of the thresholding are performed to determine the type of the nut. Each value of the thresholding is set according to the result of scripts which analyse pictures of the empty conveyor belt and samples of pictures of colours which have to be distinguished from background and from each other. The technique of the threshold setting is based on the analysis of channels of given colours in the colour space. We determine a middle point between distinguished colours (treated as a set in the three-dimensional colour space) separately for each channel. As the final threshold we take the value of the middle point in one channel, where the selected channel is related to the biggest distance between compared colours.

There are two ways to define a middle point, first is the middle between centres of sets in the colour space corresponding to given colours. Second approach uses a middle point between nearest points from two different sets. Additionally the script checks, as a necessary condition of separation of colours, whether sets do not intersect. As a descriptor of separation quality we use distances between centres of colour sets expressed in multiples of standard deviations evaluated for a given colour. To improve the quality of separation we consider RGB colour space as well as the HSV system which is more resistant to unfavourable lighting conditions. Next improvement of the separation goes from a quantile-based filtering of initially segmented pixels of a nut. In presented algorithm to determine the type of the nut we consider values between 15% and 85 % quantiles. The quantiles filtering is often necessary to remove intersections between distinguished colours in the colour space (Fig. 5).

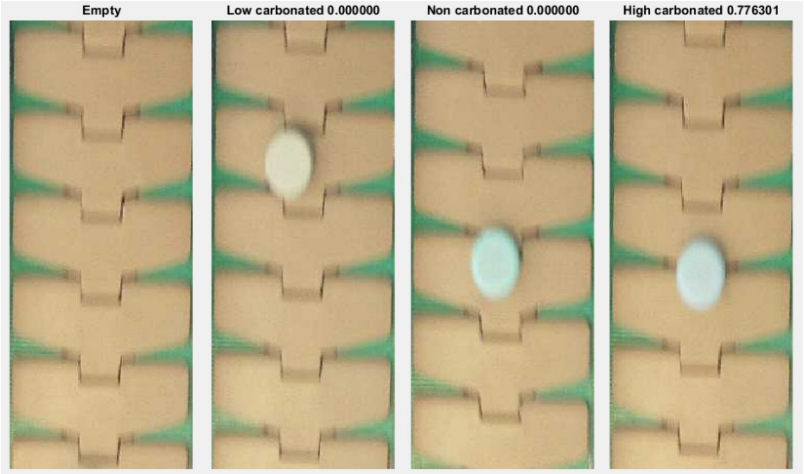


Fig. 5. Example of results of the algorithm of determination of type of bottle nut.

Examination of appearance of nut on a bottle

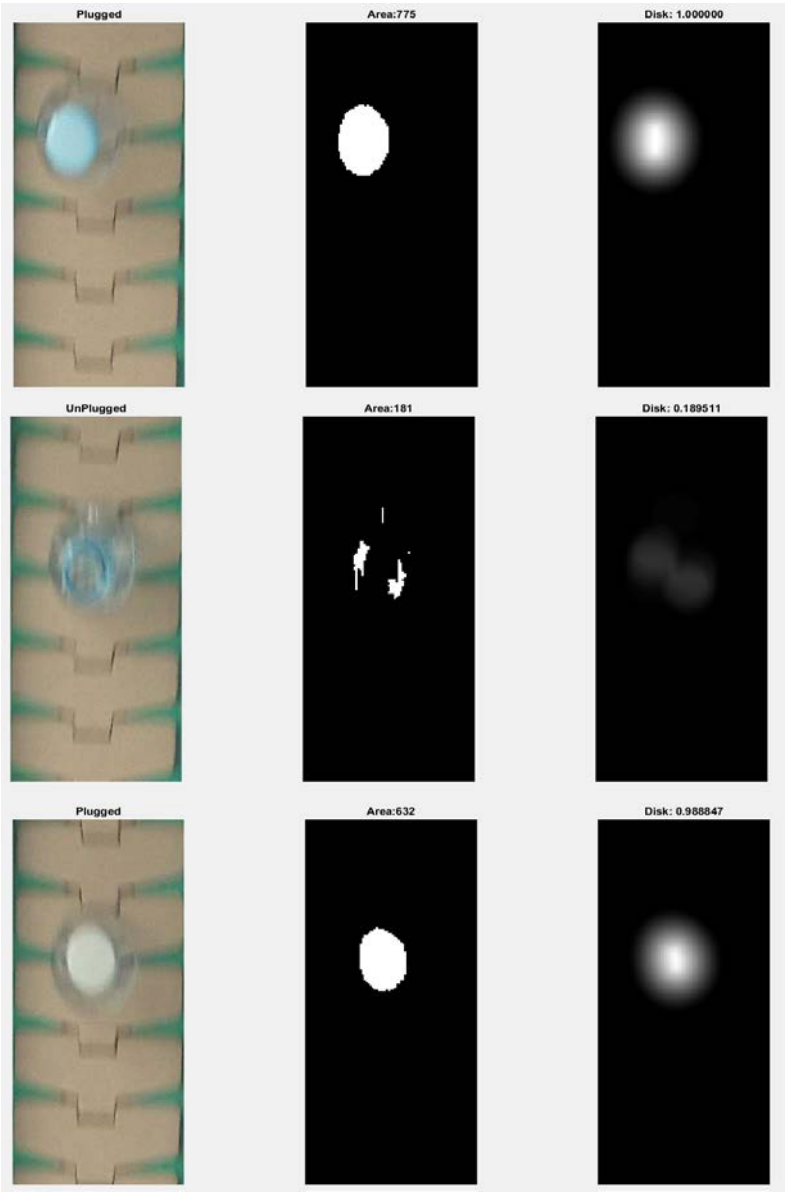


Fig. 6 Example of the result from application of determination presence of nut on the bottle.

Second application solves the problem of examination whether a bottle is plugged or not. Testing setting is the same as in the previous problem. To determine presence of the bottle in a picture we use colour thresholding of backgrounded conveyor belt from first application. Next, we resize whole image to standardize sizes of found objects. If an object is found in picture, then two-dimensional convolution of binary mask of the picture is performed to evaluate similarity of object on mask to circular shape of the nut. The range of numerical descriptor of similarity is interval [0,1] where descriptor is equal to 1 if in binary mask there is an object which can fully contain a circle of radius at least as big as the radius of circular referential filter used in convolution. To reduce noises in the picture and improve separation of background we use Gaussian filtering (Fig.6).

Code Book based object segmentation algorithm

To perform properly previous algorithms require that there is at most one object in a frame. However, for most of industrial applications, situation of multiple objects in the frame has to be considered due to optimisation of space and speed of production. Because of that, there is a necessity of using some general methods of separation objects in the frame. Basing on the Code Book foreground-background segmentation algorithm presented originally in [2] we implemented fast method of recognition object on a moving conveyor belt. Simple frame-differentiation approach to remove background is not sufficient because of movement of the conveyor belt in the background. On the other side, regardless of the current position of the belt in the frame, the background contains the same set of colours. Due to that specification of background code book approach is found to be extremely sufficient. The only remaining problem of that algorithm is its computational complexity thus our work has focused on improving its speed. First of all, we divided the whole frame for blocks of size 12 by 12 pixels, then for each block we calculated a weighted average (separately for each channel of the frame) and that values are compared with the corresponded values in the code book, which is previously constructed with the same block-divided manner based on a series of pictures of an empty conveyor belt. Other relevant fact is that the created code book was ordered properly according to advice in [2]. Next important improvement was changing colour space from RGB to HSV, which allow us to apply a simpler way of defining a distance between compared colours. Instead of Eucliden distance we use an absolute value of differences in particular channels which is much more efficient.

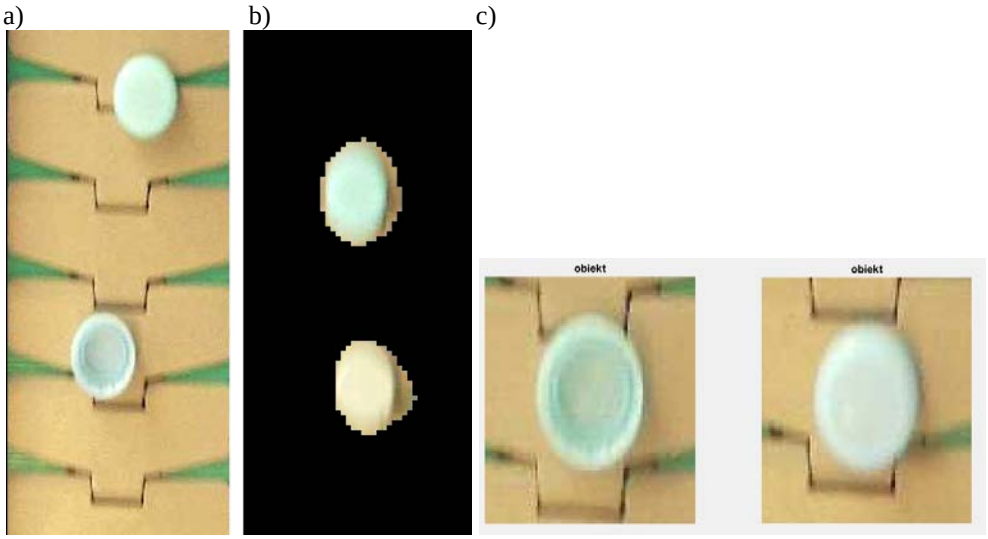


Fig. 7. Example of Code Book algorithms. Chessboard-like masked images before post-processing (a). Two-step method of increasing the resolution of segmentation (b). Code Book based multiple object separation (c).

The last improvement is an implementation of segmentation only to a half of all blocks of the frame spread out in the photo in a way similar to the arrangement of fields on a chessboard. That technique not only reduces the number of blocks to check by half but also reduces a part of the noises in processed image. After comparison an image with the code book, a full mask is obtained by the morphological post-processing to remove noises and also to merge detected blocks of the foreground into connected regions. After first using of the algorithm we can perform it once again, but with smaller blocks (4 by 4 pixels) and only to regions found in the first attempt. This approach allows us to increase final resolution of segmentation with average speed of processing about ten times faster (tens of milliseconds) compared to the one-step method with size of blocks set to 4 by 4 pixels (Fig.7).

5. Conclusion

Nowadays, optical detection plays an increasingly important in the production process. The development of new systems based on the technology of optical detection methods to a large degree can improve the production process at different stages. The challenge is to create a single optimal and efficient system ensures easy application, which can cooperate with the automation line production. It is also important to develop optimal algorithms for the image analyse and the defect detection.

References

1. Blecker T., Kersten W., Lüthje C., *Innovative Process Optimization Methods in Logistics: Emerging Trends, Concepts and Technologies*, Erich Schmidt Verlag GmbH & Co KG, 2010.
2. Kim K., Chalidabhongse T. H., Harwood D., Davis L., *Real-time Foreground-Background Segmentation using Codebook Model*, Real-time Imaging, Volume 11, Issue 3, pp. 167-256, 2005.
3. Musharavati F., *Process Planning Optimization in Reconfigurable Manufacturing Systems*, Universal-Publishers, 2010.
4. Romanowski A., Grudzień K., *Zestawienie metod rekonstrukcji obrazów i modelowania przestrzennego 2D dla tomografii procesowej*, Automatyka (półrocznik AGH), t. 9, z. 3, 621-630, 2005.
5. Rymarczyk T., *Characterization of the shape of unknown objects by inverse numerical methods*, Przeglad Elektrotechniczny, vol. 88, Issue 7B, pp. 138-140, 2012.
6. Rymarczyk T., Adamkiewicz P., Olchowy D. et al., Segmentation and Analysis of Medical Images by e-Medicus System, International Interdisciplinary Phd Workshop 2016, pp. 125-129, 2016.
7. Saravanan R., *Manufacturing Optimization through Intelligent Techniques*, CRC Press, 2006.
8. Smolik W., Kryszyn J., Radzik B., Stosio M., Wróblewski P., Wanta D., Dańko L., Olszewski T., Szabatin R., Single-shot high-voltage circuit for electrical capacitance tomography, Measurement Science and Technology, vol. 28, Issue 2, p. 025902, 2017.
9. Szumowski J., Adamkiewicz P., Rymarczyk T., Electrical Capacitance Tomography to Optimize Quality Control System, International Interdisciplinary Phd Workshop, pp. 121-124, 2016.
10. Tchorzewski P., Rymarczyk T., Sikora J., Using Topological Algorithms to Solve Inverse Problem in Electrical Impedance Tomography, International Interdisciplinary Phd Workshop 2016, pp. 46-50, 2016.
11. Wajman, R., Banasiak, R., Tunnel-based method of sensitivity matrix calculation for 3D-ECT imaging. Sensor Review, vol. 34, pp. 273–283, 2014. doi:10.1108/SR-06-2013-692.

Anita SOWA-WATRAK[1], Mariusz RZĄSA[2]

Algorithms Of Image Analysis For Diagnosis Of Welded Connections

The following paper presents developed by the authors the algorithm of image analysis supporting the process of diagnosing welded connections. Attention has been paid to the main disadvantages that arise in welded connections and to the basic difficulties when it comes to diagnosing the previously mentioned defects. Moreover, the method of computer image analysis based on the threshold of color saturation and a modification of the black-and-white image method in the RGB raw palette has also been presented. The paper reveals results of sample analyzes, focusing on the values of the image analysis method applied. This solution extends the possibilities of imaging the material defects detected, which improves the classification of welded connections.

1. Introduction

Welding is one of the basic methods of joining metals. It has been in use for many years, and although welding technology is still undergoing modernization, there is no technology to guarantee 100% repeatability of the welds. It is still necessary to carry out checks on welded connections made with the non-destructive method [6][8].

The welding process consists of combining the materials by initial heating and then melting their place of joining. In this place is formed alloy of two joined metals, which we call a shed. Welding can take place either with or without a binder. Welding is divided by the way the temperature is produced for: gas and electric thermite [2].

These welding methods does not guarantee sufficient repeatability. This is due to the presence of many factors affecting this process. The effect of high temperature causes material changes impacting the strength parameters of the material. In addition, the welded connection is not uniform throughout the cross-section. What is more, its geometry changes over its length. These factors make it necessary to control and diagnose welded joints [2].

Diagnosis of welded connection are based on the detection of defects and their impact on the strength of the weld connection. Verification of defects takes into account the type, size, severity and its position. What also should be mentioned here is the fact that the most common defects in welded connections include: no milling, tubular bubbles, gas bubbles and ridge leakage.

The most common methods for diagnosing welds include ultrasound or x-ray scans. Ultrasonic methods are cheaper, but not always let to fully assess the previously mentioned defect. Additionally, they are used for detecting defects such as inclusions, cracks of any kind, and no milling. Radiographic studies, in turn, allow for the detection and estimation of incompatibilities in a wide range. It is possible to recognize gas blisters, slag inclusions,

[1] Ph.D., student, University of Technology in Opole, Prószkowska 76, 45 758 Opole, POLAND,
tel. +(48) 697 930 849, e-mail: ani.sowa@doktorant.po.edu.pl
[2] Ph.D., University of Technology in Opole, Prószkowska 76, 45 758 Opole, POLAND,
tel. +(48 77) 449 80 71 , e-mail: m.rzasa@po.opole.pl

various spills, floods, misfolding and face inequalities [1][3]. As the process of analyzing x-ray images is time-consuming, a computer algorithm for diagnosing welds is proposed.

2. The idea of radiographic tests

To start with, radiographic studies consist of overexposing of the x-ray beam weld and then recording on the photographic film the degree of absorption of the radiation by the x-rayed material. The attenuation of the radiation beam is described by the Beer absorption equation:

$$I = I_0 \cdot e^{-\mu x} \tag{1}$$

Where: I - the intensity of the bonus X after passing through the joint,

I_0 - premium rate X issued from source,

μ - linear absorption coefficient,

x - the path traveled through the radiation beam [4].

The degree of exposure of the photographic film depends on the type of material to be examined, radiation intensity and exposure time. All types of heterogeneity in the weld are characterized by different absorption coefficients, which in turn causes different degrees of illumination of the phonographic plate. As a consequence, darker areas are observed where the absorption coefficients of the material are smaller [7].

The most commonly used are x-ray tubes with constant and unregulated intensity of radiation. The correct contrast on the radiogram is obtained by selecting the appropriate exposure time. This often requires several sequences of pictures. Computer-assisted image analysis can partially reduce the number of necessary sequences. This is particularly useful in the diagnosis of frontal and inguinal welds, where there are different thicknesses of material in the scanning plane, and choosing the appropriate exposure time is not easy.

The paper presents the method of computer aided diagnostics of welded connections. The welded pictures displayed in the work of the RTG have been made with a Scanray 250 x-ray tube powered by 250 kV, for 4 sec in the authentic process of welded connections diagnostics by employee of company.

3. The idea of computer method of image analysis assistance

To start with, the idea of the method is to set an appropriate saturation threshold for each pixel of the image. Assigning pixels to groups above and below the threshold causes the image to become a one-bit image (uninfocused). In the initial period of research, the authors of this article applied this method to x-ray grayscale images. In order to ensure better visualization of the defects found, contours of the image after the binarization have been extracted by the authors. This solution makes it possible to distinguish material defects caused by welded connections. Computer image analysis is designed to facilitate the decision to allow the weld to be commissioned by the welded joint evaluator.

Figure 1 shows the results for an exemplary welded connection. Moreover, Figure 1a shows an example x-ray image of a weld. After the binearization process, the image shown in Figure 1b was obtained. Once the edges have been isolated (Fig.1c), any discontinuities and material defects can be easily and quickly identified.

Because x-ray imaging is done in black and white, it is natural to get a grayscale image for which a suitable binarization threshold is set. As a result of this process, luminance below

the threshold is set to "0" and pixels with luminance above threshold "1". In this way, the image is one-bit.

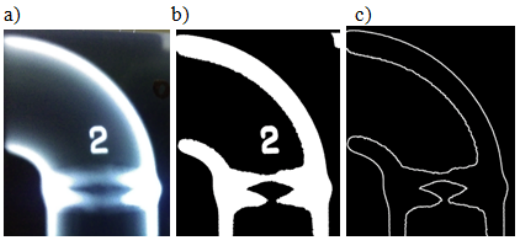


Figure. 1: The idea of the image analysis method: a) x-ray image of the weld, b) Image of the binary image, c) Image after the separation of the contours.

Extracting the contours greatly facilitates a visual assessment of the quality of the welded connection. (Fig.1c). It is then possible to quickly detect discontinuities and inclusions of foreign materials in the weld including the presence of gas bubbles. Thanks to this analysis it is possible not only to discover defects but also to measure and express them in numerical form.

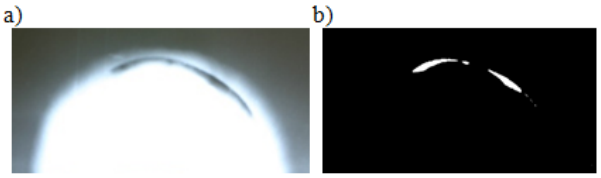


Figure. 2: x-ray photo defects in the weld: a) x-ray image, b) extracted defect.

For example, Figure 2 shows a picture of a weld with a distinct defect. Visible defect can be isolated and presented in a separate image (Fig.2b). In this image we can measure the dimensions of this defect and calculate its surface area. Based on these parameters, it is possible to diagnose more accurately the weld joint and to decide whether to allow the weld to operate.

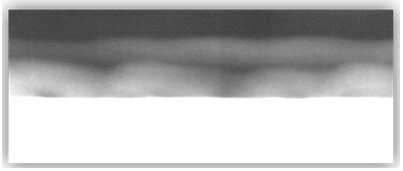


Figure. 3: x-ray photo of the thermal impact zone.

This method does not allow for proper detection of all parameters of the welded joint. An example may be the determination of a shredding zone or material change zone as a result of the heat generated during the welding process. For example, Figure 3 shows an image of a thermal zone. The visible area of the thermal effect is clearly visible, which has caused changes in the structure of the welded material. These changes are characterized by a different x-ray absorption coefficient.

Figure. 4: Thermal zone of boundary zone.

Although grayscale is clearly visible in the grayscale image, the application of grayscale thresholds only allows the boundary of its effect to be determined (Fig.4). This can sometimes give some insufficiency, as information about the unevenness of change within the zone is lost.

This is because, as a result of the threshold process, only one threshold value for grayscale saturation is taken into account. In addition, the image field which often sees the human eye as gray in fact consists of a mosaic of pixels of different saturation. This is an additional difficulty in setting the right grayscale threshold. For this reason, the authors proposed some modifications to the method. This modification aims to introduce additional values describing the pixel in order to enable more accurate image analysis. The previously mentioned solution will allow to widen the scope of welded connection analysis.

4. Application of RGB filter

The solution described in Chapter 3 does not allow for distinguishing the characteristic zones that occur in welded joints. These include the melting zone and thermal changes. Due to the strength of the weld, it is important to know the thickness and continuity of these zones. For this reason, the authors have developed their own method to widen the scope and possibilities of diagnosed welds.

In industrial conditions, x-ray devices are used to produce black and white images of the illuminated welds. At present there are opportunities to obtain color x-ray images, however, to diagnose welds of this kind still have no economic justification. The most commonly used weld detectors are black-and-white images on photographic film. The proposed method is to scan the image from the film and save it in a RGB color image format. The scanned image continues to be a black and white image, but each pixel of the image is described using three RGB values.

RGB palette is a palette consisting of three basic colors (red, green and blue) with which any color of the pixel is displayed in the image. This is because any gray tone can be expressed using three basic RGB colors. The YUV model determines the saturation intensity of basic RGB colors to achieve the appropriate grayscale value. Bearing in mind that the human eye is more sensitive to green, and the least blue to grayscale describes the relationship:

$$GS = 0{,}299 \cdot R + 0{,}587 \cdot G + 0{,}114 \cdot B \tag{2}$$

Where: GS - grayscale value,

R, G, B - RGB color saturation values [5].

Having an RGB image means that each pixel of the image in shades of gray is determined by three numbers. This solution enhances the scope and capabilities of computer image

analysis, enabling the separation of material change zones in the weld environment and the assessment of the zone of thermal changes in the weld environment.

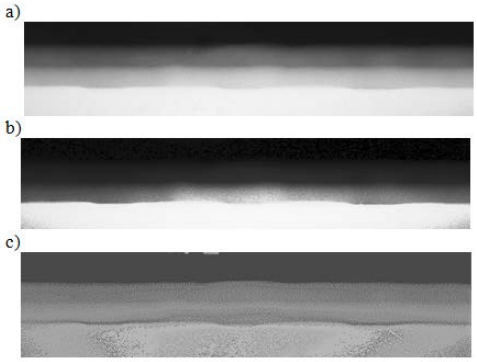


Figure. 5: Two-layer weld: a) Weld x-ray image, b) Image after filtering R and B colors, c) Image after filtering G color.

Figure 5 shows the effect of filtration using selected RGB palette colors. Figure 5a presents an image of a two-layered grayscale weld in the RGB color format. If the image was only recorded on a shade scale, it would only be possible to properly separate the edges of the first top layer. Application of blue and red color filtration caused the second layer to become less visible, thus enabling the middle zone to be correctly defined (Fig.5b). In turn, green filtration resulted in a decrease in the difference between the middle layers (Fig.5c), thus allowing the last lower zone.

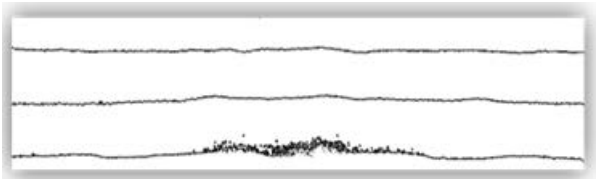


Figure. 6: Layers separated by selective RGB filtering.

Figure 6 shows the result of separating the boundaries of individual layers shown in Figure 5. Obtaining this image makes it easy and simple to measure the thicknesses of the individual layers. In addition, it is possible to automate the calculation of layer thicknesses, which significantly improves the classification of welded joints.

Figure. 7: Lay Picture of the thermal impact zone for the image in Fig. 3.

The advantages of the proposed method are well illustrated in Figure 7. It depicts an image of a thermal zone for the weld of Figure 3. In comparison to the image obtained only from grayscale analysis (Fig.4), it contains much more detail to better evaluate the effect of thermal effects on material properties.

In addition, selective RGB filtering significantly extends the scope of diagnostics of welded joints. For example, it is possible to determine the thickness of the welded connection (Fig.8). Pipe is a zone in which a weld metal alloy and weld material are formed. In order to ensure proper strength of the welded joint it is necessary to maintain not only the continuity of the melt but also its thickness [5].

Figure. 8: View of welded joints of welded joint.

5. Conclusions

The diagnostic method shown in this work streamlines the classification of welded joints. The use of selective RGB filtering method significantly widens the possibilities of computer-aided diagnostic method. The method not only allows the filtering of basic RGB colors, but also allows R-G, R-B and G-R to be filtered, which further enhances the ability of the image to detect material defects.

References

1. Czuchryj J., Papkala H., Winiowski A.: *Niezgodności w złączach spajanych*, Gliwice, 2005, pp. 15-21
2. Czuchryj J., Sławomir S.: *Niezgodności spawalnicze w złączach spawanych z metali i termoplastycznych tworzyw sztucznych*, Instytut Spawalnictwa, Gliwice, 2016, pp. 1-16
3. Czuchryj J., Walasek-Konior K., Kondoszek S.: *Problemy oceny jakości złączy próbnych wykonanych przez spawaczy podczas egzaminu kwalifikacyjnego*, Biuletyn Insytutu Spawalnictwa w Gliwicach, 2009 R. 53, no 2, pp. 59-64

4.Gamal E., Nakauchi S.: *Prediction of meat spectral patterns based on optical properties and concentrations of the major constituents*, Food Sci Nutr., March 2016, 4(2), pp. 269–283

5. Gomes J., Velho L.: *Image Processing for Computer Graphics*, Springer Science + Business Media New York, 1997, pp. 91-95
6. Lipiński T., Szabracki P.: *Diagnostyka złączy spawanych za pomocą metod NDT*, Studies & Proceedings of Polish Association for Knowledge Management, 2014, No 69, pp. 127-139
7. Senczyk D.: *Metoda pomiaru współczynnika osłabienia promieniowania rentgenowskiego przez materię*, Krajowa Konferencja Badań Radiograficznych Stary Młyn 2011, 29-31 sierpnia 2010r., http://www.badania nieniszczace.info/Badania-Nieniszczace-Nr-01-08-2011/KKBR_2011_Senczyk_%20wsp_oslabienia.df S
8. Zuev V.M., Kapustin V.I., Karpenko A.I., Van'kowa N.E.: Lipiński T., Szabracki P.: *x-ray testing and engineering diagnostics*, Russian Journal of Nondestructive Testing, March 2006, Volume 42, Issue 3, pp. 208–211

Tomasz RYMARCZYK, [12], Przemysław ADAMKIEWICZ[1], Paweł TCHÓRZEWSKI[1], Jan SIKORA[3,4]

Electrical Tomography for Innovative Imaging and ROI Monitoring Using Body Surface Potential Mapping

The paper presents the model of system idea based on electrical tomography. The system will consist of a mobile device that allows simultaneous recording of electrical potential of heart activity and ventilation of the lungs. Electrical tomography for pulmonary and heart monitoring is a non-invasive imaging method in which an unknown physical object is probed with electric currents applied on the boundary. The internal conductivity distribution is recovered from the measured boundary voltage data. The reconstruction algorithms for electrical impedance tomography were tested. New results of a reconstruction of a numerically simulated phantom are presented.

1. Introduction

Diseases of the respiratory and circulatory systems are widespread today. The development of this type of illness (both acute and chronic) is affected by many factors (including environmental pollution, stress). Early diagnosis as well as the possibility of continuous monitoring outside the ambulatory unfortunately is not possible. The main goal of the project is to create a tomographic system for 3D imaging and ROI (region of interest) monitoring using a node potential map, along with a mechanism for the interpretation of disease states [1-9]. The system will consist of a device that records the electrical potential of heart activity and lung ventilation [10-16]. The system will monitor the condition of patients and support the diagnostic process in such diseases as:

- Acute respiratory failure syndrome (ARDS)
- Chronic obstructive pulmonary disease (COPD)
- Bronchospasm
- Pneumonia
- Pulmonary hypertension
- Obstructive sleep apnea (OSA)
- Pneumothorax
- Cardiac Hemodynamics
- Aortic insufficiency
- Hypertension
- Ischemic heart disease

For the detection of multiple complex disease units, 24-hour monitoring of the vital functions of the patient is essential for normal functioning. The future of medical diagnostics is long-term patient monitoring devices - mobile devices recording a wide spectrum of diagnostics to detect pathological syndromes. The response to the needs of the medical market

[1)] Research and Development Center, Netrix S.A., Związkowa 26, 20-148 Lublin, POLAND

[2)] University of Economics and Innovation in Lublin, Projektowa 4, 20-209 Lublin, POLAND

[3)] Lublin University of Technology, Institute of Electronics and Information Technology, Nadbystrzycka 38d, 20-618, Lublin, Poland

[4)] Electrotechnical Institute, ul. Pożaryskiego 28, 04-703 Warszawa

will be a mobile system, a 3D tomographic imaging system and area monitoring using a node potential map, along with a pathology interpretation mechanism. The system will be able to monitor parameters:

- Myocardial activity
- Blood flow
- Blood pressure
- Heart rate
- Impedance
- Lung capacity
- Changes in pulmonary impedance
- Bioimpedance
- PEEP (positive end-expiratory pressure)
- ΔEELV (end-expiratory volume of lungs)
- Relative electrical permittivity
- Patient position during the test

2. Idea of Measurement System

The concept of the project is to create a mobile tomographic system for 3D imaging and field monitoring using a node potential map and a mechanism for the interpretation of disease states. The system will consist of a mobile device that allows simultaneous recording of electrical potential of heart activity and ventilation of the lungs. The system will consist of a data recording device and an aggregation and processing engine. The engine - the electronic aggregation and processing system will allow aggregation, processing and inference based on the collected data. For the purposes of medical inference - the identification of pathological syndromes and disease units on the basis of collected data will be used to reconstruct the tomographic image and developed author's algorithms. All measuring signals will be transmitted to the main module located in the central part of the belt (at the bridge height) [1,6,14,17-19]. This module will perform functions such as:

- Acquisition of EIT (electrical impedance tomography) and ECT (electrical capacitance tomography) measurement data
- Acquisition of electrocardiographic signals
- Acquisition of impedance
- Pre-processing of measurement data
- Calibration of the measuring elements (active electrodes)
- Transmission of measurement data
- Power supply of active measuring electrodes (ECT and EIT)

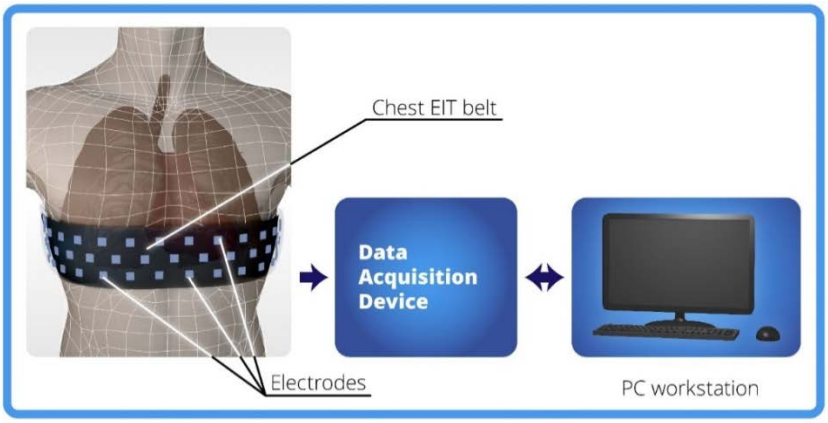


Fig. 1. Model of system idea based on electrical tomography.

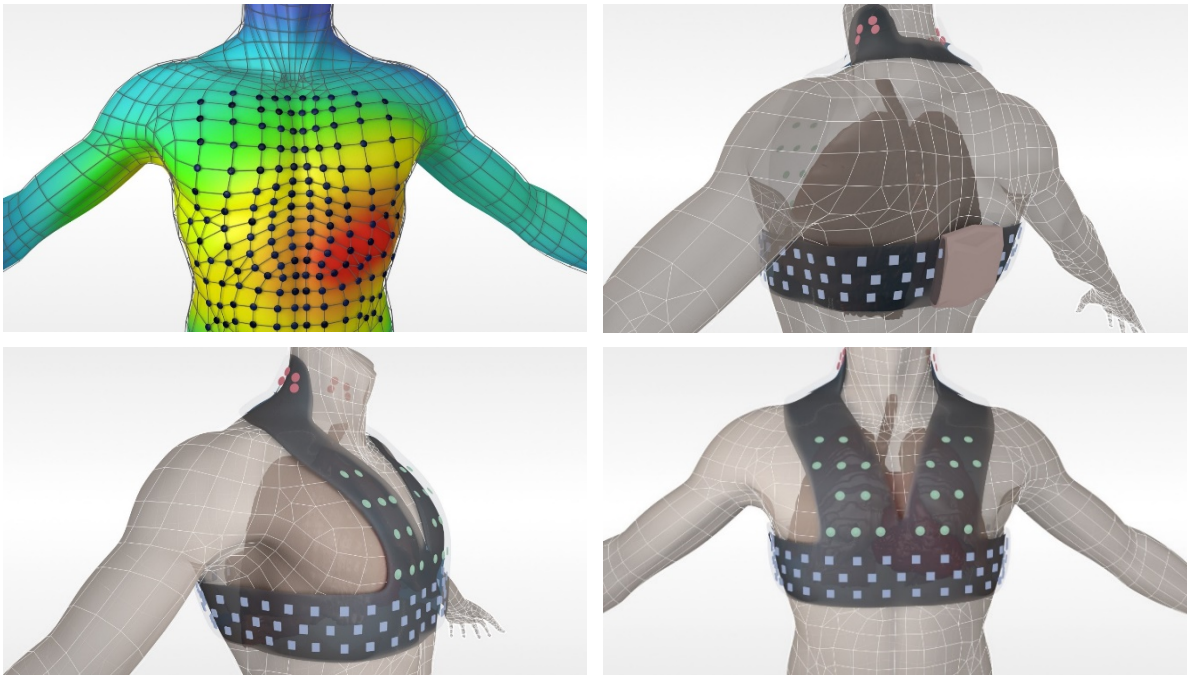


Fig. 2. Project of the measurement belts and potential map.

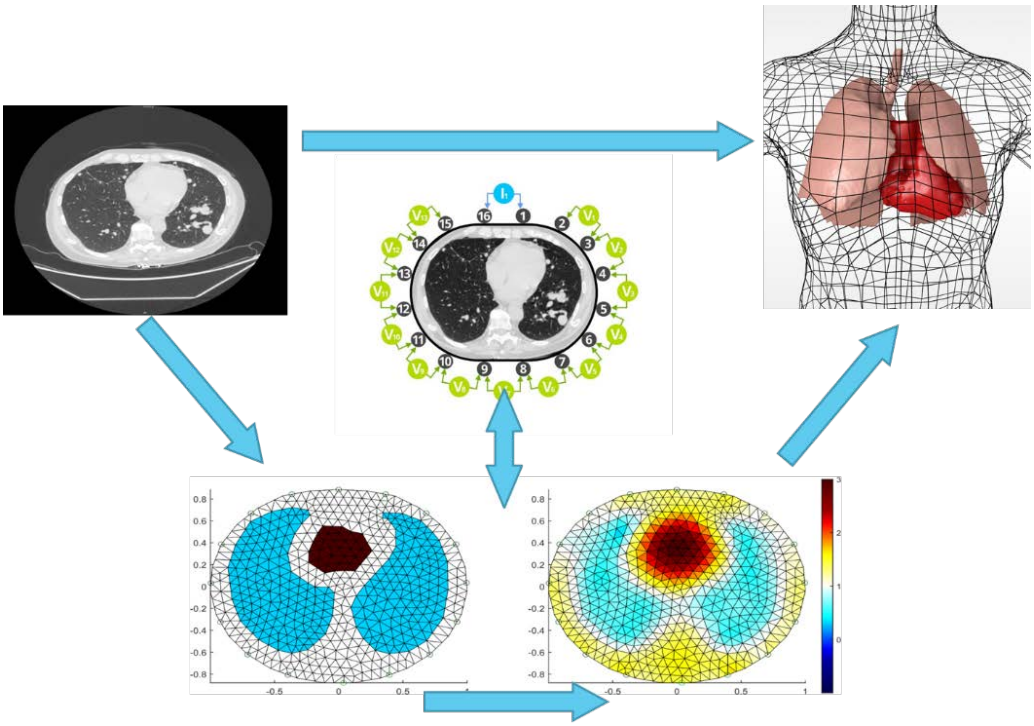


Fig. 3. The model of the imaging system

Figure 1 presents the model of system idea based on electrical tomography. The project of the measurement belts and potential map was shown in Figure 2. Figure 3 presents the idea of the imaging system with multisource collecting and analysing data using EIT, ECT, the body surface potential mapping and ECG measurements.

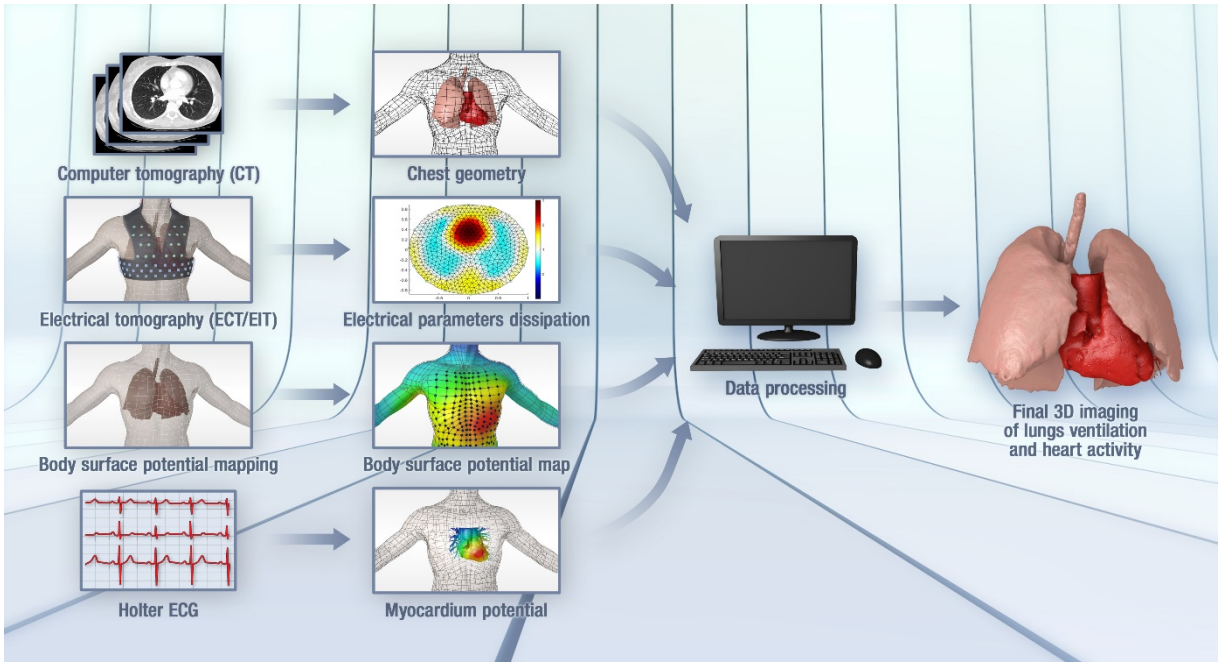


Fig. 4. Model of multisource collecting and analysing data.

3. Simulation and Reconstruction

There was prepared the image reconstruction of the lungs - 2D and 3D. Figure 5 shows the model of lungs and the image reconstruction obtained by Gauss-Newton with Tikhonov regularization, Gauss-Newton with Laplace regularization, Total Variation and Level Set Method. Figure 6 presents the 3D image reconstruction of the pulmonary model.

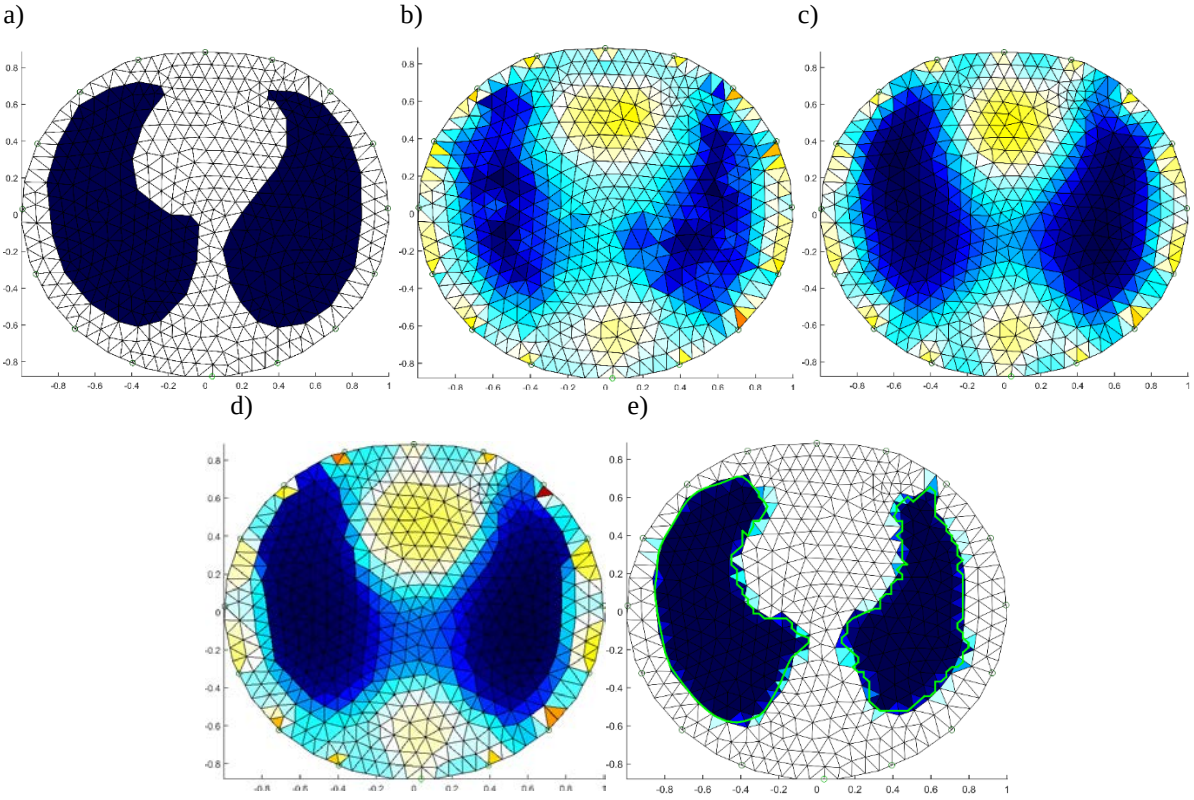


Fig. 5. The image reconstruction of lungs in a Cartesian coordinate system: a) model, b) Gauss-Newton with Tikhonov regularization, c) Gauss-Newton with Laplace regularization, d) Total Variation e) Level Set Method.

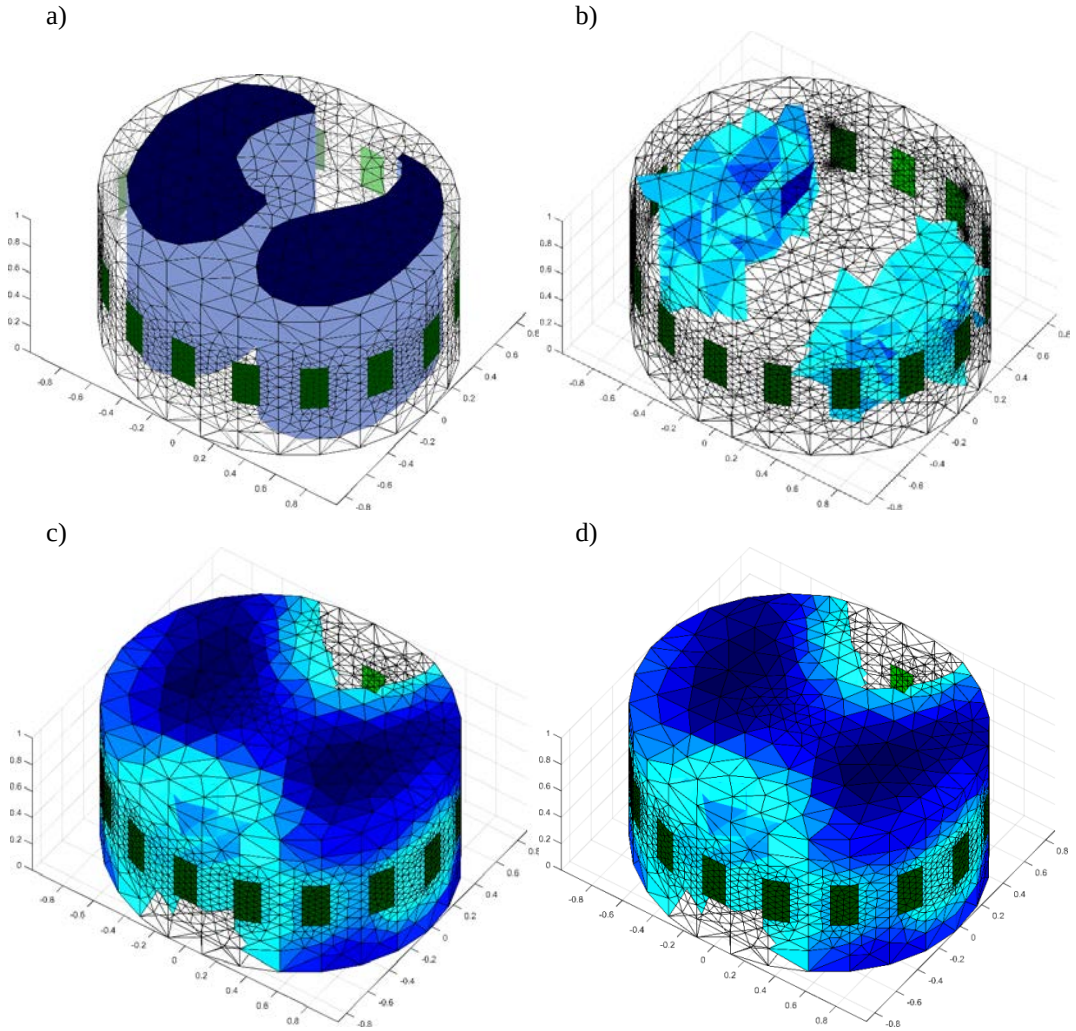


Fig. 6. The 3D image reconstruction (by 2D plate measurement electrodes) in a Cartesian coordinate system: a) model, b) Gauss-Newton with Tikhonov regularization, c) Gauss-Newton with Laplace regularization, d) Total Variation.

4. Conclusion

This paper presents the architecture of the system based on electrical tomography. In many applications of electrical tomography, such as monitoring the lungs of unconscious intensive care patients, data acquisition on the entire boundary of the body is impractical. The boundary area available for electrical tomography measurements is restricted. Physiological processes that produce changes in the electrical conductivity of the body can be monitored by hybrid algorithms. The proposed approach was based on solving the non-linear inverse problem. There was proposed an approach to EIT image reconstruction in cases where the Gauss-Newton, Total Variation and level set methods were implemented. The presented algorithms give good results. The best quality of reconstruction was achieved using the level set method.

References

1. Adler A., Lionheart W.R.B.: Uses and abuses of EIDORS: an extensible software base for EIT, Physiological Measurement, Vol. 27, 2006.
2. Alberto-P. Calder´on, On an inverse boundary value problem, (Rio de Janeiro, 1980), Soc. Brasil. Mat., Rio de Janeiro, 1980, pp. 65–73

3. Banasiak R, Wajman R, Sankowski D, Soleimani M., *Three-Dimensional Nonlinear Inversion of Electrical Capacitance Tomography Data Using a Complete Sensor Model*. Progress In Electromagnetics Research (PIER) vol. 100, pp. 219-234, 2010.
4. Filipowicz S.F., Rymarczyk T., *The Shape Reconstruction of Unknown Objects for Inverse Problems*, Przeglad Elektrotechniczny, vol. 88, Issue 3A, pp. 55-57, 2012.
5. Hamilton S.J., Siltanen S.: Nonlinear Inversion from Partial EIT Data: Computational Experiments, Contemporary Mathematics, Volume 615, 2014, pp. 105-129
6. Holder D.S.: Electrical Impedance Tomography: Methods, History and Applications Series in Medical Physics and Biomedical Engineering, London, 2005
7. Kłosowski G., Gola A., *Risk-based estimation of manufacturing order costs with artificial intelligence* [in:] Ganzha M., Maciaszek L., Paprzycki M. (eds.), Proceedings of the 2016 Federated Conference on Computer Science and Information Systems (FEDCSIS), IEEE, pp. 729-732, 2016, DOI: 10.15439/2016F323.
8. Kłosowski G., Gola A., Świć A., *Application of Fuzzy Logic Controller for Machine Load Balancing in Discrete Manufacturing System*, Lecture Notes in Computer Science, vol. 9375, pp. 256-263, 2016, DOI: 10.1007/978-3-319-24834-9_31
9. Kryszyn J., Smolik W., *2D modelling of a sensor for electrical capacitance tomography in ECTSIM toolbox* , Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 1, pp. 146-149, 2017.
10. Kryszyn J., Smolik W., Radzik B., Olszewski T., Szabatin R., *Switchless charge-discharge circuit for electrical capacitance tomography*, Measurement Science and Technology, vol. 25, Issue 11, pp. 1-9, 2014.
11. Mueller J L and Siltanen S: Linear and Nonlinear Inverse Problems with Practical Applications. SIAM 2012
12. Polakowski, K., Filipowicz, S.F., Sikora, J., Rymarczyk, *Tomography technology application*, Przeglad Elektrotechniczny, vol. 84, Issue 12, pp. 227-229, 2008.
13. Rybak G., Chaniecki Z., Grudzień K., Romanowski A., Sankowski D., *Non–invasive methods of industrial process control*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 4, Issue 3, pp. 41–45, 2014.
14. Rymarczyk T., *New Methods to Determine Moisture Areas by Electrical Impedance Tomography*, International Journal of Applied Electromagnetics and Mechanics, vol. 37, Issue 1-2, pp.79–87, 2016.
15. Rymarczyk T., Adamkiewicz P., Duda K., Szumowski J., Sikora J., *New Electrical Tomographic Method to Determine Dampness in Historical Buildings*, Archives of Electrical Engineering, vol. 65, Issue 2, pp. 273-283, 2016.
16. Rymarczyk T., Tchórzewski P., Adamkiewicz P., Duda K., Szumowski J., Sikora J., *Practical Implementation of Electrical Tomography in a Distributed System to Examine the Condition of Objects*, IEEE Sensors Journal, vol. 17, Issue:24, 2017, pp. 8166-8186, DOI: 10.1109/JSEN.2017.2746748.
17. Tchorzewski P., Rymarczyk T., Sikora J., *Using Topological Algorithms to Solve Inverse Problem in Electrical Impedance Tomography*, International Interdisciplinary Phd Workshop 2016, pp. 46-50, 2016.
18. Wajman R., Fiderek P., Fidos H., Jaworski T., Nowakowski J., Sankowski D., Banasiak R., *Metrological evaluation of a 3D electrical capacitance tomography measurement system for two-phase flow fraction determination*, Measurement Science and Technology, vol. 24, 065302, 2013.
19. Smolik W., Kryszyn J., Radzik B., Stosio M., Wróblewski P., Wanta D., Dańko L., Olszewski T., Szabatin R., *Single-shot high-voltage circuit for electrical capacitance tomography*, Measurement Science and Technology, vol. 28, Issue 2, p. 025902, 2017.

Tomasz RYMARCZYK[1,2], Karol DUDA[1], Jakub SZUMOWSKI[1], Paweł TCHÓRZEWSKI[1], Maciej PAŃCZYK[3], Jan SIKORA[3,4]

Nondestructive Methods of Data Acquisition to Examine Brick Wall Dampness

The paper presents a measurement system based on electrical tomography. It will be an innovative solution for an evaluation study of the level of dampness of the walls and the condition of the building. The prepared constructions contain 16 and 32 electrodes for measuring moisture in the brick wall on one side and two-sided. Electrical tomography, which is based on measuring potential difference, can be used to inspect historical buildings. The way in which we can define state of wall depends on the fact that every material has the unique conductance.

1. Introduction

One of the major causes of pathologies in historic buildings all over the world is the presence of moisture, particularly rising damp. Moisture transfer in walls of an old buildings, which are in direct contact with the soil, leads to a migration of soluble salts responsible for many building problems [4,12-14]. Building porous materials (e.g. brick or concrete), both natural and manufactured have pores (like a sponge) and the moisture can be pulled up against gravity (capillary effect). Figure 1 shows damp wall. Rising damp from the soil is a problem in old buildings, especially without adequate horizontal and vertical insulation of foundations. Moisture creates a danger not only to the walls, but also to human health. It promotes progress of rheumatic disorders and formation of fungus on the walls. Fungus can cause allergies and many other diseases. There are many different drainage systems (dry as watertight barriers, injection of hydrofuge products, etc.). Regardless of the method it is very important to continuously monitor the status of damp during the drying process.

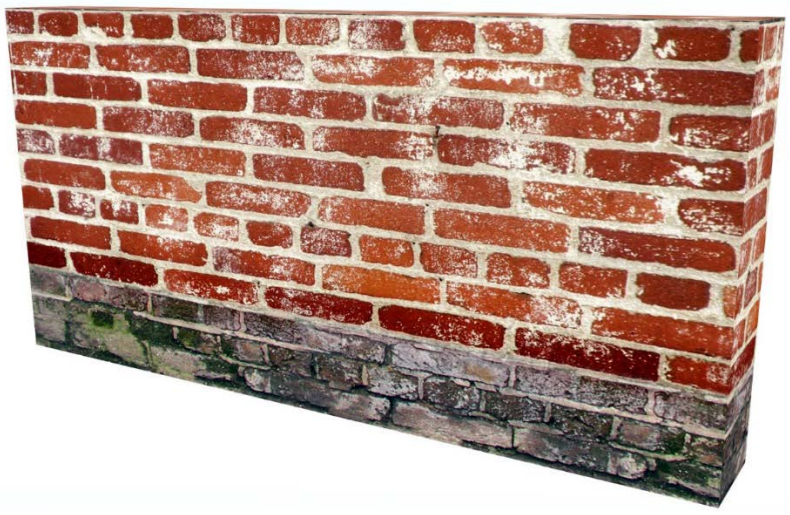


Fig. 1. Examples of excessively damp brick walls.

[1)] Research and Development Center, Netrix S.A., Związkowa 26, 20-148 Lublin, POLAND

[2)] University of Economics and Innovation in Lublin, Projektowa 4, 20-209 Lublin, POLAND

[3)] Lublin University of Technology, Nadbystrzycka 38d, 20-618, Lublin, Poland

[4)] Electrotechnical Institute, ul. Pożaryskiego 28, 04-703 Warszawa

2. Measurement systems

The electrical tomography is a technique of imaging the distribution of conductivity or permittivity inside the tested object from measurements of the distribution of potentials on the object surface [1-3,7-9,11,15-20]. Many different techniques can be used for the optimization process [5-10].

Fig. 2. Surface electrodes on the damp brick wall.

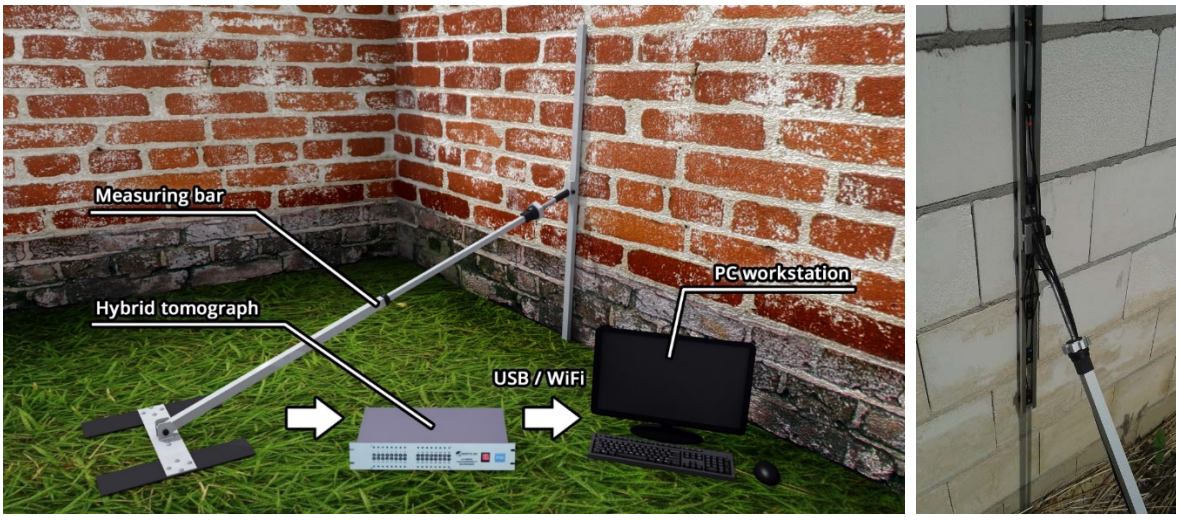


Fig. 3. The concept of measurement model.

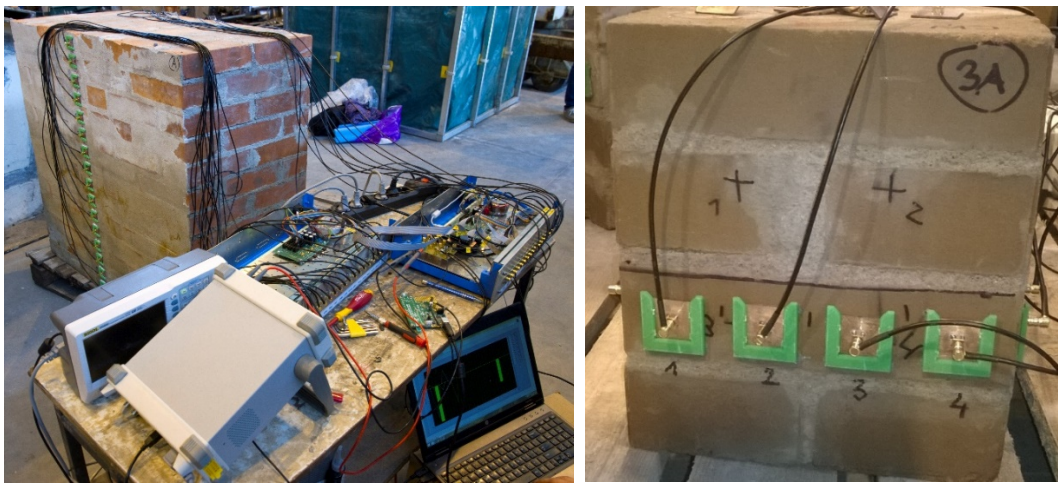


Fig. 4. The measurement system – wall I and II.

The prepared constructions contain 16 and 32 electrodes for measuring damp brick wall on one side and two-sided. Electrical Impedance Tomography, which is based on measuring potential difference. The way in which we can define state of wall depends on the fact that every material has the unique conductance. There were used necessary electrodes, meters, AC generator, multiplexer and PC with LABview and EIT/ECT devices to realize measurements. Figure 2 shows surface electrodes on the damp brick wall. The concept of measurement model is presented in Figure 3. The measurements on the brick wall are shown in Figure 3and 4. Next, let us concentrate our attention on airbrick. One should notice, that in this case we have 32 points electrodes. Electrodes have been attached along two opposite sides of airbrick. Professional test-bench have been developed in order to examine the airbrick (see Fig. 5).

Fig. 5. The laboratory measurement system with measured damp air brick.

3. Image Reconstruction

There were built and presented two special models of brick cube “wet” and “moist” with 16 electrodes. In order to solve the inverse problem, we have prepared two-dimensional finite element mesh. One should notice that surfaces of finite elements which are localized near electrodes are slight. Therefore, solution of the forward problem is accurate. Figure 5 presents the moist building brick (polarization voltage, frequency of 10kHz) - real measurements: (a) the reconstruction by Gauss-Newton method, (b) the reconstruction by hybrid method. Figure 6 shows the reconstruction in moist building brick, and Figure 7 in the airbrick. The image reconstruction in models with 16 and 32 electrodes present figure 8 and 9.

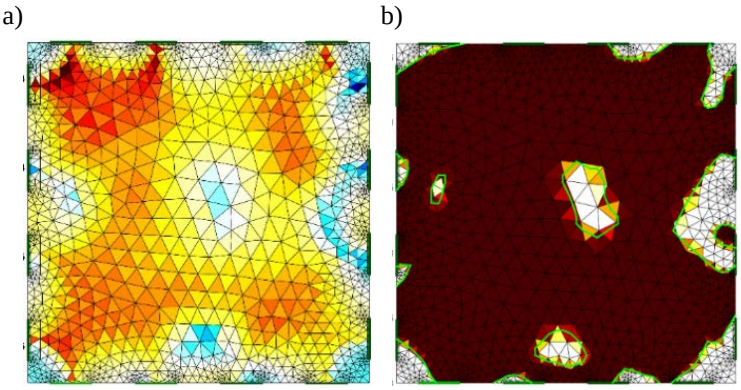


Fig. 6. The moist building brick (polarization voltage, frequency of 10kHz) – real measurements: (a) the reconstruction by Gauss-Newton method, (b) the reconstruction by Gauss-Newton Level Set.

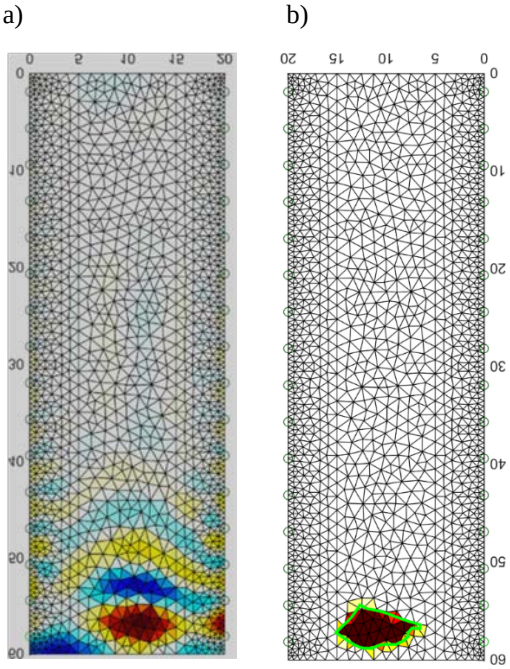


Fig. 7. The image reconstruction of the moisture in the airbrick: (a) Gauss-Newton method, (b) by Gauss-Newton Level Set.

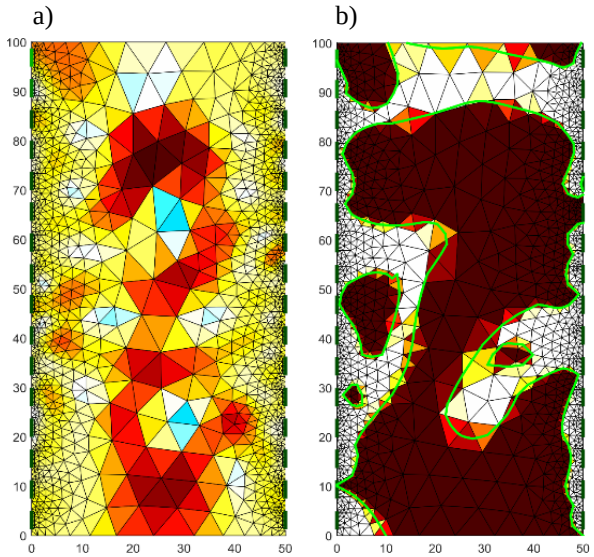


Fig. 8. The wet building brick (polarization voltage, frequency of 10kHz) – real measurement with 32 electrodes: (a) the reconstruction by Gauss-Newton method, (b) the reconstruction by hybrid method.

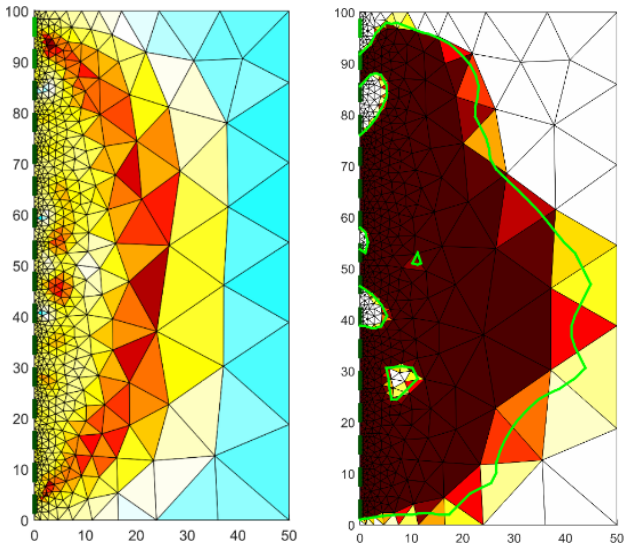


Fig. 9. The wet building brick (polarization voltage, frequency of 10kHz) – real measurement (16 electrodes): (a) the reconstruction by Gauss-Newton method, (b) the reconstruction by hybrid method.

5. Conclusion

A new nondestructive method of the inspecting the walls in historical buildings system model was presented. Parameters of the device were as there was assumed. According to assumptions it is possible to build effectively the small electrical tomography system. The such solution is measurement speed and its accuracy, however, those are still high enough for practical use. The test results for prototype devices and systems were promising. The electrical tomography is a good technique of imaging the distribution of conductivity and permittivity inside the walls and historical buildings. In conclusion, this project should be considered just a first step in a research of building effective electrical tomography systems.

References

1. Banasiak R, Wajman R, Sankowski D, Soleimani M., *Three-Dimensional Nonlinear Inversion of Electrical Capacitance Tomography Data Using a Complete Sensor Model*. Progress In Electromagnetics Research (PIER) vol. 100, pp. 219-234, 2010.
2. Filipowicz S.F., Rymarczyk T., *The Shape Reconstruction of Unknown Objects for Inverse Problems*, Przeglad Elektrotechniczny, vol. 88, Issue 3A, pp. 55-57, 2012.
3. Holder D.S., *Electrical Impedance Tomography: Methods, History and Applications*, Series in Medical Physics and Biomedical Engineering, London, 2005.
4. Hoła J., Matkowski Z., Schabowic K., Sikora J., Nita K., Wójtowicz S., *Identification of Moisture Content in Brick Walls by means of Impedance Tomography*, COMPEL, Vol. 31, Issue 6, pp. 1774-1792, 2012.
5. Kłosowski G., Gola A., *Risk-based estimation of manufacturing order costs with artificial intelligence* [in:] Ganzha M., Maciaszek L., Paprzycki M. (eds.), Proceedings of the 2016 Federated Conference on Computer Science and Information Systems (FEDCSIS), IEEE, pp. 729-732, 2016, DOI: 10.15439/2016F323.
6. Kłosowski G., Gola A., Świć A., *Application of Fuzzy Logic Controller for Machine Load Balancing in Discrete Manufacturing System*, Lecture Notes in Computer Science, vol. 9375, pp. 256-263, 2016, DOI: 10.1007/978-3-319-24834-9_31
7. Kryszyn J., Smolik W., *2D modelling of a sensor for electrical capacitance tomography in ECTSIM toolbox* , Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 1, pp. 146-149, 2017.

8. Kryszyn J., Smolik W., Radzik B., Olszewski T., Szabatin R., *Switchless charge-discharge circuit for electrical capacitance tomography*, Measurement Science and Technology, vol. 25, Issue 11, pp. 1-9, 2014.
9. Majchrowicz M., Kapusta P., Jackowska-Strumiłło L., Sankowski D., *Acceleration of image reconstruction proces in electrical capacitance tomography 3D in heterogeneous, multi-GPU system*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 7, Issue 1, pp. 37-41, 2017. DOI: 10.5604/01.3001.0010.4579
10. Polakowski, K., Filipowicz, S.F., Sikora, J., Rymarczyk, *Tomography technology application*, Przeglad Elektrotechniczny, vol. 84, Issue 12, pp. 227-229, 2008.
11. Rybak G., Chaniecki Z., Grudzień K., Romanowski A., Sankowski D., *Non–invasive methods of industrial process control*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 4, Issue 3, pp. 41–45, 2014.
12. Rymarczyk T., *New Methods to Determine Moisture Areas by Electrical Impedance Tomography*, International Journal of Applied Electromagnetics and Mechanics, vol. 37, Issue 1-2, pp.79–87, 2016.
13. Rymarczyk T., Adamkiewicz P., Duda K., Szumowski J., Sikora J., *New Electrical Tomographic Method to Determine Dampness in Historical Buildings,* Archives of Electrical Engineering, vol. 65, Issue 2, pp. 273-283, 2016.
14. Rymarczyk T., Tchórzewski P., Adamkiewicz P., Duda K., Szumowski J., Sikora J., *Practical Implementation of Electrical Tomography in a Distributed System to Examine the Condition of Objects*, IEEE Sensors Journal, vol. 17, Issue:24, 2017, pp. 8166-8186, DOI: 10.1109/JSEN.2017.2746748.
15. Tchorzewski P., Rymarczyk T., Sikora J., *Using Topological Algorithms to Solve Inverse Problem in Electrical Impedance Tomography*, International Interdisciplinary Phd Workshop 2016, pp. 46-50, 2016.
16. Wajman R., Fiderek P., Fidos H., Jaworski T., Nowakowski J., Sankowski D., Banasiak R., *Metrological evaluation of a 3D electrical capacitance tomography measurement system for two-phase flow fraction determination,* Measurement Science and Technology, vol. 24, 065302, 2013.
17. Smolik W., Kryszyn J., *Linea rover ranges iterative algorithms for image reconstruction in electrical capacitance tomoraphy*, Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, vol. 1, pp. 115-120, 2017.
18. Smolik W., Kryszyn J., Radzik B., Stosio M., Wróblewski P., Wanta D., Dańko L., Olszewski T., Szabatin R., *Single-shot high-voltage circuit for electrical capacitance tomography*, Measurement Science and Technology, vol. 28, Issue 2, p. 025902, 2017.
19. Tai C., Chung E., Chan T., *Electrical impedance tomography using level set representation and total variational regularization*, Journal of Computational Physics, vol. 205, no. 1, pp. 357–372, 2005.
20. Zhang C., *Electrical resistance tomography (EIT) by using an ECT sensor. Imaging Systems and Techniques*, IEEE International Conference, 2010.

Jarosław KAWECKI[1], Łukasz JANUSZKIEWICZ[2], Paweł OLEKSY[3], Robert KAWECKI[4]

Wearable inertial positioning system for indoor localization of people

The paper presents the wearable electronic system that localizes people inside the buildings with inertial sensor. This system contains a smartphone and a shirt with electronic modules that were sewn with conducting thread into it. The electronic modules are used to measure acceleration and to transmit the data over Bluetooth wireless system to the smartphone. The smartphone allows to collect the measurement data and to calculate the human position in the building. It can be used also to send estimated localization of the person over the internet. The results of the system tests performed inside the building are also presented.

1. Introduction

Localization systems that are able to operate inside the buildings are being intensively developed in last years. Among many commercial applications of such systems they can be used also for increasing safety of people that may be the subject of heart attack or other health problems in the unknown building. The miniaturization of electronic modules and sensors allows to use them in wearable systems. Such electronic system can benefit from being placed very close to the human body especially when it aims at monitoring human motion [1]. This is the case of inertial systems for indoor localization of people that utilizes the measurements of acceleration and rotation of human body that can be than recalculated to the distance and direction of movement. Those systems don't need the installation of additional radio transmitters what is the drawback of systems which use „Beacons" or other wireless networks WiFi [2][3][4]. The inertial systems that are designed to track people's motion can be less accurate than systems that utilize dedicated radio transmitters while using systems that analyze time difference of signal arrival. However they are still being developed because they don't need additional infrastructure which reduces the cost.

System that is presented in this article uses miniature accelerometers and Bluetooth transmission module, which were integrated with clothes. This sensors allow to measure a linear and angular acceleration and Bluetooth module transmits this data to smartphone. The program that is running on smartphone estimates user movement and display on the screen its current position. To preserve the comfort of system user it was necessary to use miniature power sources that are lightweight enough to be attached to the T-shirt. Due to their small capacity (600 mAh, 3V) the power usage of Bluetooth module had to be limited. It may result

[1]Ph.D. student, Lodz University of Technology Institute of Electronics, 211/215 Wolczanska Str., 90-924 Lodz POLAND, e-mail: jaroslaw.kawecki@p.lodz.pl
[2]Ph.D., Lodz University of Technology Institute of Electronics, 211/215 Wolczanska Str., 90-924 Lodz POLAND, e-mail: lukasz.januszkiewicz@p.lodz.pl
[3]Ph.D., Lodz University of Technology Institute of Electronics, 211/215 Wolczanska Str., 90-924 Lodz POLAND, e-mail: pawel.oleksy@p.lodz.pl
[4]Ph.D., Lodz University of Technology Institute of Electronics, 211/215 Wolczanska Str., 90-924 Lodz POLAND, e-mail: robert.kawecki@p.lodz.pl

with the loss of data link between module on the shirt and smartphone in the pocket. In the paper an analysis of path loss for different location of the module on the shirt is presented. It was used in the project for the design the placement of wireless transceiver on the shirt.

2. Wearable inertial positioning system

The wearable system that was designed to localize people inside the building is presented in picture 1. It consists of t-shirt which electronic modules and a smartphone. T-shirt modules were sawn into the shirt with electrically conductive sewing threads that played the role of electrical interconnections between modules. The central module is Adafriuit Flora microcontroller [5] that gather data from measurement moduleAdafruitBNO055 that contains accelerometer, gyroscope and magnetometer. The data processed by Adafruit Flora module are then send to smartphone with Adafruit Bluetooth LE module. To preserve the comfort of the user the miniature battery module was placed directly on the shirt. To save its energy the least computationally complex algorithm was used. The purpose of the developed algorithm implemented in the microcontroller was to estimate the user's displacement and to send real-time data to the mobile device, eg. smartphone with the android operating system, where the appropriate algorithm determines the location.

In the figure 2 the electronic modules that have been sewn into the shirt are presented. The Adafruit Flora [5] module (number 1 in Fig. 2), is an AVR Atmega 32u4 microcontroller that is compatible with the Arduino environment. The Flora microcontroller receives data from the BNO055 inertial module (number 3 in Fig. 2) and processes them accordingly to obtain the user's location. The designated relocation is then sent by radio using the Bluetooth Low Energy (LE) module (number 2 in Fig. 2) to the mobile application, where the target location calculations are performed. The use of the LE-type Bluetooth module reduced the power consumption of the entire textile system to 23 mA. The module is powered by a 4.5V cell (number 4 in fig.2).

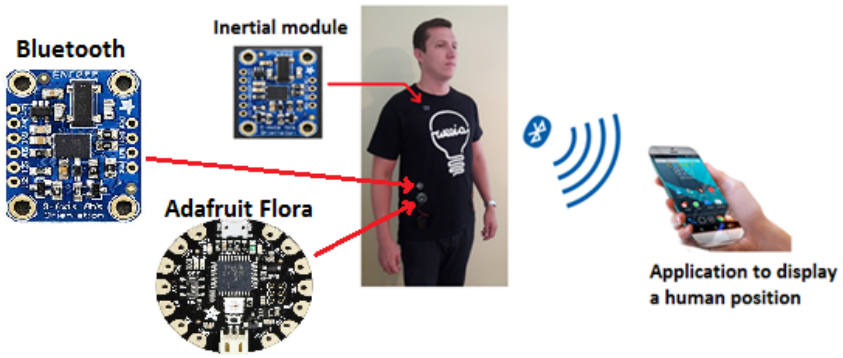


Figure. 1: Structure of indoor localization system.

Figure. 2: An electronic system sewn into the shirt: 1-microcontroller, 2- bluetooth transmitter, 3-inertial module, 4-battery.

A thread type 316L made of stainless steel [6] was used for the electrical connections between the t-shirt modules. According to this, it is possible to connect all modules and to transmit measured data from the accelerometer module to the microcontroller preserving user comfort. The linear resistance of the electro-conductive strand made of stainless steel is relatively high (30 Ω / m) compared to conductors made of silver or copper. It was used with Adafruit modules because of its resistance to corrosion under the influence of sweat and moisture.

The localization algorithm consists of a method for calculating the number of steps made while walking and for estimating the direction of human movement. The limited complexity of the calculations that should be performed in this algorithm does not require high processing power, what results with low power consumption. The proper functioning of the algorithm requires the placement of sensors in the area of the shirt that is well adhering to the body. Otherwise movements of the clothes could introduce artifacts in acceleration measurements. As a result of the experiments, it was determined that the location of the sensors in the area of the arm guarantees good adhesion to the body and does not cause user discomfort.

In the first version of the system, the Bluetooth module was placed in the area of the arm due to the proximity to the measuring module. It has been observed, that in large rooms where the building partitions were far from the system user, the connection between the Bluetooth module and the smartphone which was in the trouser pocket was broken.

Another variant of the t-shirt had a measurement module located near the shoulder, and a Bluetooth module located at the bottom of the t-shirt, directly over the trouser pocket. In this case, the length of the conductive sewing threads which connects the Bluetooth module and the measuring module was about 70cm, and the connection resistance was 22 Ω. With so much serial resistance in the data line, it was impossible to get the system working properly.

In order to select such location of the Bluetooth module, that is suitable for proper data transmission between the T-shirt module and the smartphone in the pocket and between the accelerometer and the t-shirt transmitter, the analysis of the radio link path loss in the designed system was made for wireless transmission in the proximity of the human body.

3. Path loss in the proximity of human body

The path loss measurements in the link between Bluetooth module and smartphone have been performed for different Bluetooth module placement. The wireless module works in the Bluetooth Low Energy standard in 2.4 GHz band and uses a miniature ceramic antenna. It was very important to choose the best placement for the Bluetooth module towards the smartphone because the path-loss changes significantly with even the minor changes of antenna placement on the body. Figure 3 shows two parameters which were investigated to characterize the transmission on the shirt: the distance between the accelerometer and the Bluetooth module – R and the distance between shirt and smartphone – L. It was important to properly deploy the Bluetooth module to ensure the best radio link stability (L parameter) and the minimum resistance between the inertial module and the Bluetooth module (parameter R). The position of the inertial sensor remains constant because it is the best place to measure acceleration of human body.

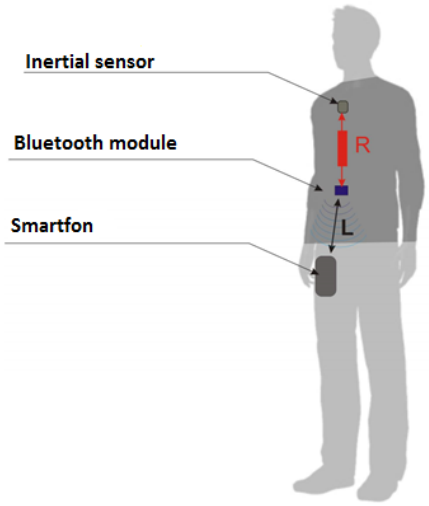


Figure. 3: Definition of parameters L and R

In the first case, it was assumed that the Bluetooth module is located on the shoulder. In this case the L parameter was equal to 70 cm, the attenuation in wireless link was 85 dB, the value of R parameter was 2 cm and the resistance of threads was 0.6 Ω. For a Bluetooth antenna on the chest at a distance of L = 54 cm from the smartphone antenna the attenuation was 78dB and the value of the parameter R = 17.6cm with a resistance of 5.3Ω. For the L = 35 cm, the attenuation was 65 dB, the R = 38.5cm, and the resistance was 11.5Ω. For the cases where the resistance value was above 10Ω, there were communication problems.

The best solution was to place the Bluetooth module 54 cm away from the smartphone and about 17 cm away from the inertial module. This allows for the best quality of the radio link and the wired connection.

4. The evaluation of localization system performance

The localization algorithm uses data from inertial sensors for tracking the movements of the person that are then recalculated for the position of the person inside the building. This task was divided into two programs: the first one that was implemented in the Adafruit Flora microcontroller which was sewn into the shirt. It is responsible for gathering data from the inertial module and calculating the displacement of the person. The second program is

implemented on smartphone and calculates position from the displacement and presents it on the building plan. The first program basing on the data from the accelerometer, calculates the distance traveled by the user within a fix time period of 1 second. If the value of acceleration is greater than the predefined threshold, it is considered that the user has taken a step. The algorithm detects how many steps the user has taken. The step length was fixed in this version of algorithm and estimated to be 70cm. The program determines the direction of the user's movement basing on the data provided by magnetometer. This allows to determine the displacement in horizontal plane (x, y coordinates) by which the user's position has changed.

The second program which works on the smartphone, is responsible for processing the displacement information received from the first algorithm. Assuming that the initial position of the person inside the building is known (what is true for the interiors with one doors) the current position is being calculated by adding the displacements provided by the first program running on microcontroller. The position of the user is then visualized on the map and can be also stored in the log-file on the Smartphone.

To verify the performance of the system the tests were carried out in the building of the Lodz University of Technology. For this purpose it was assumed that the person can move inside the zone of dimensions equal to 20m by 11.5m. Firstly the procedure of magnetometer (that is a part of the inertial module) calibration was launched, before starting the experiment. To correctly calibrate it should be rotated by 360 ° around each axis. Figure 4 presents mobile application screen which shows the path walked by the person in the building. The yellow mark on the map shows the actual path while the red line shows the path that was estimated by the localization program. Red points mark the place where that the user has taken a step.

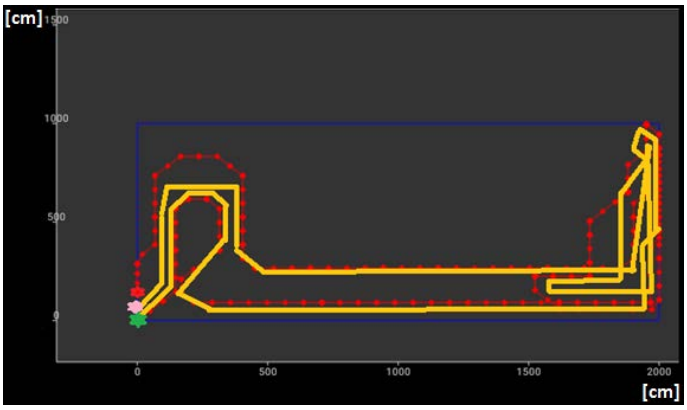


Figure. 4: The mobile application screen recorded during test (the red series is the path walked by the user, estimated from the algorithm, yellow is actual path walked by the person).

The starting point of walk was placed in the center of coordinate system x=0m, y=0m. As it was presented in figure 4, the path estimated from the algorithm (red series) was not the same as the reference line (yellow series). This shows the error of localization system. It can be caused by the difference of average value of the step length fixed in the algorithm and the length of real steps made by person. Also the limited precision of accelerometers had influence to the localization error. The destination point of the system user obtained from the algorithm was similar to the real localization of the end point of the walk. A person with a textile shirt reached the similar to designated point at x = 0m, y = 2m. The designed point was x=0m, y=0m.

5. Conclusions

In this paper a system for people localization that utilizes accelerometers integrated into a t-shirt was presented. The incorporation of miniature electronic modules into the T-shirt preserved the users comfort. Another advantage of this solution is the stable placement of accelerometers towards human body what decreases the measurement errors from a random movements of sensors. The power supply provides at least 10 hours of system time. RSSI Attenuation measurements have shown that the Bluetooth module is appropriately positioned relative to the smartphone to achieve a stable data link [7].

The design of system that combine textiles and electronic technologies requires the analysis of many issues such as electrical resistance of threads and radio signal attenuation in the proximity of human body. Taking into account both aspects, it is possible to select the location of the modules that provide the required energy budget of the wireless link and the reliable operation of the wire connection. The system that was developed will be used for further investigation on localization algorithms that utilize data from inertial sensors.

References

1. Yang C., Hsu Y., *„A Review of Accelerometry-Based Wearable Motion Detectors for Physical Activity Monitoring"*, Sensors 2010, 10(8), 7772-7788
2. Hsu H. H., Peng W. J., Shih T. K., Pai T. W. and Man K. L., *"Smartphone Indoor Localization with Accelerometer and Gyroscope",2014 17th International Conference on Network-Based Information Systems, Salerno*, 2014, pp. 465-469.
3. Zhang R. , Hoeflinger F. , Gorgis O. and Reindl L. M.,., *"Indoor localization using inertial sensors and ultrasonic rangefinder*", 2011.
4. Hsu Y. L., Wang J. S., Chang C. W., "*A Wearable Inertial Pedestrian Navigation System with Quaternion-Based Extended Kalman Filter for Pedestrian Localization*," in *IEEE Sensors Journal* , vol.PP, no.99, pp.1-1
5. https://www.adafruit.com/product/659
6. https://www.adafruit.com/products/641
7. Hausman S., Januszkiewicz Ł.,*"Impact of Indoor Environment on Path Loss in Body Area Networks",Sensors*2014, *14*, 19551-19560.

Paweł OLEKSY[1], Łukasz JANUSZKIEWICZ[2], Robert KAWECKI[1], Jarosław KAWECKI[1],

The measurements of human body impact on transmission channel in wireless body area networks

Continuous development of the Wireless Body Area Networks and wireless localization systems resulted in an increasing number of transceivers that are located inside the buildings. In the indoor environment the number of people and their position can vary greatly. Moreover, in the currently used GHz bands, human body attenuates the electromagnetic waves significantly. It is therefore necessary to take into account the influence of people on the propagation channel in indoor wireless systems. In the paper a test-bench for path loss measurement in wireless body area networks has been presented. The results obtained for a few measurement scenarios are also presented.

1. Introduction

Design process of the Wireless Body Area Networks (WBAN) requires the analysis of the impact of human body on the propagation of electromagnetic waves. Typically, the impact of human body on the antennas placed on body or inside body (implants) is considered because human body introduces a strong absorption of electromagnetic waves in microwave band, and as a result significantly affects the impedance matching of the antenna and its radiation pattern [1]. These effects depend strongly on the spatial orientation of the antenna and human body [2], and on the distance between antenna and the body. These are particularly important issues in on-body networks, in which the transmitter and receiver are located in the proximity of human body.

The impact of the human body on the propagation channel is also not negligible in case of off-body networks where transceivers are placed on the body and communicate with the wireless devices located in a distance from the body. In those systems, the additional obstacle that are present in transmission channel, such as people situated between the transmitter and the receiver, may increase the path loss by a few or even several dB [3]. This phenomenon is particularly difficult to investigate in the case of multi-path propagation because the direct signal from the transmitter can be attenuated as well as signals reflected by the walls in the room.

In the indoor environment also the designers of wireless localization systems have to face the problem of additional attenuation in the transmission channel introduced by the human body. It is important because those systems often use a reference field strength distribution

[1]Ph.D. student, Lodz University of Technology Institute of Electronics, 211/215 Wolczanska Str.,90-924 LodzPOLAND, e-mail: pawel.oleksy@p.lodz.pl
[2]Ph.D., Lodz University of Technology Institute of Electronics, 211/215 Wolczanska Str., 90-924 Lodz POLAND, e-mail: lukasz.januszkiewicz@p.lodz.pl
Ph.D. student, Lodz University of Technology Institute of Electronics, 211/215 Wolczanska Str., 90-924 Lodz POLAND, e-mail: robert.kawecki@p.lodz.pl
[2]Ph.D. student, Lodz University of Technology Institute of Electronics, 211/215 Wolczanska Str., 90-924 Lodz POLAND, e-mail: jaroslaw.kawecki@p.lodz.pl

maps which attempts to match the position of the mobile terminal (e.g. in fingerprinting techniques) based on the received power [4]. Presence of people in the room or their movement causes that the power received at a given point will be different from the reference pattern. For this reason, it is important to examine the influence of people on the transmission channel to appropriately set filter parameters implemented in the location algorithm.

The design of wireless systems that are operating in the proximity of human body is usually made with computer programs that perform simulations with numerical models of the human body. The on-body systems can be simulated with the finite difference time domain method or finite element method that are capable of simulating entire body using the computer resources that are accessible in typical computer. For an off-body channel that covers potentially quite large environments those methods cannot be applied because the model is too large (relative to the proportion to wavelength) and the computer resources needed for the simulation exceed available memory size. It is therefore important for designers to investigate the problem basing on measurements of channel parameters in a variety of scenarios that take into account also the effect of obstacles.

To characterize the propagation channel, a significant number of measurements is needed to observe and compare the results for different scenarios of experiment. The measurement time can be greatly reduced by automating the measurement process. In this paper a measurement system designed to automate investigation on transmission channel characteristics in off-body networks is presented.

2. Measurement system design

The general concept of the system for measuring the effect of the human body presence on the transmission channel assumes that the transmitter has fixed position and uses omnidirectional antenna and the receiver is located on the moving person and uses wearable antenna. is The portable spectrum analyzer (Rohde&Schwarz FSH8) was attached to the person to measure the power received from the transmitter. The received power measurements were performed at various distances from the transmitter.

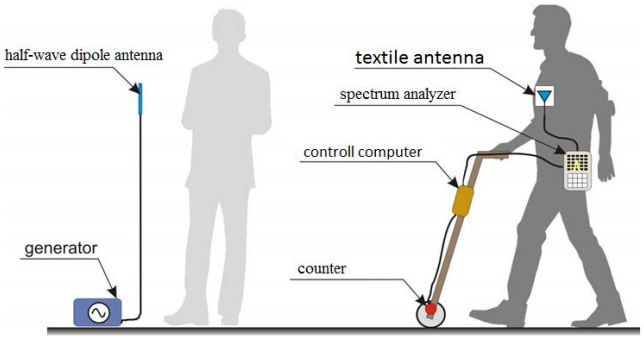
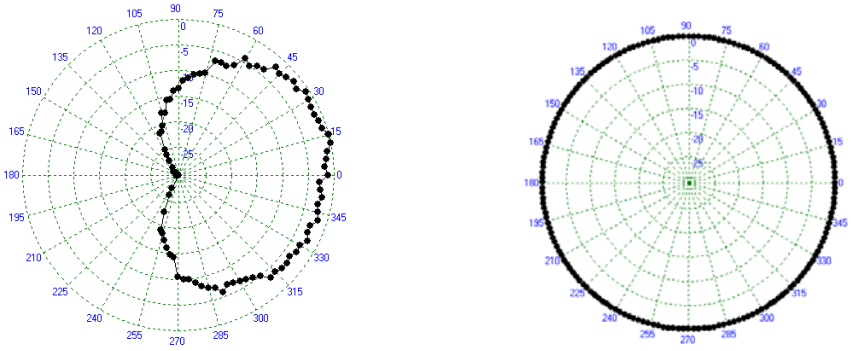


Figure. 1: Measuring system

The measurement system is presented in Figure 1. In the experiments the Rohde &Schwarz SMB 100A generator was used as a signal source (transmitter) and the transmitting antenna was half-wave dipole operating at 2.45 GHz frequency. The receiving antenna was attached to the chest of person performing the measurement and connected to spectrum analyzer. For this purpose the textile V antenna [5] was used as it has a good impedance matching in the 2.4 - 2.5 GHz band for on-body location. The radiation pattern of the antennas used in this experiment are presented in figure 2. The distance measuring system was designed to synchronize

the measurements of received power with the distance between the transmitting and receiving antenna. It uses the measuring wheel equipped with a rotating counter placed at the end of the operator's rod. The central unit of the measuring system is a control computer that records data from the spectrum analyzer and measured distance.



a. b.

Figure. 2: Radiation pattern of antennas used for measurements, horizontal plane, normalized to maximum gain G_{max}: a – wearable antenna (G_{max}=4.5 dBi), b – half wave dipole (G_{max}=2.15 dBi)

The control computer was made with the OrangePi platform and the LUbuntu operating system has been used. Thanks to its small size, the device is handy and the measurement can be carried out without difficulty by one person. An important feature of the chosen platform is the availability of Ethernet, USB and GPIO ports, enabling the use of a spectrum analyzer, external memory, communication interface and sensor wheel. To ensure full mobility of the developed tool, the computer is powered by a miniature Power-bank device with a maximum current of 2A

The control program was written in Python. The script runs automatically when the device is turned on, after connection with peripheral devices is established and configuration procedure is launched. Sending the appropriate request to the spectrum analyzer was performed using the Ethernet interface. For each command error handling has been implemented so that the user is informed about possible problems related to e.g. data transmission. The program is responsible for recording the received power value and measured distance between the transmitting and receiving antennas. Impulse from a magnetic sensor located near the wheel triggers a new measurement every ¼ turn. A 745 mm diameter wheel was used to build the device, which allowed to achieve resolution of 585 mm. An additional feature that facilitates the operation of the instrument is the possibility to record measurements on the external disk.

3. Examples of measurement results

The aim of the experiment was to test the usability of measurement system. It was configured to investigate the influence of people on the path loss in WBAN networks operating in the off-body scenario. The received power was measured for various distances between the transmitter and the receiver antenna located on the chest of the moving person. Measurements were made for a sinusoidal signal with a frequency of 2.45 GHz and a transmit power of 10 dBm.

The path-loss in off-body wireless body area networks operating in indoor environment depends on many factors like the frequency of the signal, structure of the building and presence of obstacles (people). Due to this reason, the measurements were made for 4 different scenarios that covered different environment and the height of the transmission antenna mast. The walls of the rooms in which the measurements were performed, were made of typical

materials that are gypsum board and brick, while the floor of the rooms was made of concrete. The furniture in the rooms (scenarios I-III) were made of wood. For each scenario, a measurement was repeated for 5 different variants of the arrangement of persons in the transmission channel, as illustrated in Figure 3. The parameters of measurement scenarios are presented in table 1.

Table 1. The measurement scenarios.

Scenario number	Length of the room [m]	Width of the room [m]	Transmit antenna height [m]	furniture
1	11.6	6.6	1.5	yes
2	11.6	6.6	1.8	yes
3	7.8	3.8	1.5	yes
4	11.6	6.4	1.5	no

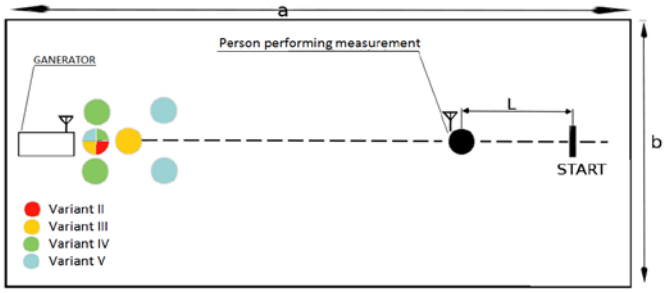


Figure. 3: People location during measurements. Colors are assigned to different location variants. L- distance from the measurement starting point, a - room length, b - room width

Measurement scenarios were repeated for 5 different variants of obstacles that were corrupting the propagation channel:

I. no person is in the proximity of transmitter,

II. one person is in the proximity of transmitter,

III. two people are in the proximity of transmitter (one after the other),

IV. three people are in the proximity of transmitter (side by side),

V. three people are in the proximity of transmitter: one directly between the transmitter and the receiver and two persons on the sides.

If people were placed towards the transmitter, the measurement was performed only in the distance up to the first person. In the presented experiment, the measurement starting point was 5.3 m from the transmit antenna. The person performing the measurement was facing the signal source with wearable antenna (what refers to angle 0° in fig. 2a) so that his body did not introduce any further attenuation.

Figure 4a shows the results obtained for the first scenario. Significant differences in received power between variants occur only when person performing the measurement was about 2m from the transmit antenna. Consequently, it was decided to compare the median of received power only for this distance range. The results obtained for all other scenarios and variants are shown in figure 4 and compared in Table 2.

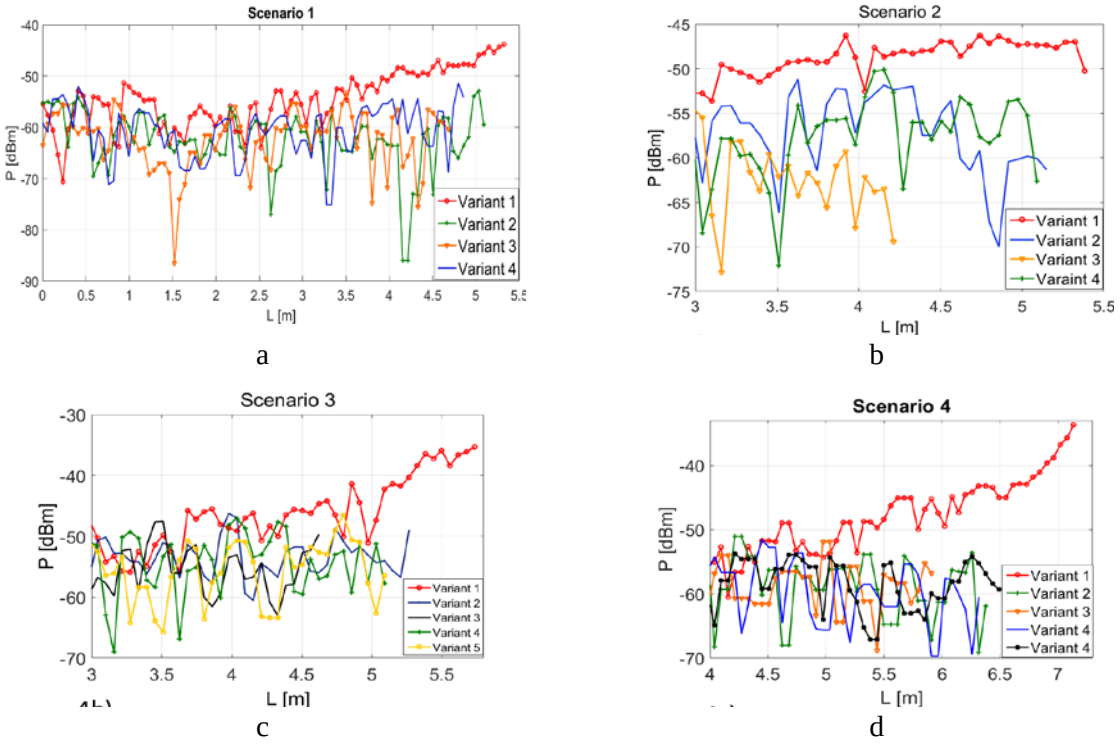


Figure. 4 Measurement results for different scenarios: a - scenario 1, b - scenario 2, c –scenario 3, d –scenario 4

Table 2: Median of received power in the 2m range from transmit antenna –different variants of person location

	Scenario I	Scenario II	Scenario III	Scenario IV
variant I	-48.9 dBm	-47.9 dBm	-45.6 dBm	-45 dBm
variant II	-62.8 dBm	-57.5 dBm	-53.1dBm	-59.3 dBm
variant III	-60.4 dBm	-62.8 dBm	-56 dBm	-58.4 dBm
variant IV	-57.1 dBm	-55.9 dBm	-53.4 dBm	-61.1 dBm
variant V			-54 dBm	-59.5 dBm

The results of measurements show that the presence of people in the room where the off-body WBAN is operating influences the propagation channel. The received power level in wireless link was approximately 10 dB lower in the proximity of the transmitter that was shadowed by the person. It can be seen in figure 4 especially at distance of 1.5-2 m from the transmitter. The biggest attenuation can be observed for variant III where two persons were placed in a line between transmitter and receiver. It is also important to note that environment has significant impact on obtained results what is visible in results presented in figure 5 that were obtained for different environments but without additional obstacles.

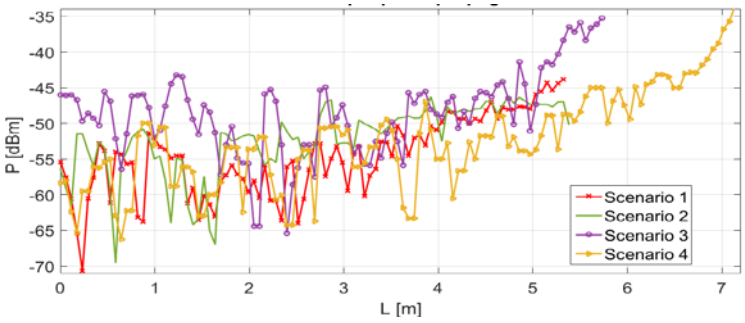


Figure. 5: Results comparison for variants without people in propagation channel

4. Conclusions

The measurements that were carried out with the designed measurement system show that proposed solution can be useful for further experiments. It was able to gather 600 measurement points in less than 1 min. In addition, the designed tool is versatile enough to allow measurements not only for BAN networks, but also to measure the electrical field strength distribution in a building for WiFi networks.

The results show that the impact of human body on received power level in off-body networks is not negligible and it should be further investigated. Attenuation effect is most pronounced when the transmitter is located close to the people. The results show that one person can attenuate signal by -15dBm, which can lead to significant errors, especially in localization systems based on fingerprinting method. It is therefore essential to conduct research on an off-body propagation channel. Further research will be carried out to investigate the impact of the position of the antenna on the human body to the path loss.

References

1. Pandian P. S., Safeer K. P., Gupta Pragati, Shakunthala D. T., Sundersheshu B. S. and Padaki V. C. *"Wireless Sensor Network for Wearable Physiologi-cal Monitoring"*, Journal of Networks, Vol. 3, No. 5, May 2008, str. 21 - 29
2. Januszkiewicz Ł. and Hausman S., "*Simplified human phantoms for wireless body area network modelling*," *2015 9th European Conference on Antennas and Propagation (EuCAP)*, Lisbon, 2015, pp. 1-4.
3. Benaissa S., Plets D., Tanghe E., Verloock L., Martens L., Stevens N., Hoebeke J., Sonck B., Tuyttens F., Vandaele L. & Joseph, *„Off-Body Channel Modelling and Network planning at 2.4 GHz for Dairy Cows on Pastures"*, Leuven, Belgium, 22/05/16 - 24/05/16, .
4. Piao Xinglin , Zhang Yong , Li Tingshu, Hu Yongli, Liu Hao, Zhang Ke and Ge Yun *„RSS Fingerprint Based Indoor Localization Using Sparse Representation with Spatio-Temporal Constraint" Sensors*2016, *16*, 1845.
5. Januszkiewicz Ł, Hausman S., Nowak I., Krucińska I., *„Textile veeantena made with PVD process, International Journal of Applied Electromagnetics and Mechanics"*, Vol. 46, Issue 2, 2014, pp.361-365
6. Amini N., Xu Wenyao, Li Zhinan, Ming-Chun Huang and Sarrafzadeh M., "Experimental analysis of IEEE 802.15.4 for on/off body communications," *2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*, Toronto, ON, 2011, pp. 2138-2142.
7. El Azhari M., Nedil M., Ben Mabrouk I. and Talbi L., "*Characterization of a NLOS Off-Body Channel at 2.45 GHz Using Patch Antenna Inside a Mine*," *2015 IEEE International Conference on Ubiquitous Wireless Broadband (ICUWB)*, Montreal, QC, 2015, pp. 1-5.
doi: 10.1109/ICUWB.2015.7324423.
8. Ambroziak *S. J.*, "*An Off-Body Channel Model for Body Area Networks in Indoor Environments*," in *IEEE Transactions on Antennas and Propagation*, vol. 64, no. 9, pp. 4022-4035, Sept. 2016.

Robert KAWECKI[1], Paweł OLEKSY[2], Jarosław KAWECKI[3], Krzysztof PIWOWARCZYK[4], Sławomir HAUSMAN[5], Piotr KORBEL[6]

Application of particle filter algorithm for indoor terminal positioning

The paper shows the application of the particle filter to wireless indoor positioning. Localization system employs Bluetooth Low Energy beacons and mobile device to determine the position of the user within the test area. The proposed system uses particle filtering algorithm to estimate the user's current position from RSSI measurements and knowledge of reference electromagnetic field distribution computed for each beacon located in the building. The article presents the results of localization with the use of the proposed algorithm. Furthermore, it indicates the measurements of the positioning error and compares them to the accuracy obtained with the use of the popular k-NN classification method.

1. Introduction

Accurate and fast localization of people and objects indoors is vital for deployment of numerous systems aimed at improvement of people's comfort and safety. A growing number of publications concerning this subject proves that this is a challenging task and it leads to various solutions depending on the application scenario [1]. Among many technologies that are being applied for this purpose, one of the most widely used are radio beacons [2-3]. It is based on the analysis of the parameters of radio links that allows to find the distance between the user terminal and the transmitters of known locations. Another popular technology employs inertial sensors (accelerometers and gyroscopes) to continuously track the motion of the object.

The development of inexpensive solutions suitable for use with popular mobile devices has led to the development of numerous applications that use location based services (LBS). Such systems are used, for example to provide marketing information to the user, or to improve the security in public places. Due to poor signal quality, satellite systems cannot be used to provide reliable and accurate positioning services indoors. However, the widespread availability of Wi-Fi and Bluetooth Low Energy short range transmitters makes it possible to use their signals for positioning purposes [3].

[1] M.Sc., Lodz University of Technology, Institute of Electronics, Wólczańska 211/215, 90-924 Łódź, POLAND, tel. +48 42 631 26 26, fax: +48 42 636 22 38, e-mail: robert.kawecki@p.lodz.pl
[2] M.Sc., Lodz University of Technology, Institute of Electronics, Wólczańska 211/215, 90-924 Łódź, POLAND, tel. +48 42 631 26 26, fax: +48 42 636 22 38, e-mail: pawel.oleksy@p.lodz.pl
[3] M.Sc., Lodz University of Technology, Institute of Electronics, Wólczańska 211/215, 90-924 Łódź, POLAND, tel. +48 42 631 26 26, fax: +48 42 636 22 38, e-mail: jaroslaw.kawecki@p.lodz.pl
[4] M.Sc., Lodz University of Technology, Institute of Electronics, Wólczańska 211/215, 90-924 Łódź, POLAND, tel. +48 42 631 26 26, fax: +48 42 636 22 38, e-mail: krzysztof.piwowarczyk@p.lodz.pl
[5] D.Sc., Lodz University of Technology, Institute of Electronics, Wólczańska 211/215, 90-924 Łódź, POLAND, tel. +48 42 631 26 36, fax: +48 42 636 22 38, e-mail: slawomir.hausman@p.lodz.pl
[6] Ph.D., Lodz University of Technology, Institute of Electronics, Wólczańska 211/215, 90-924 Łódź, POLAND, tel. +48 42 631 26 12, fax: +48 42 636 22 38, e-mail: piotr.korbel@p.lodz.pl

One of the most commonly used radio signal parameters in localization systems is the Received Signal Strength (RSS). According to majority of standards and technical specifications, the Received Signal Strength Indicator (RSSI) has to be implemented in almost all commercially available radio systems. There are also available solutions using the signal propagation time measurements. These time-based systems usually offer significantly better positioning accuracy than those using RSSI, but they are more expensive, require dedicated infrastructure, and thus are not compatible with majority of popular user devices [1].

Knowing the distribution of the electromagnetic field intensity in the area in which the mobile terminal user moves [6], various methods [1] [4] [6] to estimate its location can be used. The most commonly used ones are correlation-based methods, where the positioning results depend on the similarity of the parameters of the signals recorded by the user terminal to some reference values stored in the system database. Using this class of methods requires calibration of the system (Figure 1). During this stage, at selected points of the operating environment of the system, called reference points, the values of parameters of locally observed signals are measured and stored in a reference database. During the essential phase, the parameters of the signals received by the terminal are compared with the results stored in the reference database. Then, basing on the results of this comparison, terminal location is determined by the most likely location. The following sections describe the general concept of operation of the localization system intended for mobile devices. Finally, they set out the proposed particle filter algorithm procedure. The following paragraphs present the results and verification of the operation of the developed system and concise summary as well.

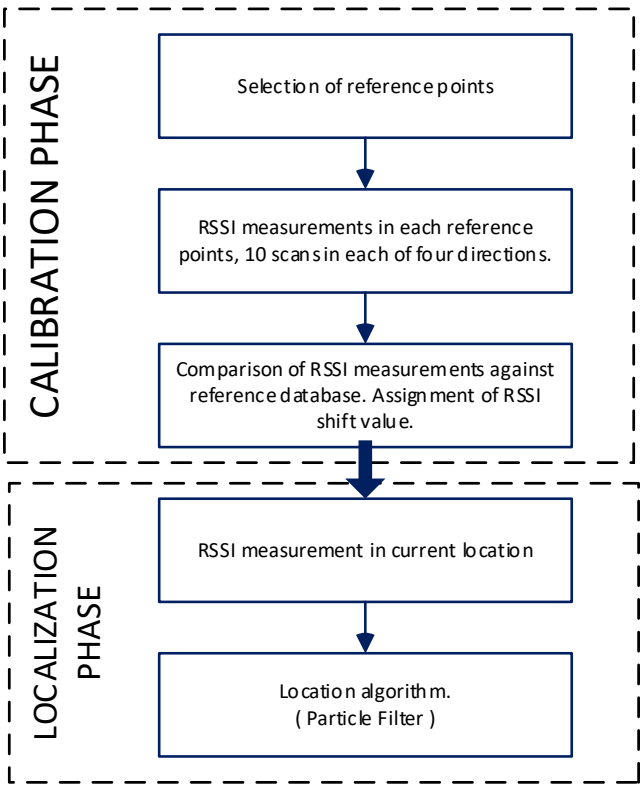


Figure. 1: The flowchart of the proposed indoor positioning algorithm.

2. The proposed localization system

System architecture considered in the article includes Bluetooth Low Energy transmitter network and user terminals equipped with dedicated mobile application implementing algorithms for determining the position of the user and presenting user location on the building plan.

Because of the growing popularity of mobile devices based on Android system, it was decided to use them as a platform, on which software aiming at localization of the users and visualisation of the estimated locations of the objects was implemented. The program requires cooperation with a terminal equipped with a Bluetooth Low Energy receiver and active access to the Internet.

The system works according to the following scenario:

1. Bluetooth transmitters broadcast messages containing unique identifiers, allowing to unambiguously determine the sources of the signals.
2. Application installed on the mobile device searches for Bluetooth tags in the vicinity, and registers their identification numbers and the signal strength (RSSI).
3. Based on the information about the observed identifiers and RSSI values, the implemented algorithm determines the estimate of the localization of the terminal. The procedure of updating the terminal localization estimate is performed every time RSSI values are different from the values recorded in previous algorithm iterations.

4. Particle filter algorithm

The localization of the terminal is based on the particle filter algorithm shown in Figure 2, and inspired by the version implemented in [2]. The algorithm comprises the following steps:

A) Loading reference data (interpolation of data measured at reference points in a 0.75 m grid) of spatial signal strength distributions from individual BLE tags;

B) Loading sequence of RSSI readings for the path passage (historical data for simulated mode or live measurements during application);

C) Low-pass filtering of measurement results to eliminate short-term deep signal loss;

D) Initialization of localization procedure – determination of initial position estimation of the terminal;

E) Initialization (for N particles) of coordinate vectors (X_P, y_P) for random values in range of building coordinates;

F) Comparison of measured power values received to the reference base value (power map received for individual BLE tags) – selection for all points satisfying the condition:

$$P_{xy} - P_{delta} < P_{refmap} < P_{xy} + P_{delta} \quad (1)$$

where:

P_{refmap} – power values in the reference base;

P_{xy} – measured power value received at the coordinates (x, y),

P_{delta} – maximum admissible deviation of the power values from the measured RSSI value.

Searching for points meeting criterion (1) is performed in the area limited to the last position estimate of $\pm xy_{delta}$, which is based on the assumption of the maximum speed of terminal movement;

G) Indicating the position estimate as a point with coordinates corresponding to the median of the points coordinates meeting criterion (1). Designating, on the basis of the calculated estimate, the displacement of x_{delta} and y_{delta} with reference to the position in the previous algorithm iteration.

H) Designation of a new set of particles for the position estimation:

$$X\_P = X\_P + x_{delta} + G_{delta} \quad (2a)$$

$$\mathrm{Y\_P} = \mathrm{Y\_P} + \mathrm{y_{delta}} + \mathrm{G_{delta}} \quad (2b)$$

where G_{delta} is the variance of the particles distribution model;

I) Assigning new weights to individual particles: i.e., particles that are outside the area of the location (e.g. outside the building) remain assigned weight 0 and particles whose positions do not match the conditional points (1) for at least two sources, also remain assigned a weighting factor of 0;

J) Normalization of particle weights – sum of weights of all particles should be equal 1;

K) Calculation of the degeneration coefficient and check, if it exceeds the experimentally chosen threshold value G_THR = 0.9;

L) Update of the position estimation – determination of the average value of the coordinates of the particles (x_P, y_P);

M) Calculation of position estimation error (affects only simulated mode).

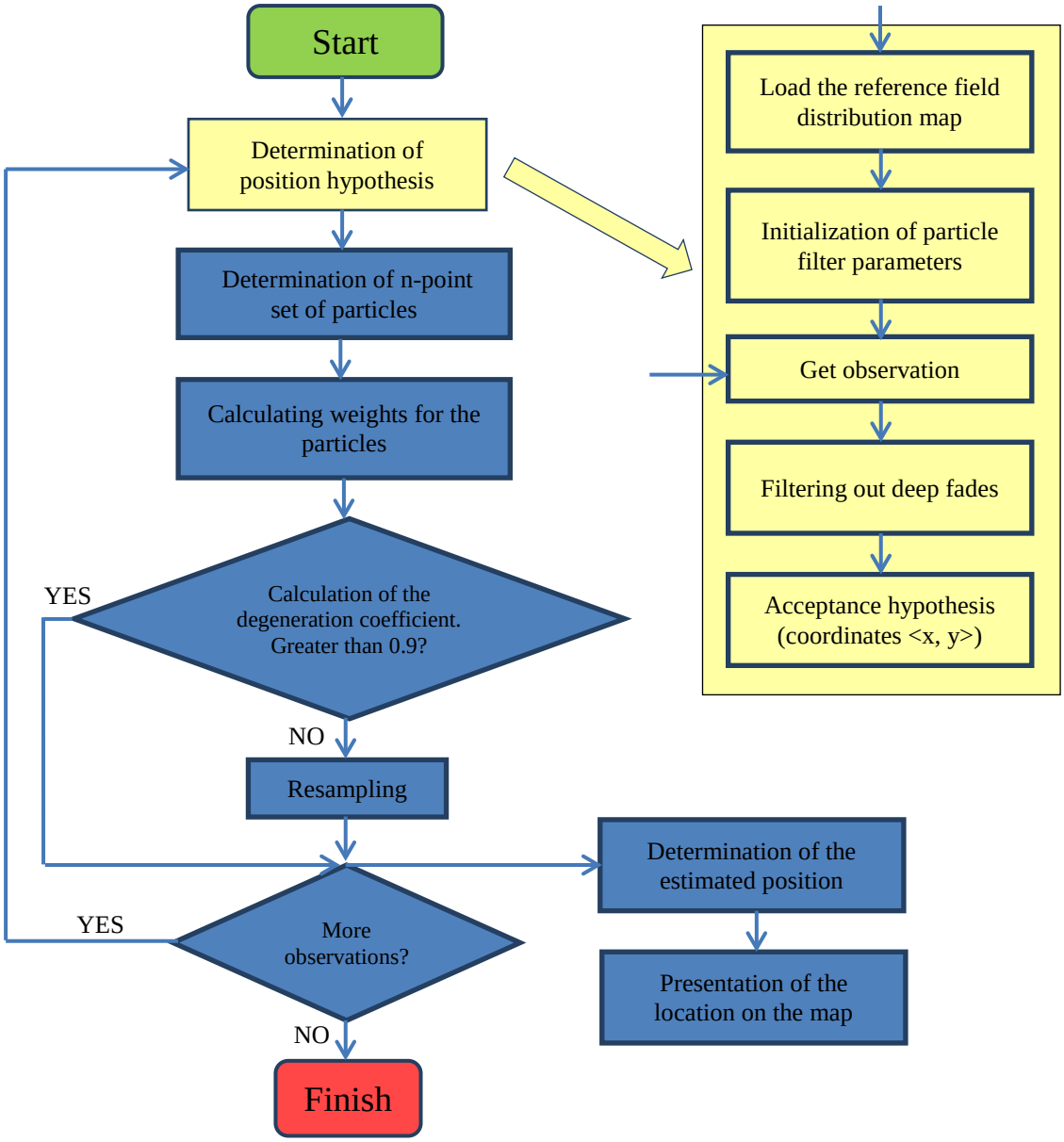


Figure. 2: Block diagram of the particle filtration algorithm.

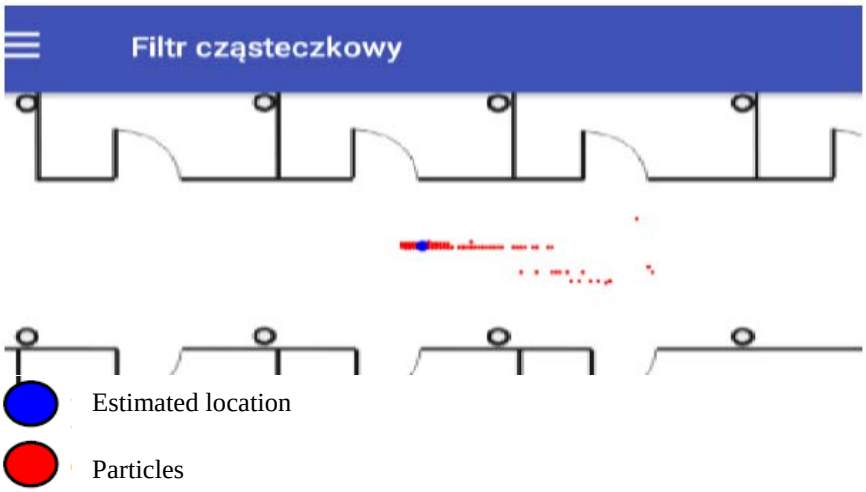


Figure. 3: Screenshot of the mobile application window with visualization of the performance of the particle filtration algorithm – view of the estimated terminal location and particles.

5. Test setup and experimental results

To test the properties of the proposed method, a series of experiments were carried out. As a source of signals, ten BLE tags from Kontakt.io, compatible with Bluetooth 4.0 Low Energy and low power consumption and a maximum transfer rate of about 1 Mbps are used. The operating range of the modules provided by the manufacturer does not exceed 100 m in the open area. For testing purposes, the tags were installed in an office building, in the premises of the Telecommunications Division of the Institute of Electronics. Dimensions of the area where the measurements were taken are approximately 40 × 18 meters. Figure 4 shows a section of a building plan with the marked location of the BLE beacons during testing and broadcasting identifiers.

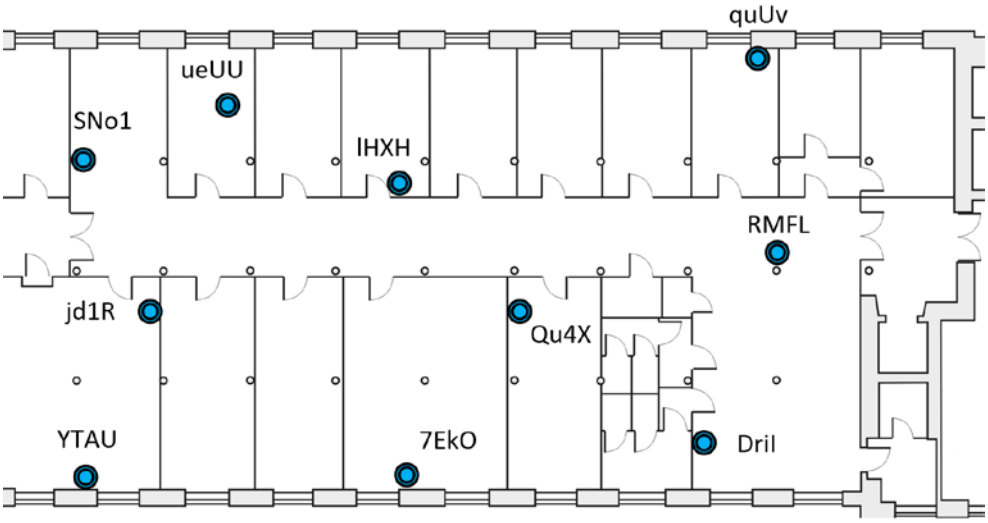


Figure. 4: Arrangement of BLE transmitters in the building

The module mounting height was about 1.4 m above the floor surface. The beacons were configured to broadcast location packets at maximum frequency, i.e., an advertising packet was sent every 20 ms. Transmission power was set at 0 dBm. Signal power measurements were performed using LG Nexus 5 terminals. The mobile application developed for the purpose of the study enables to record parameters such as RSSI power level from individual tags and time of measurement. The first stage of the research involved the preparation of a reference measurement database. For this purpose, the measurements were conducted in the points of the reference grid within the test premises. The distance between the neighbouring grid points was 0.75 m. In each of the points, 40 power measurements were obtained (10 measurements for four consecutive settings varying in terminal direction by 90 degrees, which is dictated by the need to take into account the non-isotropic radiation pattern of the antenna terminal and at the same time its dependence on the position relative to the body of the user) [7]. The results of the measurements lead to the creation of the reference radio map. Figure 5 shows an example of spatial distribution of the power received from one of the beacons.

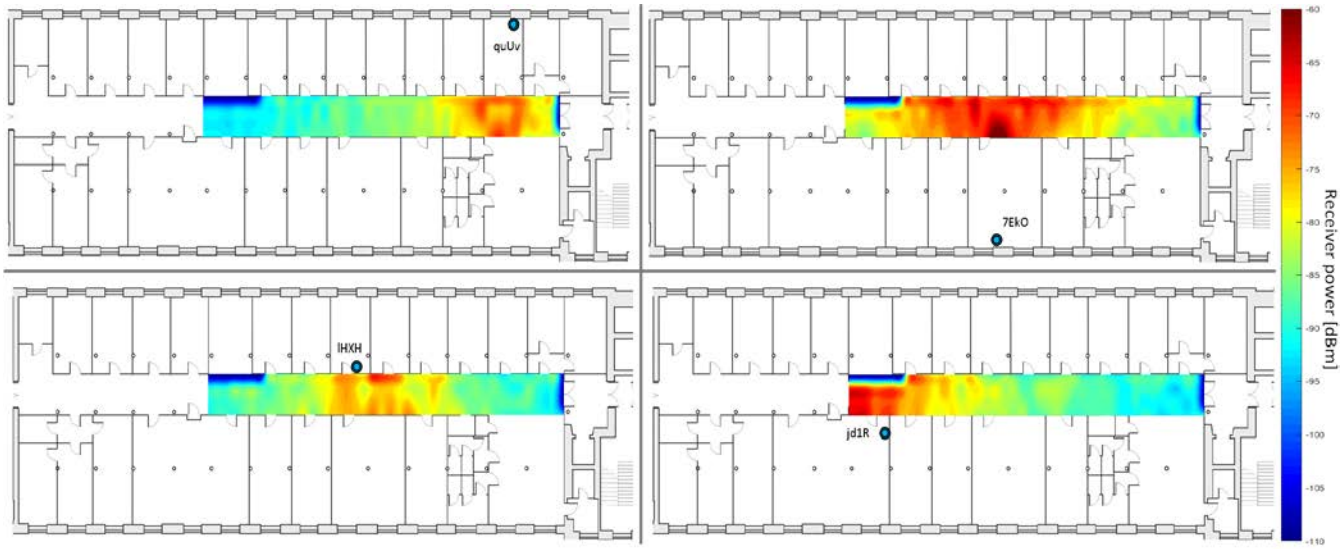


Figure. 5: Spatial distribution of received power measured in the corridor of the building for 4 example beacons.

The evaluation of the proposed method was divided into two main stages. In the first stage, an algorithm was implemented in a Matlab computational environment and studies were carried out using a set of data collected during reference measurements. The second stage of the method evaluation was the implementation of an Android-based mobile application algorithm and location accuracy testing using live signal power measurements. As a reference for the algorithm evaluation, the k nearest neighbours (k-NN) method was adopted. Although the reference method does not belong to the same class of methods (it does not take into account the history of the measurements), it is among the most commonly cited ones in the literature on indoor localization systems [5] [6], and establishes a good reference for evaluation of benefits coming from the use of more complicated methods.

During the simulation tests, three examples of terminal movement routes were considered, illustrated in Figure 6 (routes marked with letters A, B, and C). Subsequently, power readings corresponding to the points along the certain transition paths were selected from the measurement data set. The remaining data served as the reference set used by the localization algorithm. These experiments allowed to determine the parameters of the algorithm before its implementation in the mobile application.

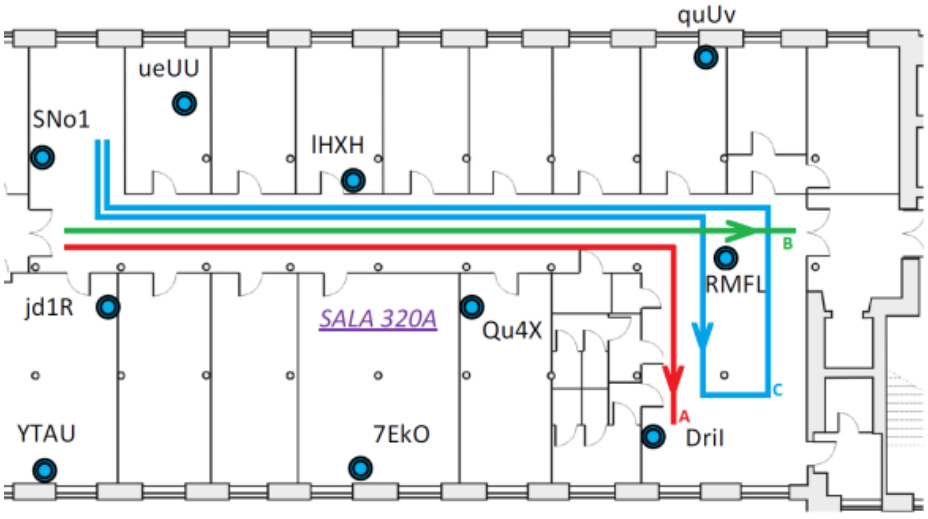


Figure. 6: The user test path.

The accuracy test of the method using the current RSSI power measurements was performed for the test paths as in the case of simulative tests. In order to facilitate the assessment of the accuracy offered by the method, measurements using a mobile application were used to validate the current location. Figure 7 shows an example route for comparing the results of localization using a mobile application with the results obtained using the k-NN classifier. Table 1 summarizes the comparative analysis of the accuracy of the proposed method with k-NN (k = 15) classifier for the considered test scenarios.

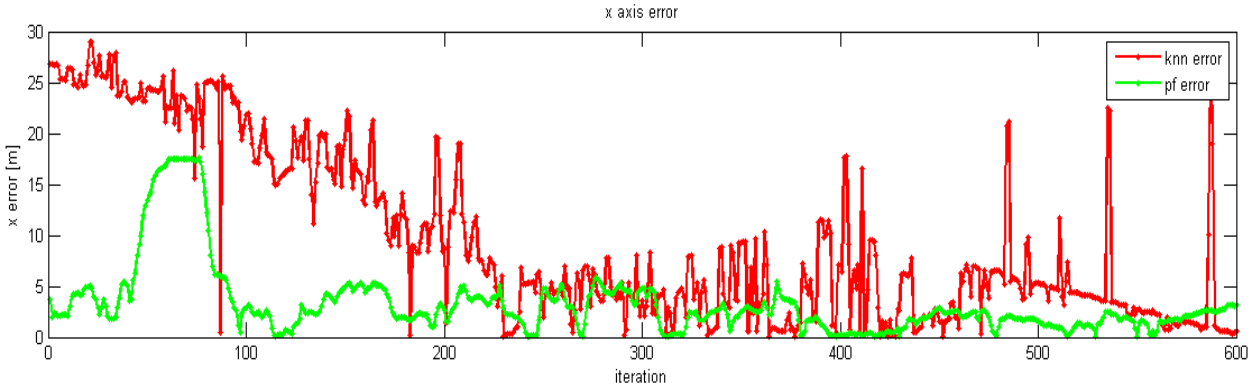


Figure. 7: Estimation error of the location of the x and y coordinates for the particle filter algorithm and k-NN for path B.

Table. 1 Comparison of positioning accuracy for the particle filtration algorithm (PF) and the k-NN classifier.

Parameter	*PF*	*k-NN*
Average error	3.37	8.40
Maximum error for 50 % of all cases [m]	2.43	4.19
Maximum error for 75 % of all cases [m]	3.95	12.02
Maximum error for 90 % of all cases [m]	5.35	25.46

6. Summary

This article presents the use of a particle filtration algorithm for localization inside buildings. The system, using the particle filter and reference maps of electromagnetic field intensity distribution, allows to localize the user on the basis of RSSI readings recorded with a mobile device. The results of the proposed algebraic localization using particle filtration were compared with the popular k-NN approach of nearest neighbours. The particle algorithm enables the improvement of positioning accuracy of a user equipped with a mobile terminal, compared to the k-NN method. It has been verified that, despite the high complexity of the computational localization method using the particle filter algorithm, it is possible to implement it on mobile devices, allowing real-time positioning.

Further work on the development of a particle filtration algorithm for mobile devices involves seeking modifications designed to reduce computational complexity while preserving the positioning accuracy offered.

References

1. Gu Y., Lo A., Niemegeers I. *"A Survey of Indoor Positioning Systems for Wireless Personal Networks"*. IEEE Communications Surveys & Tutorials, No. 1: 13-32, 2009
2. Barański P., Polańczyk M., Strumiłło P. *"Fusion of data from inertial sensors, raster maps and GPS for estimation of pedestrian geographic location in urban terrain"*. Metrology and Measurement Systems, Vol. 18, No. 1: 145-158, 2011
3. Faragher R., Harle R. *"Location Fingerprinting with Bluetooth Low Energy Beacons"*. IEEE Journal on Selected Areas in Communications, vol. 33, no. 11: 2418-2428, 2015
4. Bahl, P., Padmanabhan, V.N. *"RADAR: an in-building RF-based user location and tracking system"*. INFOCOM 2000, Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, Proceedings IEEE, vol.2: 775-784, 2000
5. Hui Liu, Darabi, H., Banerjee, P., Jing Liu. *„Survey of Wireless Indoor Positioning Techniques and Systems"*. 2007. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol.37, no.6: 1067-1080, 2007
6. Januszkiewicz Ł., Kawecki R *„Zastosowanie modelu Multi-Wall do wspomagania projektowania systemu lokalizacji we wnętrzach budynków"*. Przegląd telekomunikacyjny i Wiadomości telekomunikacyjne, nr 6/2016: 373-376, 2016
7. Wawrzyniak P., Hausman S., Korbel P., *"Area based indoor tracking algorithm based on sequence detection and maximum likelihood metrics"*. EuCAP 2016, 10th European Conference on Antennas and Propagation (EuCAP), Davos, 2016

Piotr KRUPSKI[1]

The effectiveness of plasma active cooling system designed for atmospheric pressure *plasma jet* reactor

The paper presents an active plasma cooling system which is designed for decreasing the temperature of plasma produced in atmospheric pressure plasma jet reactors. A special focus was put on a nozzle construction, which has two functions: zero-potential electrode and evaporator in cooling circulation system. The nozzle, in this case, has two functions: zero-potential electrode and evaporator in cooling circulation system. The plasma produced in those reactors with the described system has significantly lower temperature than the one produced in jets without a cooling unit. The featured system is characterized by non-thermal plasma which has a low temperature level.The temperature of cooled plasma has been proved by measurements with the use of thermal imaging camera and temperature sensors. The solution has been developed exclusively by the author and has currently been subject to a patent procedure, in the course of which it has received a polish patent application number P.422159.

Introduction

Among the applications of nothermal plasma in biomedicine, there occurs the subject of living structures treatment in the purpose of influencing the pathogens existing on them. An important parameter, in this case, is the level of plasma temperature which is required to obtain the needed properties of plasma decontamination, and which also influences the stability ofdischarge products in gas, such as ozone and active gas oxides. At the same time, together which decreasing temperatures of the produced plasma, the negative thermal influence on living tissues is being minimalised. An increased temperature exposes both of the sensitive hydrogen and hydrophobic bonds, causing the protein to lose its spatial structure which is relevant for its function played in living tissues. An exemplary plasma jet reactor is shown in figure 1.

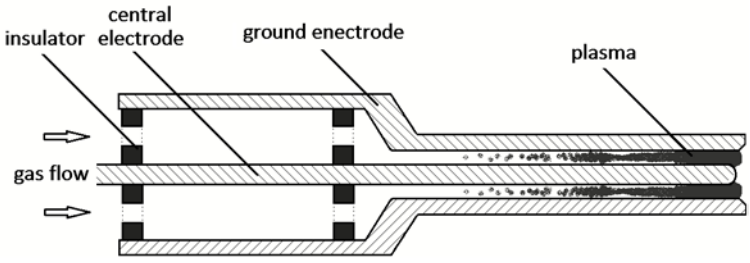


Fig. 1. A cross-section of an exemplary *plasma jet*.

[1] Ph.D., Lublin University of Technology, Institute of Electrical Engineering and Electrotechnology, Nadbystrzycka str. 38 A, 20-618 Lublin, POLAND, tel. 81 538-42-87, e-mail: piotr-jl@o2.pl

1. Temperatures of plasma and temperature decreasing methods

The use of the term of temperature is allowed only when the distribution of energy of kinetic particles composing a specific configuration can be described by a Maxwell distribution. In the state of thermodynamic equilibrium in plasma, the energy is evenly distributed into all the degrees of freedom of the particles and, eventually, it leads to their mean kinetic energy values being alike [1]. This state is obtained only very rarely as it excludes any directed movement of particles. Such a movement can be caused for example via the occurrence of an external electrical field. Far more often does the situation occur in which plasma only locally exhibits the state of thermodynamic equilibrium. In this state, under the influence of external electrical field, electrons acquire energies that are far greater than the ionic and neutral particle energy. Due to the fact that the exchange of energy between the electrons and ions takes place more slowly than between particles witch comparable size and mass, in low-condensed plasma there might persist a state in which Maxwell distribution will be established for electrons and ions separately, while the mean values of both distributions will not be equal. Therefore, the measurement of some values linked with the term of temperature, such as electrons, ions, and the mean temperature of gas, is possible. Regarding the different properties of each of the mentioned particles, in order to measure their temperature, various methods are used. [2, 3, 4]. Similarly, to determine the level of temperature in plasma jet reactors, it is possible to use different methods.

There is also a number of methods used for decreasing temperature in AAPJ (*Atmospheric-Pressure Plasma Jet*). The easiest way to do that is to increase the stream of working gas which flows through the reactor. A significant change of temperature occurs when the flow acquires turbulent character. A different way is to provide working gases in liquid state, so that the phase transition could happen only in the plasma generator.

2. Active temperature decreasing system

This paper shows an active cooling system involving process gas and plasma area. Non-thermal plasma properties are obtained inside a specially-constructed jet which is made from copper and has two functions at the same time: zero-potential electrode and evaporator in the cooling system. The cross section of the cooling jet is presented on the figure 2.

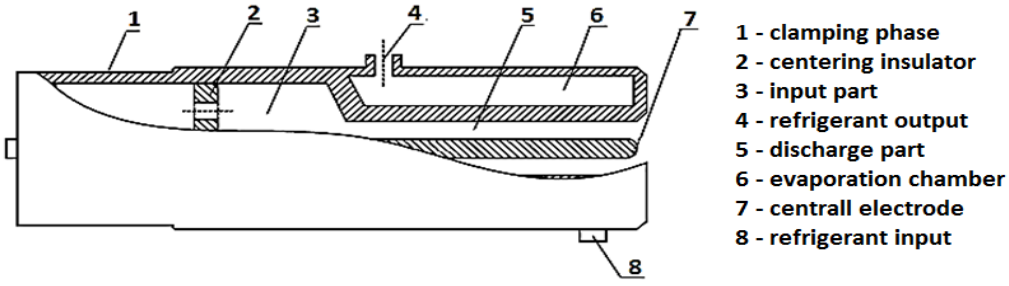


Fig. 2. A cross-section of the cooling jet.

The evaporation chamber is connected to a closed Linde's refrigeration circuit (Fig.3). In this system the plasma and the refrigerant are not miscible, they are however, powerfully thermally conjoined. Thus, heat migrates from plasma to over-heated steams of refrigerant.

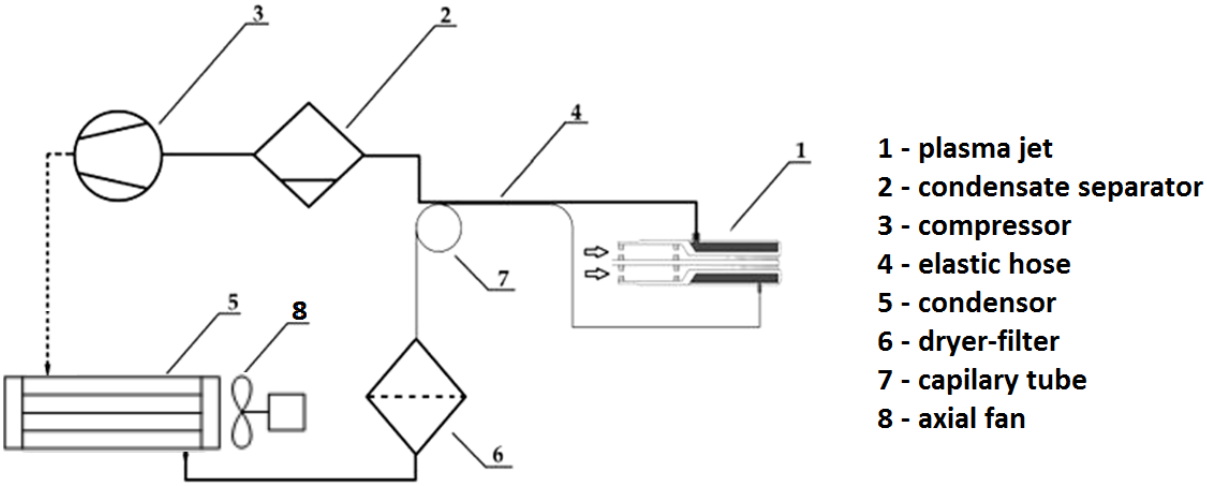
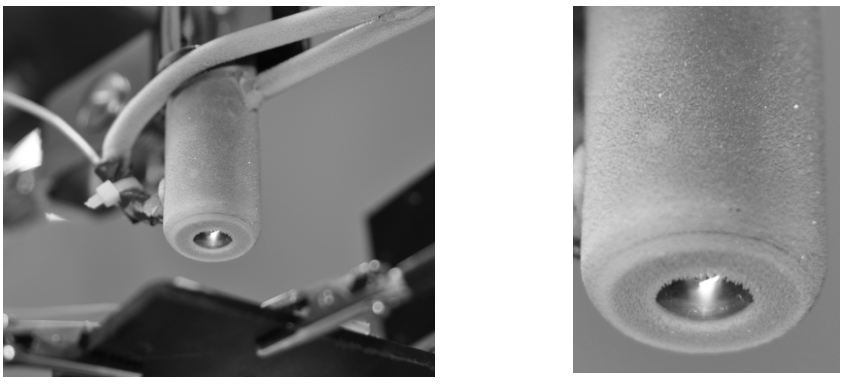


Fig. 3. Diagram of refrigeration cycle designated for *plasma jet*.

3. Methodology and results

The experiment studied the efficiency of reducing plasma temperature during the treatment of polymer surface. The reactor worked with helium and nitrogen gases with volume streams 7.1 m^3/min and 4.1 m^3/min. Sinewave voltage of 14.84 MHz was provided to the central electrode. A 50x50x2mm ABS (*Acrylonitrile Butadiene Styrene*) plate with a temperature sensor on its surface was the object of plasma treatment. The plate was clipped to a base with alligator clips. The distance between the tip of plasma output and the plate surface was equal to 1 inch. This situation is presented on figure 4.



a) b)

Fig. 4. a) *Plasma jet* directed perpendicularly to the polymer plate, b) hoarfrost on plasma jet's surface.

After providing the power of 45 W, the reactor worked until reaching a fixed high temperature level. The next step was switching on the cooling system and injecting refrigerant (*1,1,1,2-tetrafluoroethane*) into the evaporation chamber. Relative pressure in the chamber was equal to -5 psi. After achieving that, the cooler worked until the temperature was set on the lowest possible level. Then, the cooler was switched off, but the whole system was still working. The temperature was measured on both- the surface of the plate and the jet. The results are shown on figure 5.

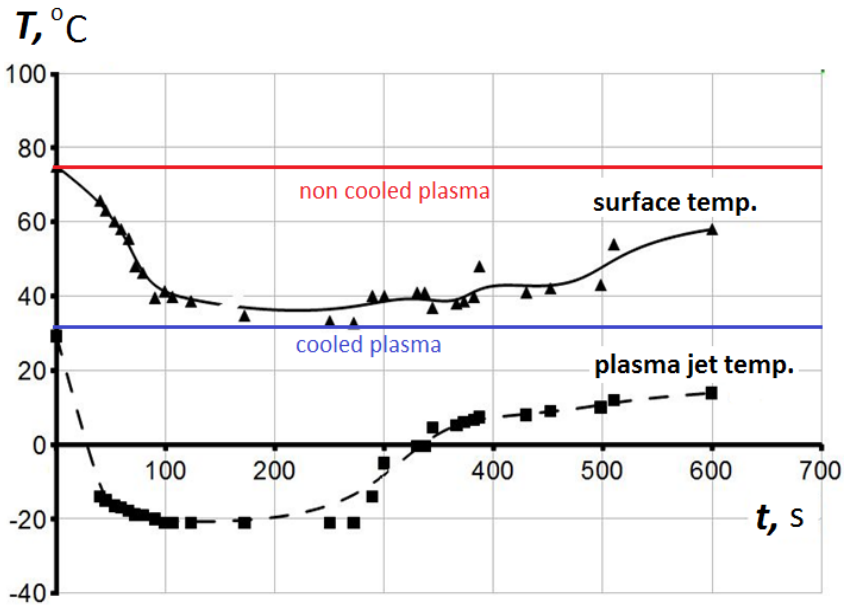
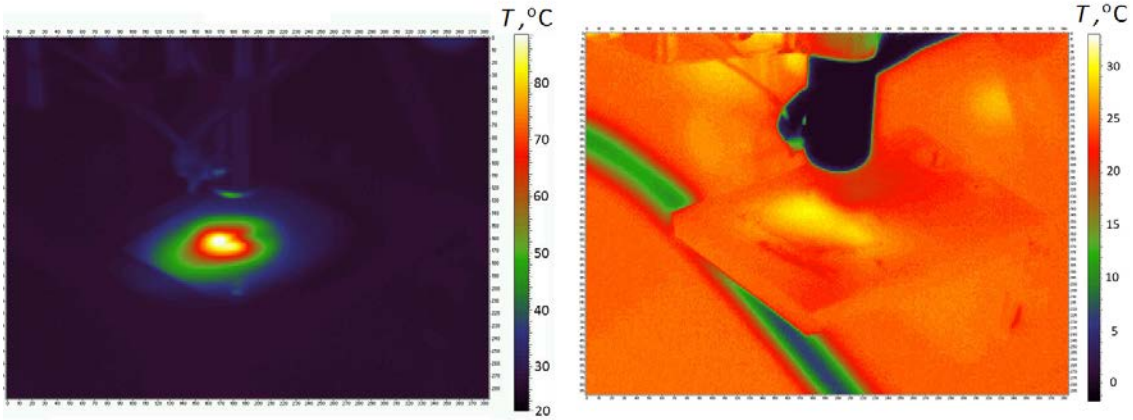


Fig. 5. Temperatures of plasma jet and treated surface during the experiment in time .

In the second part of the experiment all of the treatment conditions were repeated, but the supply power was reduced to 40 W. Thermal imaging was used to show the difference in the distribution of temperatures of both- the cooled and the non-cooled plasma in the state of stability (fig. 6).



a) b)

Fig. 6. Thermal images of laboratory work station during plasma treatment a) without cooling, b) witch active cooling system.

4. Remarks and conclusion

The object of this paper was to show the efficiency of active plasma jet cooling system. The experiment was made in the conditions involving nitrogen and helium gases. The cross-section of the jet was provided as well as its whole active cooling system.

The results were collected in two steps (for two different power values) and then presented in both: a graph and as thermal images.

The research proved the cooling system to be highly efficient which is visible on both- the graph and thermal images. The temperature of the treated plate was decreased up to 42 degrees Celsius and the result is unobtainable for any other known cooling methods.The proposed solution shows an applicatory potential not only in medical and biotechnological fields, but also during the treatment of high-temperature-susceptible materials which are represented by a low threshold of thermal plasticisation.

Due to the innovative character of the solution, the temperature can be lowered very effectively in the way that so far has not been encountered in any literature reports. Therefore, Lublin University of Technology, applied for a patent in Patent Office of the Republic of Poland. This application has been made for this particular solution. The title is: *The plasma jet cooling circuit and method of plasma cooling.* The application received the following number: P.422159.

References

1. Terebun P., Krupski P., Kwiatkowski M., Samoń R., Diatczyk J., Pawłat J., Stryczewska H.D.: *Wybrane metody pomiaru temperatury plazmy w reaktorach typu jet*, INFORMATYKA, AUTOMATYKA, POMIARY W GOSPODARCE I OCHRONIE ŚRODOWISKA, Tom 3, 2015,
2. Diatczyk J., Stryczewska H. D., Komarzyniec G.*: Modeling of the Temperature Distrbution in Arc Discharge Plasma Reactor,* JOURNAL OF ADVANCED OXIDATION TECHNOLOGIES, Vol. 9, no 2, July 31, 2006,
3. Mahmood S., Shaikh Nek M., Kalyar M. A., Rafiq M., Piracha N. K., Baig M. A.: *Measurements of electron density, temperature and photoionization cross sections of the excited states of neon in a discharge plasma*. JOURNAL OF QUANTITATIVE SPECTROSCOPY AND RADIATIVE TRANSFER, 110 (17), 2009,
4. Stryczewska H. D., Diatczyk J., Pawłat J.: *Temperature distribution in the gliding arc discharge chamber*, JOURNAL OF ADVANCED OXIDATION TECHNOLOGIES, Volume 14, Number 2, July 2011.

Krzysztof TOMCZYK[1]

Approximate Functions for Determining the Integral Square Error Generated by Voltage Output Accelerometer

Abstract This paper presents some approximation functions that allow easily determining the integral square error generated by a voltage output accelerometer in response to an exciting signal with one constraint concerning the amplitude. Both the mechanical construction and mathematical relations of the accelerometer are also presented. The procedure for determining the integral square error is described and the values of this error as a function of the time for six values of the damping ratio are determined. Then the coefficients of the linear function which approximates the obtained errors are calculated. Nonlinear equations reflecting the relation between the coefficients of the linear function and the damping ratio are determined on the assumption of a constant value of the voltage sensitivity and cut-off frequency.

1. Introduction

The voltage output accelerometers belong to a group of sensors intended for non-electrical measurements. They play a very important role in many areas of technology, e.g. in tests of machines' vibrations, in aviation, and recently in mobile electronic devices [1].
Accelerometers are calibrated by measuring their frequency characteristics in response to a harmonic excitation. This calibration is most often used to determine the value of the voltage sensitivity, in accordance with legal regulations. In this case, the results of the calibration are specified only by the dynamic properties of the sensor [2,3]. However, in measuring dynamic quantities, we mainly deal with undetermined measurement signals which are unpredictable. The dynamic error at the output of the accelerometer, as for other sensors intended for measuring signals changing over time, depends not only on the dynamic properties of the sensor but also on the shape of the input signal. This is due to the properties of convolution, which is one of the basic mathematical operations in the field of signal processing and is an integral relation that joins the sensors output signal with both its impulse response and the input signal [4,5].
Taking into account that the signal at the output of the sensor changes over time, it is indispensable to use a quantitative error criterion. For accelerometers, it is convenient to use the integral-square criterion, the most popular measure of the energy. Thus, depending on the shape of the input signal and for constant dynamic properties of the sensor, one can obtain significant differences in the value of the integral-square error.
In [4,5] the calibration procedures based on maximum dynamic errors were discussed in detail. According to these procedures and both for a given mathematical model of the sensor and in relation to a chosen model of the standard, the input signal maximizing the error at the output is determined. There are considered those signals constrained only in magnitude (rectangular shape) or simultaneously in amplitude and rate of change (trapezoidal shape) [6]. Any other signal that is contained in the above constraints will generate an error less then or at

[1] Ph.D., Cracow University of Technology, Warszawska 24, 31-155 Krakow, POLAND,
tel. ++48 12 628 25 48, e-mail: k.tomczyk@cyfronet.pl

most equal to this maximum [4]. In [7,8] a single value of the error were determined based on a complicated algorithm applied for different types of accelerometers. Finally, the significant uncertainties of the results were obtained.
The study of the impact of a change in the damping ratio on the integral square error is presented in this paper. An approximate functions, which have not been published so far, are proposed here. These functions have been obtained based on the algorithm for determining signals constrained only in amplitude, that maximize the error. This algorithm was recently discussed in detail in [9] and was developed on the basis of the theoretical solutions presented in [10–12].

2. Model of Voltage Output Accelerometer

The mechanical construction of the voltage output accelerometer is shown in Fig. 1.

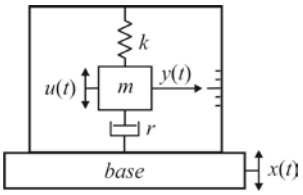


Fig. 1. Mechanical construction of the voltage output accelerometer

The construction above is represented by the following differential equation

$$m\ddot{u}(t) + r\dot{y}(t) + ky(t) = 0, \tag{1}$$

where: $u(t)$, $y(t)$, $x(t)$, $m[kg]$, $r[kg/s]$, $k[N/m]$, $m\ddot{u}(t)$, $r\dot{y}(t)$ and $ky(t)$ are the absolute mass displacement, relative mass displacement, vibration (excitation), seismic mass, dumping coefficient, spring constant, moment of inertia, moment of dumping and moment of elasticity, respectively.
Taking into account in (1) the following relation

$$u(t) = x(t) + y(t), \tag{2}$$

we obtain the differential equation as below

$$m\ddot{y}(t) + r\dot{y}(t) + ky(t) = -m\ddot{x}(t). \tag{3}$$

This equation describes the mathematical model of the voltage output accelerometer, represented by

$$K_a(s) = \frac{-S\,\omega_0^2}{s^2 + 2\beta\omega_0 s + \omega_0^2}, \tag{4}$$

where S denotes the voltage sensitivity, while

$$\omega_0 = 2\pi f_0 = \sqrt{\frac{k}{m}} \tag{5}$$

and

$$\beta = \frac{r}{2\sqrt{km}} \tag{6}$$

are the non-damped natural frequency and the damping ratio.

3. Determining the Integral Square Error

The integral square error I_2 is represented by

$$I_2 = \int_0^T y(t)^2 dt = \int_0^T \left[\int_0^T \mathcal{K}(t,\tau)\, x(\tau)\, d\tau \right] x(t)\, dt, \tag{7}$$

where $x(t)$ and $y(t)$ are the input and output signals, while

$$\mathcal{K}(t,\tau) = \int_0^T k(t,v)\, k(v,\tau) dv \tag{8}$$

and

$$k(t) = k_a(t) - k_s(t) \tag{9}$$

is the difference between the impulse responses of the accelerometer and its standard for $t = 0, \Delta_t, \ldots, T,$ where Δ_t is the discretization step of time. The time T is the period for which the error is determined.

The impulse responses are calculated by

$$k_a(t) = \mathcal{L}^{-1} K_a(s) \tag{10}$$

and

$$k_s(t) = \mathcal{L}^{-1} K_s(s) = 2cf_c Sinc[2\pi\, f_c(t - t_0)], \tag{11}$$

where $c,$ t_0 and f_c are the amplification coefficient, time delay and cut-off frequency, respectively.

The algorithm for determining the signal $x_0(t)$ constrained in amplitude and maximizing the error is as follows:

1. Calculate the function $\mathcal{K}(t,\tau)$ based on (8).
2. Assume the initial input signal

$$x^0(t) = A \cdot sign[\mathcal{K}(t,\tau)], \tag{12}$$

where A denotes the amplitude constraint of the signal.

3. Determine the new signal $x^{i+1}(t)$ based on $x^i(t)$ and by utilizing the iterative algorithm

$$x^{i+1}(t) = A \cdot sign\left[\int_0^T x^i(v)\, \mathcal{K}(v,t)\, dv \right],\ i = 0,\ 1,\ 2, \ldots \tag{13}$$

until the differences between the successive time-switchings of the signals $x^{i+1}(t)$ and $x^i(t)$ are no greater than the discretization step Δ_t assumed in the calculations. Then, we assume that $x^{i+1}(t) = x_0(t)$ is the maximizing signal.

The integral-square error is calculated by utilizing the signal $x_0(t),$ as follows [10-12]

$$I_2(x_0) = \int_0^T \left[\int_0^t k(t-\tau)\, x_0(\tau) d\tau \right]^2 dt. \tag{14}$$

The values of integral square error $I_2(x_0),$ calculated for T from 0 to 10ms. with step Δ_T equal to 1ms. and for the damping ratio β from 0.1 to 0.6 with step Δ_β equal to 0.1 are reported in Table 1.

The following values of the parameters which occur in (4), (5) and (11) are assumed: $S = 1\,V/(ms^{-2}),$ $f_c = 1kHz,$ $c = S$ and $t_0 = T/2.$

For the assumed cut-off frequency f_c and for different damping ratios $\beta,$ the non-damped natural frequency f_0 is calculated by solving the equation

$$\lambda = \frac{S}{\sqrt{[1-(f_c / f_0)^2]^2 + 4\beta^2 (f_c / f_0)^2}} \tag{15}$$

which results from the amplitude characteristic of accelerometer, where

$$\lambda = S + \frac{S}{100\%} \cdot \varepsilon \tag{16}$$

and ε denotes the tolerance of the amplitude characteristic. The value of this tolerance was assumed as equal to 10% for the voltage sensitivity obtained for zero frequency value. Then λ is equal to $1.1\,V/(ms^{-2})$.

As the solution for the frequency f_0, the lowest positive values of Eq. (15) were taken. These values are given in Table 1 below β.

The maximum value of T exceeds the steady-state of the impulse response $k(t)$ by about ten times.

Table 1. Integral square error $I_2(x_0)$ vs. time T

	$I_2(x_0) = f(T)$ $[Vs^2]$				$I_2(x_0) = f(T)$ $[Vs^2]$		
T [ms]	$\beta = 0.1$ ($f_0 = 3280Hz$)	$\beta = 0.2$ ($f_0 = 3167Hz$)	$\beta = 0.3$ ($f_0 = 2966Hz$)	T [ms]	$\beta = 0.4$ ($f_0 = 2649Hz$)	$\beta = 0.5$ ($f_0 = 2115Hz$)	$\beta = 0.6$ ($f_0 = 1092Hz$)
0	0	0	0	0	0	0	0
1	$7.703 \cdot 10^{-3}$	$3.159 \cdot 10^{-3}$	$1.593 \cdot 10^{-3}$	1	$8.116 \cdot 10^{-4}$	$5.545 \cdot 10^{-4}$	$1.937 \cdot 10^{-4}$
2	0.025	$7.854 \cdot 10^{-3}$	$3.603 \cdot 10^{-3}$	2	$1.627 \cdot 10^{-3}$	$1.273 \cdot 10^{-3}$	$4.412 \cdot 10^{-4}$
3	0.044	0.013	$5.567 \cdot 10^{-3}$	3	$2.665 \cdot 10^{-3}$	$1.983 \cdot 10^{-3}$	$6.913 \cdot 10^{-4}$
4	0.064	0.018	$7.539 \cdot 10^{-3}$	4	$3.530 \cdot 10^{-3}$	$2.598 \cdot 10^{-3}$	$9.319 \cdot 10^{-4}$
5	0.083	0.022	$9.482 \cdot 10^{-3}$	5	$4.327 \cdot 10^{-3}$	$3.358 \cdot 10^{-3}$	$1.187 \cdot 10^{-3}$
6	0.103	0.027	0.012	6	$5.299 \cdot 10^{-3}$	$4.068 \cdot 10^{-3}$	$1.440 \cdot 10^{-3}$
7	0.122	0.032	0.014	7	$6.366 \cdot 10^{-3}$	$4.706 \cdot 10^{-3}$	$1.671 \cdot 10^{-3}$
8	0.142	0.037	0.016	8	$7.275 \cdot 10^{-3}$	$5.445 \cdot 10^{-3}$	$1.928 \cdot 10^{-3}$
9	0.161	0.042	0.018	9	$8.177 \cdot 10^{-3}$	$6.204 \cdot 10^{-3}$	$2.150 \cdot 10^{-3}$
10	0.180	0.046	0.02	10	$9.035 \cdot 10^{-3}$	$6.873 \cdot 10^{-3}$	$2.402 \cdot 10^{-3}$

Analysis of the error values from Table 1 indicates that their relations with the time are very close to linear.

4. Determining the Approximate Functions

Let the time and error reported in Table 1 be represented by vectors

$$\mathbf{X} = [T_0, T_1, \ldots, T_{N-1}]^T \tag{17}$$

and

$$\mathbf{Y} = [I_2(x_0)_0, I_2(x_0)_1, \ldots, I_2(x_0)_{N-1}]^T, \tag{18}$$

where N denotes the number of determined errors.

Let us approximate the error by the linear function

$$y_a(T) = a_0 + a_1 T, \tag{19}$$

where

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \mathbf{A} = (\mathbf{\Phi}^T \mathbf{\Phi})^{-1} \mathbf{\Phi}^T \mathbf{Y} \tag{20}$$

and

$$\mathbf{\Phi} = \begin{bmatrix} 1 & T_0 \\ 1 & T_1 \\ \vdots & \vdots \\ 1 & T_{N-1} \end{bmatrix}. \tag{21}$$

An evaluation of the standard uncertainty of the approximation function is fulfilled by [13]

$$u_{\mathrm{A}}(y_a) = \sqrt{Sy_a^2} = \sqrt{\frac{(\mathbf{\Phi A} - \mathbf{Y})^T (\mathbf{\Phi A} - \mathbf{Y})}{N-2}} = \sqrt{\frac{\sum_{i=0}^{N-1} | y_a(a_0, a_1, T_i) - I_2(x_0)_i |^2}{N-2}}. \tag{22}$$

The standard uncertainty for the coefficients a_0 and a_1 is determined by

$$u_{\mathrm{A}}(a_i) = u_{\mathrm{A}}(y_a) \cdot \sqrt{D_{i,i}}, \quad i = 0,1, \tag{23}$$

where

$$\mathbf{D} = (\mathbf{\Phi}^T \mathbf{\Phi})^{-1}. \tag{24}$$

The relative uncertainty for the coefficients a_0 and a_1 is defined, as follows

$$\delta(a_0) = \frac{u_{\mathrm{A}}(a_0)}{|a_0|} \cdot 100\%, \quad \delta(a_1) = \frac{u_{\mathrm{A}}(a_1)}{|a_1|} \cdot 100\%. \tag{25}$$

The values of a_0 and a_1 as well as the associated uncertainties $u_{\mathrm{A}}(y_a)$, $u_{\mathrm{A}}(a_0)$ and $u_{\mathrm{A}}(a_1)$, related to the damping ratio β, are reported in Table 2. There are also the relative uncertainties $\delta(a_0)$ and $\delta(a_1)$ as well as the slope coefficients α, calculated by

$$\alpha = \tan^{-1}(a_1). \tag{26}$$

Table 2. Parameters and associated uncertainties for linear function

β	0.1	0.2	0.3	0.4	0.5	0.6
a_0	$-8.855 \cdot 10^{-3}$	$-1.081 \cdot 10^{-3}$	$-3.899 \cdot 10^{-4}$	$-1.148 \cdot 10^{-4}$	$-1.032 \cdot 10^{-4}$	$-3.140 \cdot 10^{-5}$
a_1	$1.857 \cdot 10^{-2}$	$4.725 \cdot 10^{-3}$	$2.038 \cdot 10^{-3}$	$9.159 \cdot 10^{-4}$	$6.945 \cdot 10^{-4}$	$2.433 \cdot 10^{-4}$
$u_{\mathrm{A}}(y_a)$	$3.731 \cdot 10^{-3}$	$5.266 \cdot 10^{-4}$	$2.079 \cdot 10^{-4}$	$8.138 \cdot 10^{-5}$	$5.292 \cdot 10^{-5}$	$1.505 \cdot 10^{-5}$
$u_{\mathrm{A}}(a_0)$	$2.105 \cdot 10^{-3}$	$2.970 \cdot 10^{-4}$	$1.173 \cdot 10^{-4}$	$4.590 \cdot 10^{-5}$	$2.985 \cdot 10^{-5}$	$8.492 \cdot 10^{-6}$
$u_{\mathrm{A}}(a_1)$	$3.558 \cdot 10^{-4}$	$5.021 \cdot 10^{-5}$	$1.982 \cdot 10^{-5}$	$7.759 \cdot 10^{-6}$	$5.046 \cdot 10^{-6}$	$1.435 \cdot 10^{-6}$
$\delta(a_0)$ [%]	23.8	27.5	30.1	40.0	28.9	27.0
$\delta(a_1)$ [%]	1.92	1.06	0.973	0.847	0.727	0.590
α [deg.]	1.06	0.271	0.117	0.0525	0.0398	0.0139

Fig. 2 shows a linear relation between the error $I_2(x_0)$ and the damping ratio β for the range from 0.1 to 0.6. From this figure, it is easy to observe that the smallest inclination is for the measurement points for $\beta = 0.1$ and the largest for $\beta = 0.6$. It is also necessary to note that for T from the range between 0 and approximately 1ms, the relations above are not linear. This is due to the impulse response not being steady-state in this interval.

Let us determine below the relation between the coefficients a_0 and a_1 of the linear function (19) and the damping ratio β. These relation can be easily approximated by the functions

$$a_0(\beta) = \gamma_0 \cdot \beta^{\gamma 1} \tag{27}$$

and

$$a_1(\beta) = \gamma_0 \cdot \beta^{-\gamma 1}. \tag{28}$$

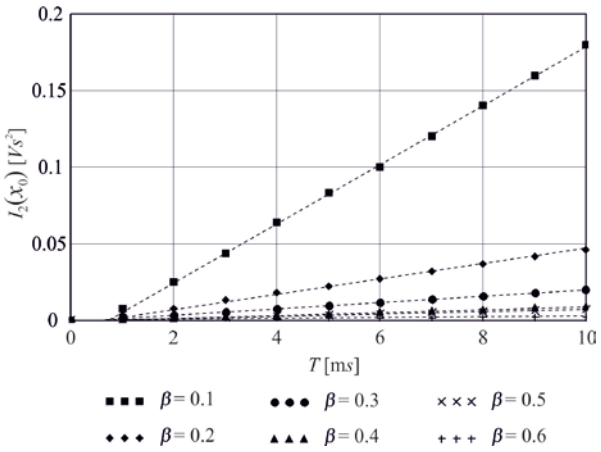


Fig. 2. Integral-square error $I_2(x_0)$ in relation to time T

The coefficients γ_0 and γ_1 that provide the best approximation can be easily determined by linearization of the equations above. For this purpose, a bilateral logarithmic transformation should be made, which in effect gives

$$\log[a_0(\beta)] = \log(\gamma_0) + \gamma_1 \log(\beta), \quad \log[a_1(\beta)] = \log(\gamma_0) - \gamma_1 \log(\beta). \tag{29}$$

By introducing the auxiliary variables

$$\log[a_0(\beta)] = A_0, \quad \log[a_1(\beta)] = A_1, \quad \log(\gamma_0) = \Gamma, \quad \log(\beta) = B \tag{30}$$

in (29), we have

$$A_0 = \Gamma + \gamma_1 B, \quad A_1 = \Gamma - \gamma_1 B. \tag{31}$$

Due to the negative values of the coefficient a_0 and the consequent impossibility of taking a logarithm, a change of their sign was made. It is necessary to change the sign on the right-hand side of Eq. (27) when converting Γ to γ_0. It is also necessary to calculate the uncertainty $u(\gamma_0)$, as follows

$$u(\gamma_0) = \frac{\partial \gamma_0}{\partial \Gamma} u(\Gamma). \tag{32}$$

After calculations based on the data from Table 2 and by using Eqs (27)–(34), we obtain the results in Table 3.

Table 3. Parameters and associated uncertainties for nonlinear functions

	$a_0(\beta)$	$a_1(\beta)$
γ_1	–3.017	–2.288
$u(\gamma_1)$	0.1774	0.1719
$\frac{u(\gamma_1)}{\lvert\gamma_1\rvert}\cdot 100\%$	5.880	7.513
Γ	–5.056	–3.958
$u(\Gamma)$	0.1040	0.1008
γ_0	$-8.784\cdot 10^{-6}$	$1.104\cdot 10^{-4}$
$u(\gamma_0)$	$2.103\cdot 10^{-6}$	$2.255\cdot 10^{-5}$
$\frac{u(\gamma_0)}{\lvert\gamma_0\rvert}\cdot 100\%$	23.94	20.42

Fig. 3 shows the nonlinear relation between the coefficients a of the linear function (19) and the damping ratio β.

Based on Eqs (27)–(28) and the parameters given in Table 3, we can easily obtain the coefficients of (19) for any damping ratio, but for constant values of the voltage sensitivity

$S = 1\,V/(ms^{-2})$ and cut-off frequency $f_c = 1\text{k}Hz$. It should be noted that for accelerometers, the damping ratio does not exceed 0.7.

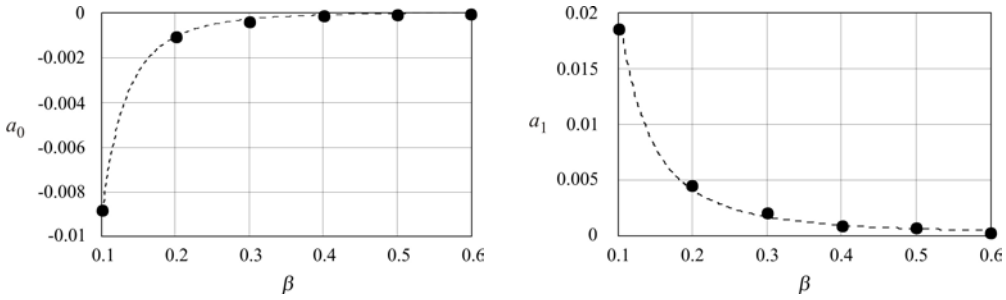


Fig. 3. Relations between the coefficients a and damping ratio β

5. Conclusion

The solution presented in this paper can be applied to the mutual comparability of the error generated by the accelerometers produced by different companies.

The solution above can be extended to a study of the impact on the error of changes both in the voltage sensitivity and cut-off frequency. Both functional and graphical relations between the error and parameters can be obtained in an analogous way as in the case of the damping ratio.

References

1 Di Natale C., Ferrari V., Ponzoni A., Sberveglieri G., Ferrari M. *Sensors and Microsystems*. Springer, 2013.

2 JCGM 101: Evaluation of Measurement Data. Supplement 1 to the Guide to the Expression of Uncertainty in Measurement - Propagation of Distributions using a Monte Carlo Method, 2008.

3 Link A., Täbner A., Wabinski W., Bruns T., Elster C. *Modelling Accelerometers for Transient Signals using Calibration Measurement upon Sinusoidal Excitation*. Measurement. Vol. 40, pp. 928–935, 2007.

4 Layer E. *Non-standard Input Signals for the Calibration and Optimisation of the Measuring Systems*. Measurement. Vol. 34, pp. 179–186, 2003.

5 Layer E., Tomczyk K. *Measurements, Modelling and Simulation of Dynamic Systems*. Springer-Verlag. Berlin Heidelberg, New York, 2010.

6 Rutland N. K. *The Principle of Matching: Practical Conditions for Systems with Inputs Restricted in Magnitude and Rate of Change*. IEEE Transactions on Automatic Control. Vol. 39, pp. 550–553, 1994.

7 Tomczyk K. *Special Signals in the Calibration of Systems for Measuring Dynamic Quantities*. Measurement. Vol. 49, pp. 148–152, 2014.

8 Tomczyk K., Layer E. *Accelerometer Errors in Measurements of Dynamic Signals*. Measurement. Vol. 60, pp. 292–298, 2015.

9 Tomczyk K. *A New Approach for Determining Signals with One Constraint that Maximize the Dynamic Error at the Output of Measuring Instrument*. Submitted to Measurement in May, 2017.

10 Wyner A. *Spectra of Bounded Functions in Open Problems in Communication and Computation*. T. M. Cover and B. Gopinath, Eds. Springer Verlag, New York, pp. 46–48, 1987.

11 Honig M. L., Steiglitz K. *Maximizing the Output Energy of a Linear Channel with a Time and Amplitude Limited Input*. IEEE Transaction on Information Theory. Vol. 38, no 3, pp. 1041–1052, 1992.

12 Elia M., Taricco G., Viterbo E. *Optimal Energy Transfer in Bandlimited Communication Channels*. Seizieme Colloque Gretsi, Grenoble, 1997.

13 Dorozhovets M. *Ocena wpływu oddziaływan systematycznych na parametry niepewnosci podczas aproksymacji metoda najmniejszych kwadratow*. PAK, vol. 12, pp. 22–25, 2006.

Agnieszka LESZCZYŃSKA[1], Aleksandra KOWALSKA[2], Krzysztof GRUDZIEŃ[3], Andrzej ROMANOWSKI[4], Dominik SANKOWSKI[5]

Pneumatic dense flow modelling and analysis based on ECT data

The paper presents statistical analysis of pneumatic dense flow with applied flow modelling method. The simplified flow model allows to define slug/plug flow pattern occurring in horizontal and vertical during pneumatic conveying of granular material. The described model was built based on ECT data gathered during measurements of material concentration changes during flow. The proposed model allow to define parameters of slug/plug shape and also frequency distribution of slug/plug appearing in time. In the paper were presented results of application of the developed software with interpretations of obtained of statistical parameters of plug/slug phenomena for two different flow characteristics.

1. Introduction

Efficient granular material transport plays a significant role in the industry factories. The most common method of granular material transport is the pneumatic conveying [1] [11]. Pneumatic conveyors are mainly used to transport granular material and powders. This type of transport allows moving material on considerable distances in between loading, unloading, and storage areas using pressurized air along pipeline installations. It is widely used in the agricultural, food, chemical, pharmaceutical, textile, paper and construction industries, as well as for transhipment in rail, road or water transport. The advantages of this type of transport are: high level of safety, low operating costs, flexibility of installation and ease of use. However, investigations aimed to improve the control efficiency of pneumatic flow are still conducted. The main flow parameters: carrying air velocity and amount of granular material loaded into pipeline, both can be modified during on-going flow in order to achieve steady flow pattern. Insufficient air velocity causes stationary slugs in the horizontal sections of pipelines, whereas in vertical sections material cannot overcome the gravity force and falls downward causing pipeline blockages [16]. Unblocking may require demounting of the pipe sections and cause substantial time lasting production breaks, resulting in considerable economic losses. On the other hand too high airflow velocity may cause material degradation and escalate disadvantageous electrostatic phenomena. The proper relationship between these two parameters, namely: air velocity and amount of granular material, in association with bulk material properties; allow keeping flow process in safety range [11] [12]. Control of the pneumatic transport process minimizes the occurrence of undesirable events and thereby

[1] Ph.D. student, Institute of Applied Computer Science, Lodz University of Technology, Poland, tel. +48426312750, fax: +420 541 146 276, e-mail: a.leszczynska@iis.p.lodz.pl
[2] Ph.D. student, Institute of Applied Computer Science, Lodz University of Technology, Poland, tel. +48426312750, fax: +420 541 146 276, e-mail: akowalska@iis.p.lodz.pl
[3] Ph.D., Institute of Applied Computer Science, Lodz University of Technology, Poland, tel. +48426312750, fax: +420 541 146 276, e-mail: kgrudzi@kis.p.lodz.pl
[4] Ph.D., Institute of Applied Computer Science, Lodz University of Technology, Poland, tel. +48426312750, fax: +420 541 146 276, e-mail: androm@kis.p.lodz.pl
[5] Professor., Institute of Applied Computer Science, Lodz University of Technology, Poland, tel. +48426312750, fax: +420 541 146 276, e-mail: dsan@kis.p.lodz.pl

increasing the overall efficiency of the pneumatic system, minimizing energy consumption, and reducing the loss of transported material.

Two main different types of pneumatic flow regime, dilute and dense flow can be distinguished depending on these parameters. During dilute flow particles are fully suspended and the solids concentration is less than 5%. In case of dense flow particles are not suspended and the solids concentration is higher than 30%. In the latter case interactions between particles and pipeline wall have significant influence on flow patterns. Dens flow can be observed in form of continuous slug or plug flow [9] [11]. Better the understanding of the flow phenomena results in application of new knowledge about particle materials behaviour into the algorithms of measurement data processing, which in turn can improve quality of control systems. In order to fulfil these needs new measurements and data processing methodologies are developed, which allow observing changes of the flow behaviour in temporal and spatial domains. In last few decades rapid development was observed in application of tomography systems to diagnosis and control of industrial processes. Process tomography is one of the most dynamically developing methods of industrial diagnostics. It has achieved the status of a modern visualization method and is competitive to the other imaging techniques. Its uniqueness is closely related to the feature of non-invasiveness. It can be also in charge of control and monitoring of some processes in the industry. It is primarily used in high-risk industries such as chemical reactors and coal-fired power plants as well as for other industries demanding to monitor the behaviour of the material flowing through the pipelines and other hardly accessible conveyors and receptacles. In order to understand and interpret process tomography data describing numerically the flow behaviour we should consider vast amount of the data yet it is difficult to gather and process such large amount of data at one time. Hence, we propose a simplified flow model that enables to study various parameters of flow.

2. Electrical capacitance tomography systems

Electrical capacitance tomography (ECT) is dedicated for monitoring of process, where primarily monitored material is nonconductive [14] [15] [17]. Based on electrical capacitance measurements between pairs of electrodes located on the vessel periphery is possible to visualise the permittivity distribution inside the measurement space. The permittivity distribution is directly related to the material concentration distribution, which allows tracking changes of material distribution during flow. The defined number of electrodes N (N=8, 12, 16) determines number of measurements L (L=28, 60, 132), which are used for image reconstruction. In literature can be found a lot of methods of the image reconstruction [2] [3]. Depend on time of numerical calculations and expected of image quality can be chosen different type of reconstruction methods. The presented results were obtained by application of simple linear back projected (LBP) methods [4].

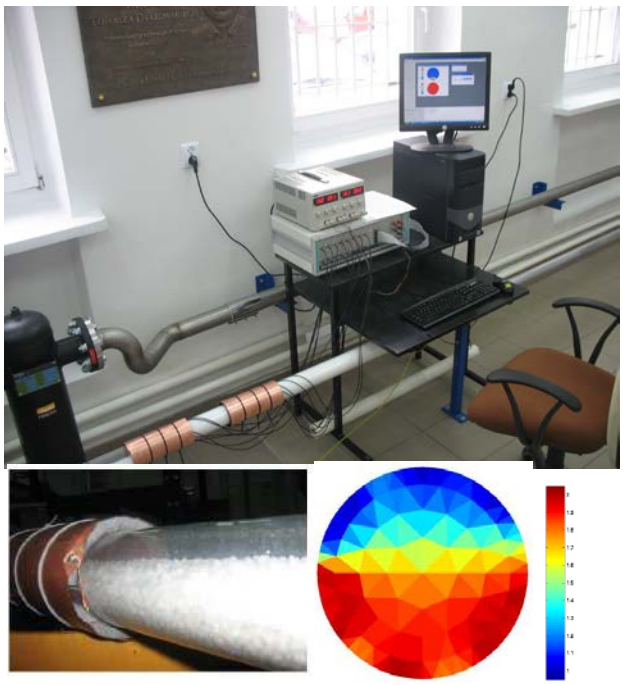


Fig .1. Photography of ECT system, with material distribution inside ECT twin-plane sensor. Upper photo reveals a segment of a flow test rig with twin plane ECT sensor on a pipeline and the measurement equipment with resulting two images of reconstructed measurement space cross sections. Bottom pictures show a zoom in half-occupied sensor (left hand-side) and a corresponding image reconstruction (right hand-side). Bottom left corner shows permittivity scale for the reconstructed image where dark blue color corresponds to empty and dark red color corresponds to fully occupied measurement space.

The use of capacitive electrical tomography allows visualizing the process in real-time and additionally gives opportunity for comprehensive research. Since the ECT measurement acquisition units offer capturing speed of several dozens or even greater than 100 frames per second it possible is to analyse dynamic flows based on tomography data; both using images as well as measurement records. Another aspect of ECT tomography nature is connected with capacitance measurements types. Two different approaches can be found, namely: AC or DC [5] [6]. From the current research point of view this element, although important, is not analysed.

The main advantage of ECT system is non-invasive way of industrial process visualization and monitoring. While spatial resolution is relatively low in comparison to X-ray tomography systems [18] [19], yet the temporal resolution is at a level sufficient to monitor dynamic processes' states [7] [13] [15] [20].

3. Related works

The applications of ECT measurements systems and data processing are developed in different directions in terms of both qualitative and quantitative process assessment methods. The following aspects are often recalled when analysing these fields of development:

- flow visualisation,
- flow pattern tracking,

- solid concentration measurement,
- solid velocity measurement,
- mass flow rate measurement.

The flow visualisation methods, including 2D and 3D imaging, play important role in flow investigation. Understanding flow phenomena, based on off-line image analysis, allows ultimately to implement better control systems. Flow visualisation and analysis of images provides useful information about e.g. reaction of flow behaviour to different conditions, including pipeline material properties, diameter of pipeline, humidity of material etc. On the other hand the development of ECT systems allows on-the-fly tracking of flow patterns and analysing changes during the on-going flow as it aims to diagnose the unstable flow phenomena. The possibility of pattern flow tracking allow to elaborate more robust methods also for predicting unwanted phenomena during flow, e.g. pipeline blockage. The last three measurement values (concentration, velocity and mass flow rate) are dedicated to development of more accurate flow meters. They are the monitoring parameters revealing the instantaneous flow rate. The method of measuring flow velocity based on correlation techniques is known for many years. However, it still needs improvements; choice of algorithm parameters make difficult to achieve proper accuracy with such flow meters. Velocity measurement, and also mass flow rate, is highly sensitive to flow pattern changes [2] [3] [19]. The appropriate selection of algorithm parameters during flow pattern changes is a challenging task. The accuracy of the estimation of material concentration is connected with spatial resolution of the ECT measurement system as well as with the applied permittivity model selection (parallel/series/Maxwell), [8]. However, spatial resolution is strongly limited; hence this element influences the accuracy of measurements.

As described above the main challenge during pneumatic conveying is controlling the flow regime, which has influence on accuracy of velocity and mass flow rate measurement and also allow to predict unwanted phenomena such as the pipeline blockage. Moreover, unstable flow patterns can result in higher power consumption or/and wear rate, product degradation. It would normally be expected that industrial scale pneumatic conveyors would operate with discontinuous dense phase flow regimes. This may take the form of discrete plugs of material, rolling dunes or a combination of both. In reality, a pneumatic conveying system may sequentially exhibit several flow regimes throughout its duration. In addition, the identification of the flow regime at critical sections of the pneumatic conveyor has fundamental meaning in order to estimate void fraction upon which many standard measurements, such as solids mass flow rate, depend [9] [10] [11].

The previous investigations showed potential of ECT data application to flow process state analysis. Tracking of the flow pattern can be conducted based on the statistical analysis of ECT data [12]. The different flow conditions generate different value of standard deviation, mean, median of ECT data. Interpretation of these parameters allows to determine the height of the solid/gas interface level within the cross-section of the pipe as well as slug length and distribution and slug frequency. These analyses give possibility of distinguishing among different flow patterns. In order to better understand information hidden in ECT data it is necessary to conduct wider analysis of data gathered during different flow experiments of pneumatic dense flow. Since this is difficult to collect and analyse extensive experimental database, a simplified model of ECT pneumatic dense flow is proposed. Based on this model it is possible to generate data for different types of material flow both for the stable as well as unstable slug flow with different characteristic of slug shape and frequency.

4. ECT dense pneumatic flow model

The model of ECT data pneumatic flow is based on temporal changes of material concentration, which are observed during dense flow in ECT measurement records. Fig. 2 presents examples of plug (fig. 2a) and slug (fig. 2b) flow types occurring in vertical and horizontal section of flow installation respectively. The material during slug/plug flow is proceeding along the pipelines in discrete-alike portions. The main difference between slug and plug patterns is visible after slug/plug has just left the ECT sensor space. For horizontal section material builds-up a stationary layer, where material occupies lower part of the pipeline. ECT data show (for presented example) level concentration in range <0.1, 0.2>. Next Fig. 3 presents zoomed-in plug plot coupled with selection of reconstructed images.

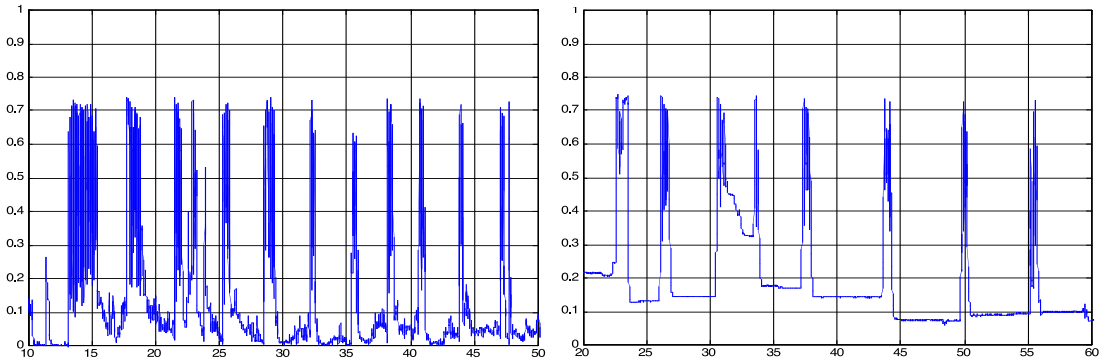


Figure. 2: Average of material concentration changes during pneumatic flow in a) vertical section, b) horizontal section of pipeline

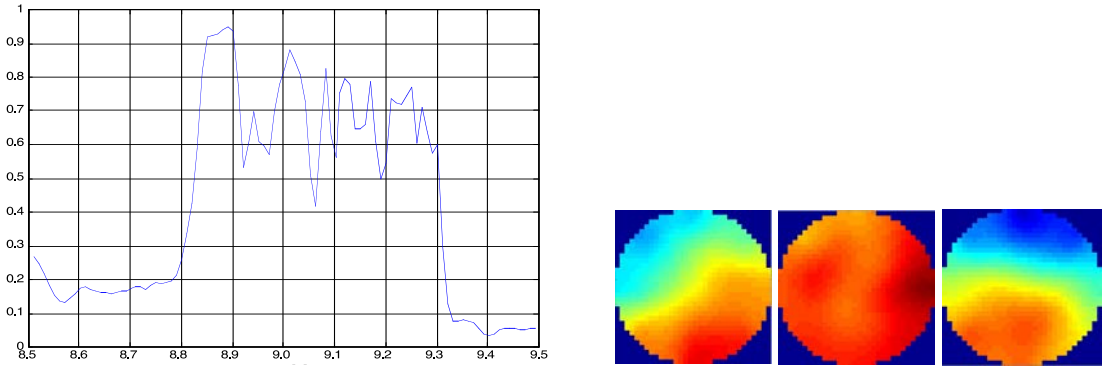


Figure. 3: Average of material concentration during single plug with examples of reconstructed images

Proposed model of slug/plug flow is based on knowledge gathered form ECT measurements. Fig.4 presents simplified scheme of slug/plug, which consist of 5 parts.

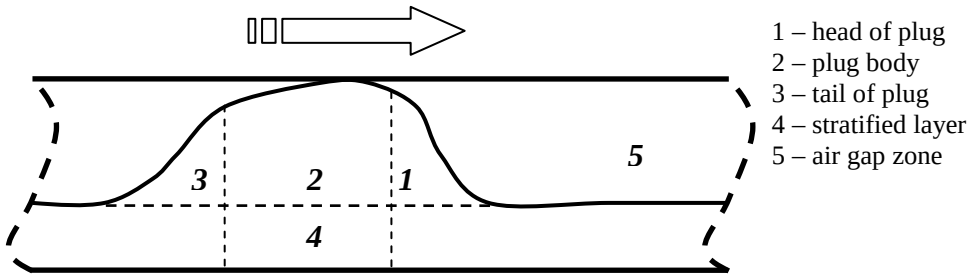


Figure. 4: Model of slug/plug flow

In order to define specific properties of slug/plug model it is necessary to offer a set of 6 parameters. Fig. 5 reveals a graphical explanation, which elements of slug/plug model can be modified to obtain expected type of flow in the proposed model. Besides the parameters connected with definition of single slug/plug shape there is additional possibility to define frequency of material building-up the slug/plug.

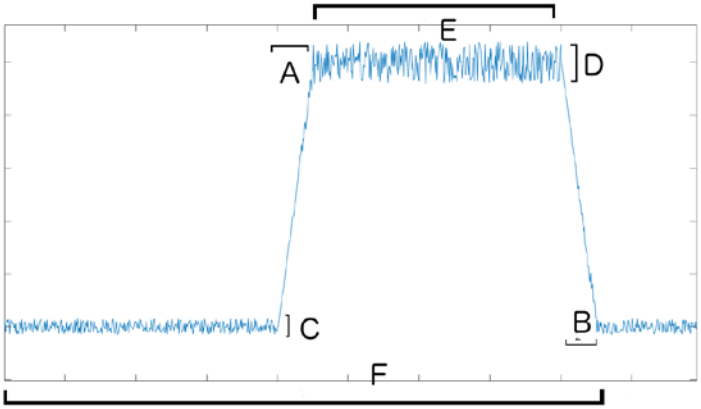


Figure. 5: Definition of slug/plug parameters of proposed flow model, A, B) slop head and tail part, C, D) level of stationary layer, D) level of occupied by material, E) time of slug/plug, F) distance between successive plug/slug.

5. Results

The development of methods to better understanding and monitoring pneumatic flow conditions were conducted based on implementation of proposed simplified model of plug/slug phenomena. As the ECT measurements system for which the proposed data processing method is dedicated provides useful information about temporal changes of flow conditions, the implementation of the flow simulation analysis was based on temporal statistical parameters calculation within a fixed time window. Figure 6 presents a screenshot view of application GUI. Left hand-side panel allows to define the parameters of window time; including length of window in context of correlation calculation and initial and time point. Right hand-side panel provides information about changes of calculated flow parameters in time as well as about the changes of material concentration for modelling or real flow.

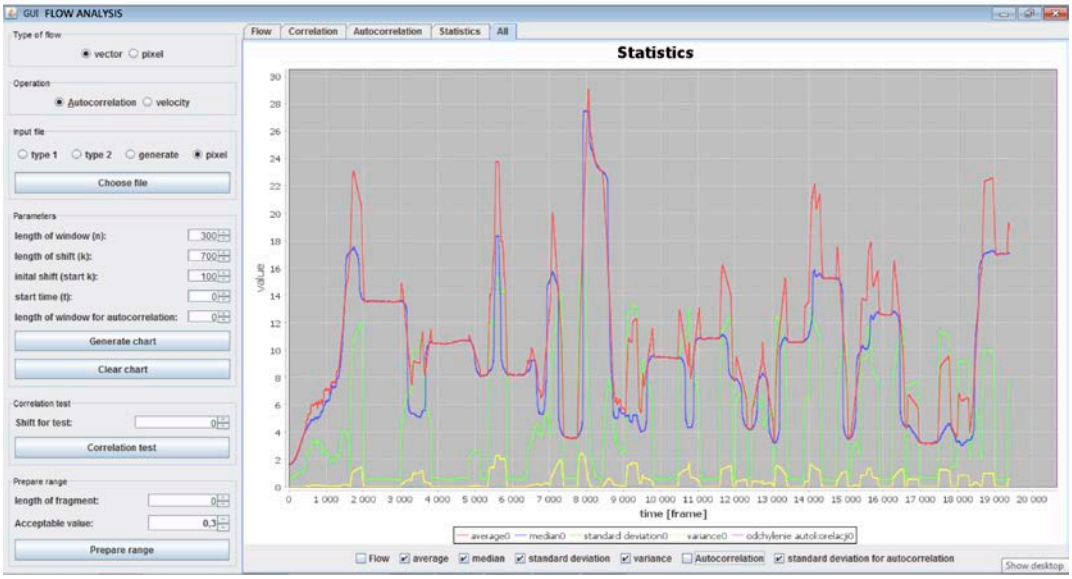


Figure. 6: Frontend view of application of GUI – flow analysis view example.

Examples of results for different user-defined flow models and flow characteristics are presented on figures 7 and 8. Figure 7a shows results of modelled stable slug flow. The slugs appear with the same time interval (1400 frames). Figure 7b presents example of calculations for average, median, standard deviation and variance for time window equal of 256 frames. As the flow was modelled as the stable one with constant slug flow time interval the obtained results show typical flow patterns.

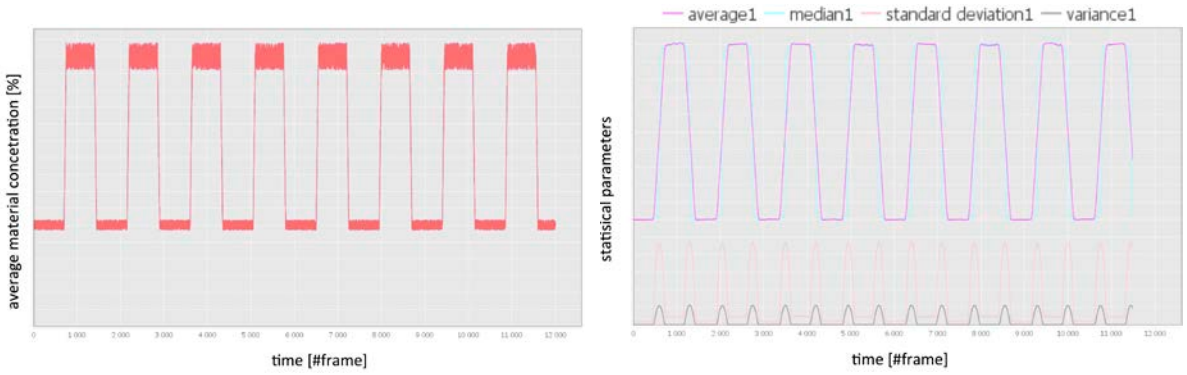


Figure 7. Examples of results of stable flow modeling. Left hand-side shows average material concentration vs. time plot. Right hand-side shows a set of statistical parameters vs. time plots.

The un-stable modeled flow in contrast to a stable flow type is characterized by the slugs appearing at variable time interval (Fig. 8a). Material is flowing in form of slug, however time distribution of slugs is heterogeneous. For this case parameters of the modeled slug flow are presented on Fig. 8b. The difference between stable (fig. 7a) and un-stable flow (fig. 8a) can be monitored based on the statistical parameters of time distribution values, where for unstable flow these values doesn't demonstrate homogeneous distribution in contrast to the stable flow case.

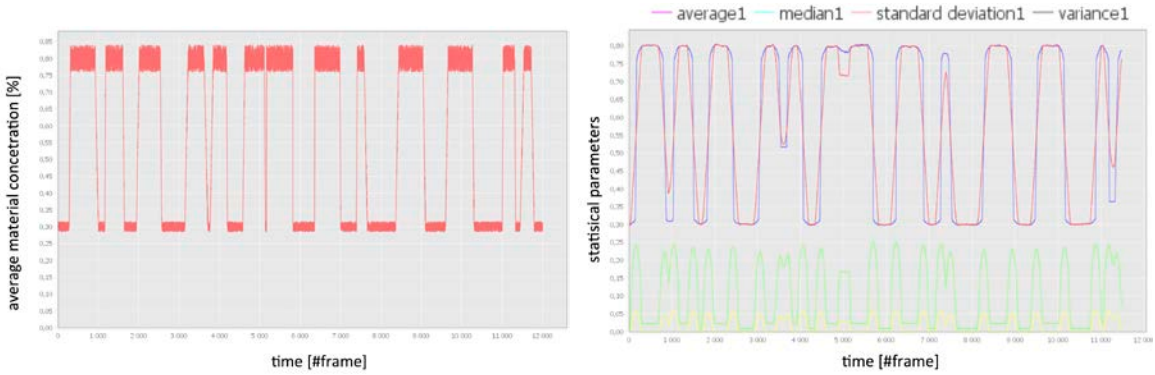


Figure 8. Examples of results of un-stable flow modeling. Left hand-side shows average material concentration vs. time plot. Right hand-side shows a set of statistical parameters vs. time plots.

2. Conclusions and future work

The obtained results and proposed method of flow modelled are under development yet the results presented in this paper shows large potential in practical use for ECT measurement data analysis applications. The results which are achieved with the aid of the proposed simplified flow model are dedicated to flow monitoring with ECT system employment. The way of flow modelling is developed under assumption that temporal resolution of ECT system is higher than for spatial resolution, however possibility to widen input data with tomography images should increase implications of results. In future work the developed application will be amended with correlation based data analysis techniques.

The preliminary results incorporating the understanding of data processing interpretation methods into the design of proposed flow modelling constitute the first step on the way to obtain robust, flexible and at the same time simplified flow monitoring system. In order to achieve this ultimate aim it is necessary to prepare complex measurement processing system, with assistance of other types of measurements (not only tomography data) that would automatically recognize flow pattern and detect any tendency to pipeline blockage. Such system can use machine learning methods which allow to improve effectiveness of control systems.

References

1. Konrad K.*: Dense Phase Pneumatic Conveying: A Review*, Powder Technology, 49, pp. 1-36, 1986.
2. Mosorov V.: *A method of transit time measurement using twin-plane electrical tomography*, Measurement Science and Technology, 17(4), p. 753, 2006.
3. Mosorov V.: *Flow Pattern Tracing for Mass Row Rate Measurement in Pneumatic Conveying Using Twin Plane Electrical Capacitance Tomography*, Part Part Syst Charact, 25, pp. 259-265, 2008.
4. McKeen TR., Pugsley T. S.: *The influence of permittivity models on phantom images obtained from electrical capacitance tomography*, Measurement Science and Technology, 13, 12, pp. 1822-1830, 2002.
5. McKee SL., Dyakowski T., Williams RA.: *Solids Flow Imaging and Attrition Studies in a Pneumatic Conveyor*, Powder Technology, 82, p. 105–113, 1995.

6. McKee S. L., Dyakowski T., Williams RA., Bell TA., Allen T.: *Solids flow imaging and attrition studies in a pneumatic conveyor*, Powder Technology, 82, pp. 105-113, 1995.

7. Yang W. Q., Lui S.: *Role of tomography in gas/solids flow measurement*, Flow Measurement and Instrumentation, 11, 3, pp. 237-244, 2000.

8. Jiamin Y., Wang H., Yang W.: *Characterization of Electrical Capacitance Tomography Sensors with Different Diameter*, IEEE Sensors Journal, 14, 7, 2014.

9. Neuffer D., Alvarez A., Owens D. H., Ostrowski KL., Luke SP., Williams RA.: *Control of Pneumatic Conveying Using ECT*, 1st World Congress on Industrial Process Tomography, pp. 71-76, 1999.

10. Ostrowski K. L., Luke S. P., Bennet MA.: *Real Time Visualisation and Analysis of Dense Phase Powder Conveying*, Powder Technology, 102, pp. 1-13, 1999.

11. Jaworski J., Dyakowski T.: *Application of electrical capacitance tomography for measurement of gas-solids flow characteristics in a pneumatic conveying system*, Measurement Science and Technology, 12, 8, pp. 1109, 2001.

12. Dyakowski T., Luke S.P., Ostrowski KL., Williams RA.: *On-line monitoring of dense phase flow using real time dielectric imaging*, Powder Technology, 104(3), pp. 287–295, 1999.

13. Romanowski A., Grudzien K., Aykroyd RG., Williams RA.: *Advanced Statistical Analysis as a Novel Tool to Pneumatic Conveying Monitoring and Control Strategy Development*, Part Part Syst Charact ,vol. 23, 3-4, pp. 289-296, 2006

14. Zhang W., Wang C., Yang WQ., Wang CH.: *Application of electrical capacitance tomography in particulate process measurement – A review*, Advanced Powder Technology, 25, 1, pp. 174-188, 2014.

15. Rymarczyk T., Szumowski J., Adamkiewicz P., Duda K., Sikora J.: *ECT Measurement System with Optical Detection for Quality Control of Flow Process*, Przeglad Elektrotechniczny, 2, pp.157-160, 2016.

16. Romanowski A., Grudzien K., Chaniecki Z., Wozniak P.: *Contextual processing of ECT measurement information towards detection of process emergency states*, 13th International Conference on Hybrid Intelligent Systems (HIS), pp. 291-297, 2013, DOI: 10.1109/HIS.2013.6920448

17. Chen C., Woźniak P., Romanowski A., Obaid M., Jaworski T., Kucharski J., Grudzien K., Zhao S., Fjeld M.: *Using Crowdsourcing for Scientific Analysis of Industrial Tomographic Images*, ACM Trans. Intell. Syst. Technol. 7, 4, pp. 1-25, 29016, DOI: https://doi.org/10.1145/2897370

18. Jelliti I., Romanowski A., Grudzien K.: *Design of crowdsourcing system for analysis of gravitational flow using X-ray visualization*, Federated Conference on Computer Science and Information Systems (FedCSIS), Gdansk, pp. 1613-1619, 2016.

19. Sankowski D., Mosorov V., Grudzien K.: *Mass flow measurements based on a virtual channel concept*, 5th International Conference on Intelligent Systems Design and Applications (ISDA'05), pp. 274-279, 2005, doi: 10.1109/ISDA.2005.62

20. Grudzien, K; Romanowski, A; Sankowski, D; Aykroyd, RG; Williams, RA.: *Advanced statistical computing for capacitance tomography as a monitoring and*

control tool , in Proc. of 5th International Conference on Intelligent Systems Design and Applications (ISDA 2005), pp. 49-54, 2005.

Jakub SZAFARZ[1] Mikołaj WOŹNIAK[2] Andrzej ROMANOWSKI[3], Krzysztof GRUDZIEŃ[4]
Zbigniew CHANIECKI[5]

The design and implementation of interactive visualisation system for ECT data analysis

This paper presents a system for tomographic data visualisation. Authors present multi-touch surface based desktop application, enhancing the ease of browsing, manipulating, examining and interpreting data, both in form of raw measurement records and reconstructed visuals, acquired through ECT tomography for a process of pneumatic conveying. User interface enables easy navigation and analysis even for large sets of measurement data. The most significant advantage of the system is to provide comfortable navigation among variety of datasets and different forms of collected data along with clear and informative visualisation for substantial collections of measurement records. Proposed tool has been qualitatively evaluated for efficiency and comfort of exploring and understanding of knowledge behind the ECT data.

1. Introduction

Process tomography is one of the most compelling techniques of analysis for multiple industrial trials. It enables investigation of the processes taking place in examined objects in non-inavsive and non-destructive way [18] [20] [22]. Moreover, reconstruction of images is possible even if there is no access to the interior of the object [6] [16], as well as being an ultimate solution for offline examination and simulation of processes. It is commonly used in chemistry, pharmacy, geology and environment protection. The most well-known approach to tomography are hardfield techniques (eg. X-ray tomography), ordinarily used in medicine. However, in numerous cases such approach is unsuitable due to physical features of monitored substance, high cost of implementations and low acquisition speed of the measurement units. In order to fulfil the needs of industrial implementations, where accurate measurement of dynamic processes is very important for their proper control, tomography modalities based on analysis of electrical field properties inside investigated objects are developed [17] [5]. An example of such tomographic system is electrical capacitance tomography (ECT), which is useful for monitoring state of granular flows [23].

Nowadays, development of ECT systems can be discussed in several aspects such as development of data acquisition, algorithms for better reconstruction of images, tomography data processing for deep analysis of monitored processes or adaptation of ECT approach in control

[1]M.Sc., Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND,

[2]International Faculty of Engineering, Lodz University of Technology, ul. Żwirki 36, 90-924 Łódź, POLAND, e-mail: mikolaj@pawelwozniak.eu

[3]Ph.D., Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, e-mail: androm@iis.p.lodz.pl

[4]Ph.D., Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, e-mail: kgrudzi@iis.p.lodz.pl

[5]Ph.D., Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, e-mail: z.chaniecki@iis.p.lodz.pl

systems. Nonetheless, it is a vital issue to properly view, visualise, analyse and interpret both the raw measurements and the reconstructed data. It is essential to provide informative and appealing visual forms of data presentation, as it highly simplifies analysis and contribute to better understanding of the problem itself [12] [7] [11].

Presented system is an approach to face the challenge of providing clear and intuitive tool for tomographic data analysis. The main goal is to enable the user to easily and efficiently navigate through raw measurement records as well as create visual presentation and interpretations of collected data. This goal is realized via desktop application using multi-touch surfaces. This paper focuses on applying ECT tomographic approach for purposes of investigating the pneumatic flow processes [14] [9].

2. Design goals and requirements

It is a commonly applied method to analyse and understand complex processes through graphical representations. For several purposes, it is way easier to extract proper information and derive conclusions from visuals than from bare numerical data [8]. Visualisations can be used to display the data in comprehensive and understandable way, which may vastly improve the process of creating a mental depiction of an investigated process by the final user [1]. However, to provide desired effect, visuals need to answer several requirements.

Graphical displays ought to:

- show the data
- induce the viewer to think about the substance rather than about methodology, graphic design, technology of graphic production or anything else
- avoid distorting what the data have to say
- present many numbers in a small space
- make large data sets coherent
- encourage the eye to compare different pieces of data
- reveal the data at several levels of detail, from a broad overview to the fine structure
- serve a reasonably clear purpose: description, exploration, tabulation or decoration
- be closely integrated with the statistical and verbal descriptions of a data set

Having fulfilled those conditions, graphics can be way more precise and revealing than conventional statistical computations [21].

The importance of proper graphical design is emphasized through the most fundamental aspect of display: visualisation always requires user involvement, since the core processes of information analysis take place in human mind [1]. The cognitive pattern of data analysis is the crucial one, which is why the design of the system shall avoid posing additional challenges in this area and let the user focus on evaluating the presented information [15].

3. ECT tomography data

In order to test and present the practical operation of the system, it was applied to deal with the data acquired from ECT tomography. The investigation concerned a granular material flow in a non-transparent instalation for pneumatic conveying.

Considering pneumatic conveying of solid, estimation of the main parameters of flow is based on material concentration profile. The measurement of the velocity and the mass flow rate is directly linked to the material concentration distribution in pipeline during pneumatic conveying [14]. The proper understanding of data requires familiarity with ECT tomography setup, concerning both hardware and software. Such system consists of:

- ECT sensor, consisting of several electrodes located on investigated object boundary
- data acquisition unit - heart of ECT system
- data/image reconstruction software module
- software module of ECT data processing and process parameter calculation

An example of such system has been depicted in figure 1.

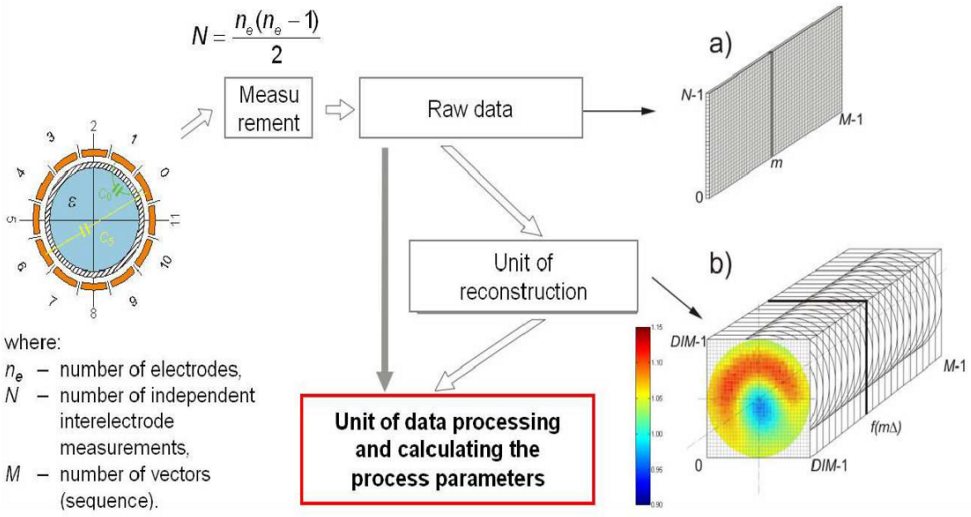


Figure 1: Exemplary setup of ETC tomography system, a) measurement data analysis, b) reconstructed data analysis [4].

Obtained measurements are reconstructed to the form of image, representing information on material concentration in the container. The data is usually normalized and presented as the fraction of total ECT sensor-occupied space by the material. Determining of the concentration profile enables further analysis through calculating the velocity profile and the mass flow rate. All those parameters provide extensive knowledge on the monitored process [19].

The typical way of ECT data processing for both measurement data and reconstructed images, is based on temporal and spatial changes of material distribution in ECT sensor space. Current state of the process can be represented as single reconstructed image or set of measurements. Analysis of the material concentration distribution based on image is supported with its

temporal changes profile, which may be represented through analysis of measurements or the sequence values of image pixels.

The proposed solution provides tools, which notably improve the experience of concentration profile analysis and enable deriving comprehensive conclusions, in comparison to traditional approach.

4. System design and results

The solution is designed to operate as the last unit of basic ECT system, utilising the functionality of previous modules to generate proper input data, which will be visualised, analysed and interpreted using the proposed solution. The main goal is to provide several tools for analysis of measurement data. It is necessary to visualise the data in different arrangements, while properly handling both navigation through the unprocessed input data, aside to graphical interpretations. This task is fulfilled through supporting various forms of data presentation: images, graphs and topograms. Direct approach to each feature is described in detail further this paper.

4.1. Interaction design

Considering the goals set to design of the proposed system, the greatest emphasis has been put on intuitive and comfortable manner of browsing, exploring and managing the data. In order to obtain the ease of data examination, multi-touch surface has been applied. Operating the application is based on simple and cognitively intuitive gestures. The gestures and their applications are briefly described below.

To obtain possibly intuitive and effortless interface, basic type multi-touch interactions have been integrated (fig. 2). The operation of the system is realised through simple gestures, such as tapping, scrolling, panning, zooming and pinching. Tangible attitude vastly boosts the comfort of work with visuals [13] [21].

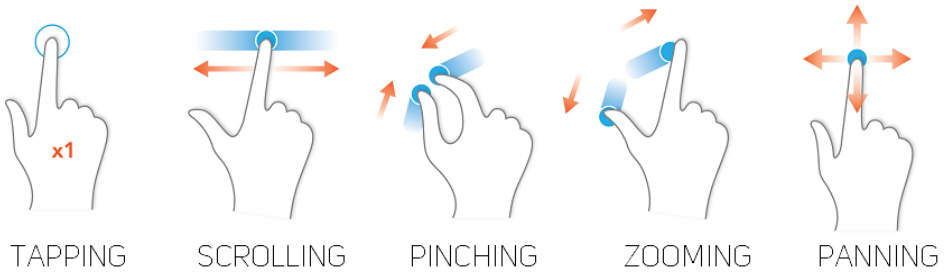


Figure 2: Gestures used for system operation. Source: GRPH3B18 CC BY-SA 3.0

Even though the destined user may be considered as an expert, the interaction pattern needs to be as intuitive and understandable as possible [13]. The significant difficulty of the task realised imposes the need to design user - friendly tool, which will be easy to master even for non-specialist user. To avoid unnecessary complexity on UX pattern, the tangible manner of operation was chosen as the most appropriate. It bypasses the demand for graphically accessible navigation tools - like slidebars or icons, while maintaining full functionality in terms of solving tasks and providing convenient usability [24].

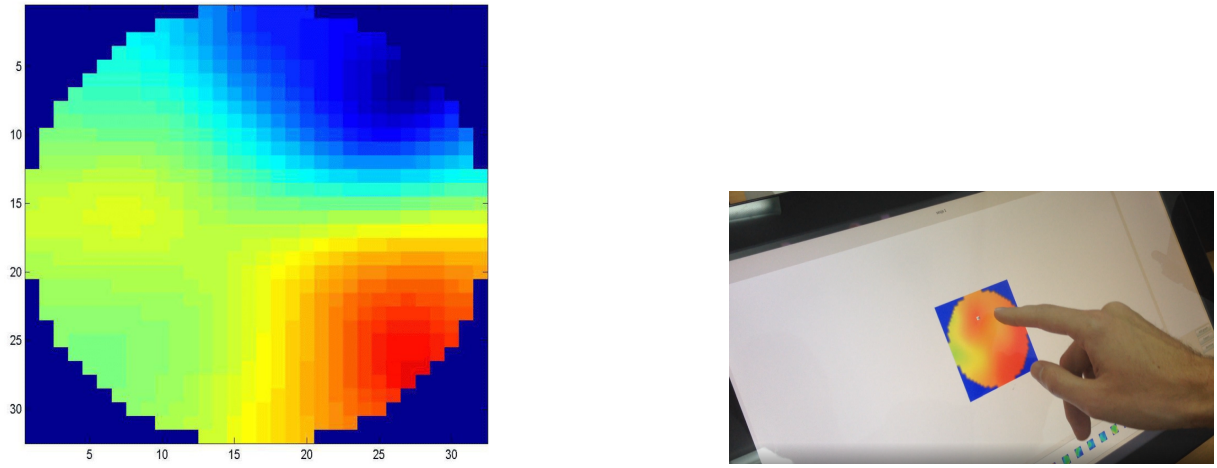
The tangible interactive system has been integrated into java-based desktop application. This approach provides the ease of maintenance and customization of the application to the individual needs. The only fixed requirement is providing a multi-touch display surface.

4.2. Application for ECT data analysis

The ultimate quest for ECT data analysis is to properly measure and interpret the process inside the examined container, especially through establishing the concentration profile of the material. The raw data is reproduced to the form of images, which may be further analysed by generating corresponding graphs and topograms, enabling the user to conclude on alterations occurring during the investigated trial. One of the most desired functionalities is to provide handy transition between areal analysis of material concentration using images and depiction of concentration shifts in time. Moreover, time-related depiction of changes taking place during the process, both based on images and raw data, is required to meet the professionals' needs.

Images

The input data is reproduced to the form of images in order to generate more descriptive graphical representation. The solution provides support for enhanced navigation along the images and creating interpretations for the information from the image. The program displays scrollable image sequence, which represent measurement data. User taps an image from the downscaled range bar and can easily magnify the particular scope of the image through zooming and pinching. The display can be controlled through scrolling and panning. The key feature for image analysis is the ability to choose the range of the image to be interpreted further and instantly generate corresponding graph or topogram, as desired (fig. 3).



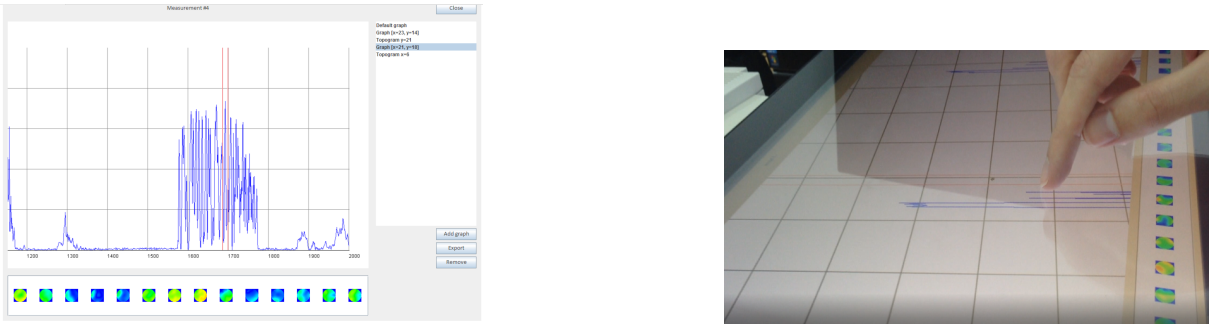
(a) (b)

Figure 3: Analysis of tomographic images: (a) - an image reconstructed from raw data, (b) - interactive examination of particular image

Graphs

Graphs can be generated for arbitrary sets of data. Plots generated for the desired range may be further analysed through magnifying, scaling and choosing more precise scope or a certain point. (fig.4a) Moreover, choosing a single spot enables finding the corresponding image. This

feature highly increases the ease of inspecting only the most interesting part of collected information and its farther examination.(fig.4b) This mechanism is supported by interactive operations such as scaling and distortion. Using simple gestures, the user is able to increase precision of examination through scaling the graph, while being dynamically supplied with corresponding images, being straight representation of raw data.

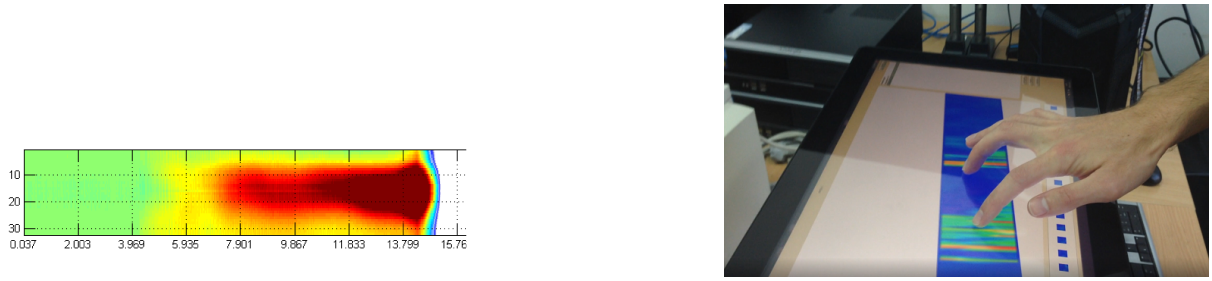


(a) (b)

Figure 4: Detailed examination of data in form of graphs. (a) - graphs generated using the solution (b) - zooming of the range enables more precise analysis

Topograms

Topograms are an interesting form of representation of tomographic data (fig.5a). They consist of a set of records, acquired from the same row or column, extracted from all images in the sequence. Such representation provides clear and understandable depiction of changes occurring in a certain area of examined object.



(a) (b)

Figure 5: Topograms and their detailed examination provides knowledge on changes occurring in the examined object. (a) - Interpretation of data range through topogram (b) - Analysis of generated topogram. Corresponding images are dynamically assigned.

The solution supports generating topograms for any specific row or column chosen from the image. In order to generate such, user pans through the desired piece of data and taps proper option on a menu bar (fig.5b).

4.3. Evaluation procedure

In order to assess the system design and performance, an exemplary task was prepared and held in controlled conditions. N=13 participants were asked to perform basic tasks related to process tomography measurement data analysis. The experiment was conducted as follows. At the beginning, users were given a 5 minute introduction to the system and the task set. Introduction covered watching a video on pneumatic flow as well as description of various types of tomography measurement data. Then, users were introduced to system interface operation.
The task scenario covered three phases. 1. Participants were asked to find segments of stable slug flow within the sequences of recorded data. 2. Then, only for those ranges of stable flow, users were asked to analyse, estimate or describe in a written form two pieces of information. (A) Percentage of the cross-section occupancy of the stationary layer within the measurement space on average within the analysed flow segments was to be determined. (B)The average level of stationary layer in slug flow was to be assessed within the analysed stable flow segments; whether it is higher or lower before or after the slug (with respect to the flow direction) passing through the ECT measurement sensor space.

4.4. Results and discussion

Right after completing the task, users were asked to give feedback on their experience with the system. The opinions were gathered through surveys, examining satisfaction with system interface and studying how challenging the task was.
In order to investigate how demanding the exercise revealed to be, Task Load Index was established. Results of this survey present that users felt successful with their performance, although the task was considered as difficult. However, their frustration on realisation is relatively low. Results of TLX survey are depicted in figure 6.

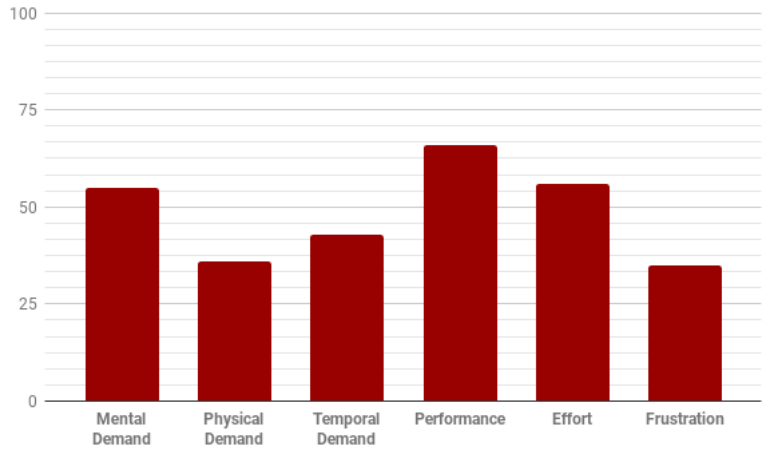


Figure 6: Evaluation of Task Load Index. Vertical axis shows to which extent the load occurred (1-100).

Another key aspect to evaluate was satisfaction with using the visuals for data analysis. Users were asked for opinion on convenience and efficiency of using different forms of graphical representations. It was not only the graphics to be assessed, but also the interactive manner of

manipulating them. Evaluation was held in form of survey, with 5-grade scaled yes/no questions on the experience. Opinions concerning usage of particular visuals are depicted in figure 7.

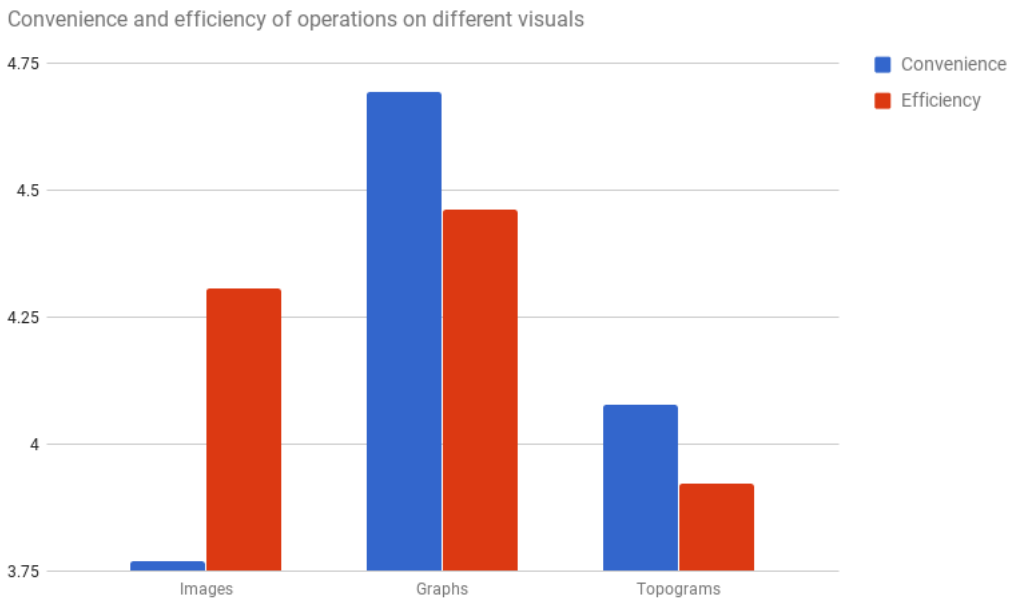


Figure 7: Convenience end efficiency assessment of different visuals. The vertical scale presents average an grade (1-5) awarded.

The survey shows that graphs are definitely the most desired form of visual representations, as they are easiest-to-use and very efficient. Analysis using reproduced images was satisfactory in terms of effectiveness, while being commented as inconvenient. Topograms are perceived as quite handy, whereas inefficient for data analysis.
Performed evaluation indicated a certain need for more accessible tools, which can significantly improve the work flow of tomographic data analysis. Nowadays, it is not only accuracy which is required, but also proper usability, so the act of analysis is no more complex and frustrating. Proposed solution faces the challenges of improving the experience of ECT data analysis, although there is still room for further improvements. The multi-touch approach appears as reasonable due to being cognitively intuitive and sufficiently efficient. It would be advisable to pay efforts on developing design of each functionality, as the choice of visualisation forms revealed to be accurate.

5. Conclusions and future work

We demonstrated the principles of operation, as well as the technical potential in terms of efficiency and usability of the system for analysis and interpretation of process tomography data. Initial results reveal that proposed multi-touch interface, coupled with carefully designed visual forms of data presentation and manipulation proved qualitatively to over-perform the classical approaches to data exploration and interpretation. In conclusion, the proposed application is a successful endeavor to answer the demand for efficient and user-friendly tool for improving the experience of tomographic data examination.

Furthermore, future research on both interface as well as the back-end side seems to be

promising in terms of data analysis further advancement. Big data analysis of experimental datasets, if properly implemented, will offer yet unseen opportunities for data mining, pattern recognition and data understanding [2]. Especially, the web-based implementation of the proposed system will enable to employ it for crowdsourcing tasks as proposed in [3] and [10].

References

1. Keim D. A. Information visualisation and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 8(1):1–8, 2002.

2. Romanowski A., Skuza M., Woźniak P., Grudzień K., and Chaniecki Z. Big data computational environment for tomography measurement data. *Process Tomography WCIPT7, Poland*, pages 537–546, 2013.

3. Chen Chen, Paweł W. Woźniak, Andrzej Romanowski, Mohammad Obaid, Tomasz Jaworski, Jacek Kucharski, Krzysztof Grudzień, Shengdong Zhao, and Morten Fjeld. Using crowdsourcing for scientific analysis of industrial tomographic images. *ACM Trans. Intell. Syst. Technol.*, 7(4):52:1–52:25, July 2016.

4. Chaniecki Z. Grudzień K. Sankowski D. Tomographic raw data - measurement and analysis, fundamental terms. 2013.

5. P. Fiderek, J. Kucharski, and R. Wajman. Fuzzy inference for two-phase gas-liquid flow type evaluation based on raw 3d ect measurement data. *Flow Measurement and Instrumentation*, 54:88—96, 2017.

6. S. F. Filipowicz and T. Rymarczyk. Measurement methods and image reconstruction in electrical impedance tomography. *Przeglad Elektrotechniczny*, 88(6):247–250, 2012.

7. K. Grudzien. Visualization system for large scale silo flow monitoring based on ect technique. *IEEE Sensors Journal*, PP(99):1–1, 2017.

8. K. Grudzien, Z. Chaniecki, A. Romanowski, M. Niedostatkiewicz, and D. Sankowski. Ect image analysis methods for shear zone measurements during silo discharging process. *Chinese Journal of Chemical Engineering*, 20(2):337 – 345, 2012.

9. A. J. Jaworski and T. Dyakowski. Application of electrical capacitance tomography for measurement of gas-solids flow characteristics in a pneumatic conveying system. *Measurement Science and Technology*, 12(8):1109, 2001.

10. Ibrahim Jelliti, Andrzej Romanowski, and Krzysztof Grudzień. Design of crowdsourcing system for analysis of gravitational flow using x-ray visualization. In M. Ganzha, L. Maciaszek, and M. Paprzycki, editors, *Proceedings of the 2016 Federated Conference on Computer Science and Information Systems*, volume 8 of *Annals of Computer Science and Information Systems*, pages 1613–1619. IEEE, 2016.

11. B. Matusiak A. Romanowski G. Rybak D. Sankowski K. Grudzień, Z. Chaniecki. Visualisation of granular material concentration changes, during silo discharging process, using ect large scale sensor. *Image Processing and Communications*, 17(4):327–338, 2013.

12. R. Mazza. Introduction to information visualisation. *Resources of Faculty of Communication Sciences, University of Lugano*, 2004.

13. T. Moscovich. *Principles and Applications of Multi-Touch Interaction*. Brown University, 2007.

14. V. Mosorov. Flow pattern tracing for mass flow rate measurement in pneumatic conveying using twin plane electrical capacitance tomography. *Particle and Particle Systems Characterization*, 25(3):259–265, 2008.

15. D. A. Norman. *Emotional Design. Why We Love (or Hate) Everyday Things.* Basic Books, NY, 2004.

16. H. Fidos T. Jaworski J. Nowakowski D. Sankowski R. Wajman, P. Fiderek and R. Banasiak. Metrological evaluation of a 3d electrical capacitance tomography measurement system for two-phase flow fraction determination. *Measurement Science and Technology*, 24(6):065302, 2013.

17. T. Rymarczyk. Using electrical impedance tomography to monitoring flood banks. *International Journal of Applied Electromagnetics and Mechanics*, 45:489–494, 2014.

18. T. Rymarczyk. *Systemy tomografii procesowej w bezinwazyjnych pomiarach diagnostycznych. in: Inżynieria Bezwykopowa vol. 53, no. 1.* Wydawnictwo INŻYNIERIA Spółka z o. o., January - March 2014.

19. D. Sankowski, V. Mosorov, and K. Strzecha. *Metody analizy obrazów z tomografów procesowych, in: Przetwarzanie i analiza obrazów w systemach przemysłowych. Wybrane zastosowania.* Warszawa, Wydawnictwo Naukowe PWN, 2011.

20. D. Sankowski and J. Sikora. *Electrical capacitance tomography: Theoretical basis and applications*. Instytut Elektrotechniki, Warsaw, 2010.

21. E. Tufte. *The Visual Display of Quantitative Information*. Cheshire, CT, Graphics Press, 1983.

22. M. Wang. *Industrial Tomography: Systems and Applications*. Woodhead Publishing, Elsevier, 2015.

23. D Watzenig and C Fox. A review of statistical modelling and inference for electrical capacitance tomography. *Measurement Science and Technology*, 20(5):052002, 2009.

24. P.W. Woźniak. *Everyday encounters with data: Exploring engaging and collaborative interactive technologies*. Licentiate of engineering thesis, Department of Applied Information Technology, Chalmers University of Technology, 2013.

Mateusz MIŁAK[1], Krzysztof GRUDZIEŃ[2], Andrzej ROMANOWSKI[3], Dominik SANKOWSKI[4]

Analysis of Slug Flows in Pneumatic Conveying of Solid based on CCD images processing

An algorithm for processing of CCD images dedicated to analyse slug flow phenomena taking place during pneumatic conveying of granular material is presented in this article. The aim of the work is development of the system providing information about characteristic flow parameters. The obtained results can be used as supporting methods for verification of the other measurement systems or as a separate system for flow analysis. This paper covers the work-flow of the proposed algorithm coupled with examples of obtained results with discussion about possible further development of the system.

1. Introduction

Pneumatic conveying of granular materials is one of the fastest and effective ways to move material among distant places [8] [12]. Therefore this type of transport is widely applied in industry. Although the pneumatic conveying is known for a long time research on its effectiveness improvement is still conducted. Too high power consumption or pipeline blockages are major problems for dense flow types. In laboratory environment different measurement or visualization techniques allowing spatial-temporal behaviour of granular material are applied in laboratory settings for various pipeline constructions and different process conditions. One of the most popular are process tomography techniques, and especially electrical capacitance tomography (ECT), [12] [13] [15] [17]. In ECT reconstructed image provides knowledge about material distribution inside the measurement sensor space during the flow. The drawback of ECT scanner solutions is that they are susceptible to interference such as electromagnetic fields and moisture. Using the ECT scanner alone results in an electromagnetic field that can interfere with nearby devices. Some other visualisation systems are based on optical images, where the core element is a fast camera [11] [14]. Both solutions employ image processing methods in order to extract information about process based on sequences of images [4] [5] [16] [21]. Image processing methods coupled with data acquisition system based parameters allow to obtain even more information about investigated process.

Image processing is often applied for the velocity measurement of pneumatically conveyed particles with aid of tracking particles methods [1]. This approach uses a CCD camera with

[1] Ph.D. student, Institute of Applied Computer Science, Lodz University of Technology, Poland, tel. +48426312750, fax: +420 541 146 276, e-mail: mmilak@iis.p.lodz.pl
[2] Ph.D., Institute of Applied Computer Science, Lodz University of Technology, Poland, tel. +48426312750, fax: +420 541 146 276, e-mail: kgrudzi@kis.p.lodz.pl
[3] Ph.D., Institute of Applied Computer Science, Lodz University of Technology, Poland, tel. +48426312750, fax: +420 541 146 276, e-mail: androm@kis.p.lodz.pl
[4] Professor., Institute of Applied Computer Science, Lodz University of Technology, Poland, tel. +48426312750, fax: +420 541 146 276, e-mail: dsan@kis.p.lodz.pl

adequate illumination to acquire tracking particle images flowing along the section of a transparent pipeline. Another use of CCD cameras is application to generate material concentration profiles using digital imaging technique [2] [6] [9] [10]. This technique is easy to setup and is not influenced by an environmental disturbances such as humidity, charge static, noise and others. The results show that the image concentration profiles obtained from this technique are better than the information derived by methods based on ECT reconstructed images, however it is only useful for flow through transparent pipeline. Combination of CCD imaging and electrostatics and ECT sensor seems to be very interesting approach to obtain complex information about flow [7] [20], similar as applied CCD imaging for verification of other methods of flow measurement [18].

Another example of CCD visualization is measurement of pneumatically conveyed particles distribution, where results were obtained by application of contour-based image segmentation; among the others, transport of coal using image processing monitoring is described in [3]. In this context a CCD camera is used to capture the images of the particulate flow field. Author claims that continuous monitoring of granular material is essential in many coal-relying industrial processes, especially where electricity generation is involved. For this purpose, it is often used to analyse the recorded images by means of special algorithms that online determine the current parameters of the material being transported.

In the presented case was used CCD camera to explore material during transport. Figure 1 illustrates simplified schema whole installation to transport materials. Installation consists of container, pipelines and mechanism to transport bulk materials inside pipeline from container to container; all working in a closed loop workflow. Major part of installation consists of opaque pipelines (indicated with light grey colour on Fig. 1). Dark grey (greenish in a coloured version) colour indicates transparent segments of pipeline installation (one in bottom, horizontal section and the other on the left hand-side in the vertical section). The bottom one is supported with a video CCD camera recording the flow. Images captured with this setup are used as input for algorithm development; with an off-line method. It means, that first measurements are collected and then in next stage this data are processed just after the image acquisition is finished. The flow imaging is limited only to horizontal part of installation.

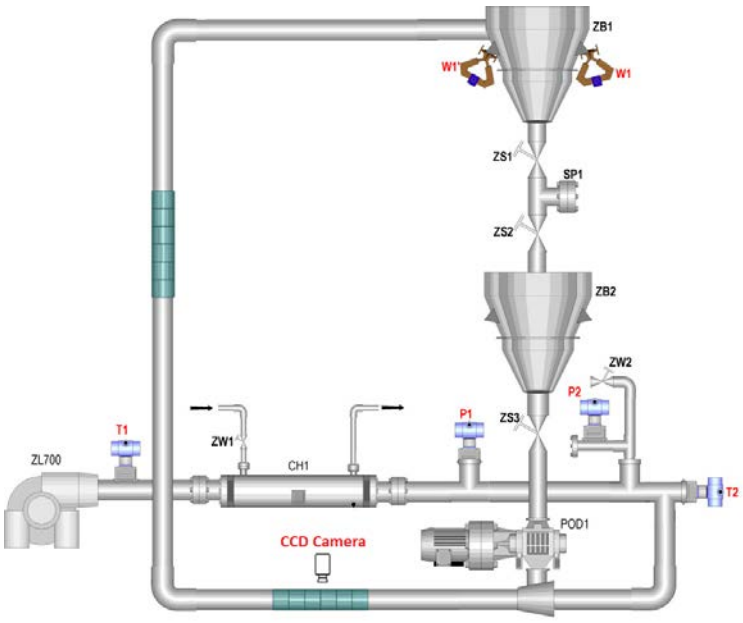


Figure 1: Scheme of the pneumatic conveying installation for transport bulk material

Figure 2: Photo of a CCD-monitored flow installation horizontal section

2. Image processing algorithm

As mentioned earlier next stage is to process the imagery data. Fig. 3 illustrates the algorithm work-flow used to analyse the flow of the material inside pipeline. This program can detect moments when segments of flowing material appear within the visible window in the pipeline, measure for how long this pipeline section is occupied by flowing material and pre-calculate volume of the material which forms the stationary layer on the bottom of the pipe.

After starting to capture CCD images the region of interest is defined; a distinct part of image area where transparent section of pipeline is visible. Such reducing of image accelerates time needed for image processing. Additionally the input image has a lot of elements, which constitute redundant noise. After only a part of a picture is selected the majority of noise is discarded. In the next step, the reduced image is thresholded in an HSV colour space that results in binarization. Fig. 4a) shows the image without threshold in situation when part of material is at the bottom of the pipe. After thresholding, Fig. 4b), presents mask only for transported material inside the pipe.

Next stage in image processing is cutting out the part of original image including only transported bulk material. In this way it is possible to observe image with quantity of materials inside the pipe. The examples of results are presented in Fig. 4a) and b), for case when material occupies majority of analysed section of the pipe and when the pipe is almost empty. This operation allows isolate only images with transported material and discarding the empty images from further processing.

Based on the binary image the boundary of flowing material is determined. In order to achieve this aim morphological operations to find contours of object (in this case the object is the part of material occupied pipeline) are applied. Results of this operation can be seen on Fig. 5a). Contour help to circumscribe the volume of the flowing material. The same effect is presented on filtered image (Fig. 5b)).

The detected contours are used to calculate volume of the bulk material in pipeline. From all contours, algorithm selects only those points, which create top contours of the detected object. Based on these points the level of volume pipeline occupied by material is calculated and determined. Fig. 5a) and 5b) give examples of images with information about the level of material occupying the pipe during the material transport. Red circles visible on Fig. 6 represent cross-section of the pipe and green horizontal lines show average level of the material volume occupying the pipe cross-section at a given moment.

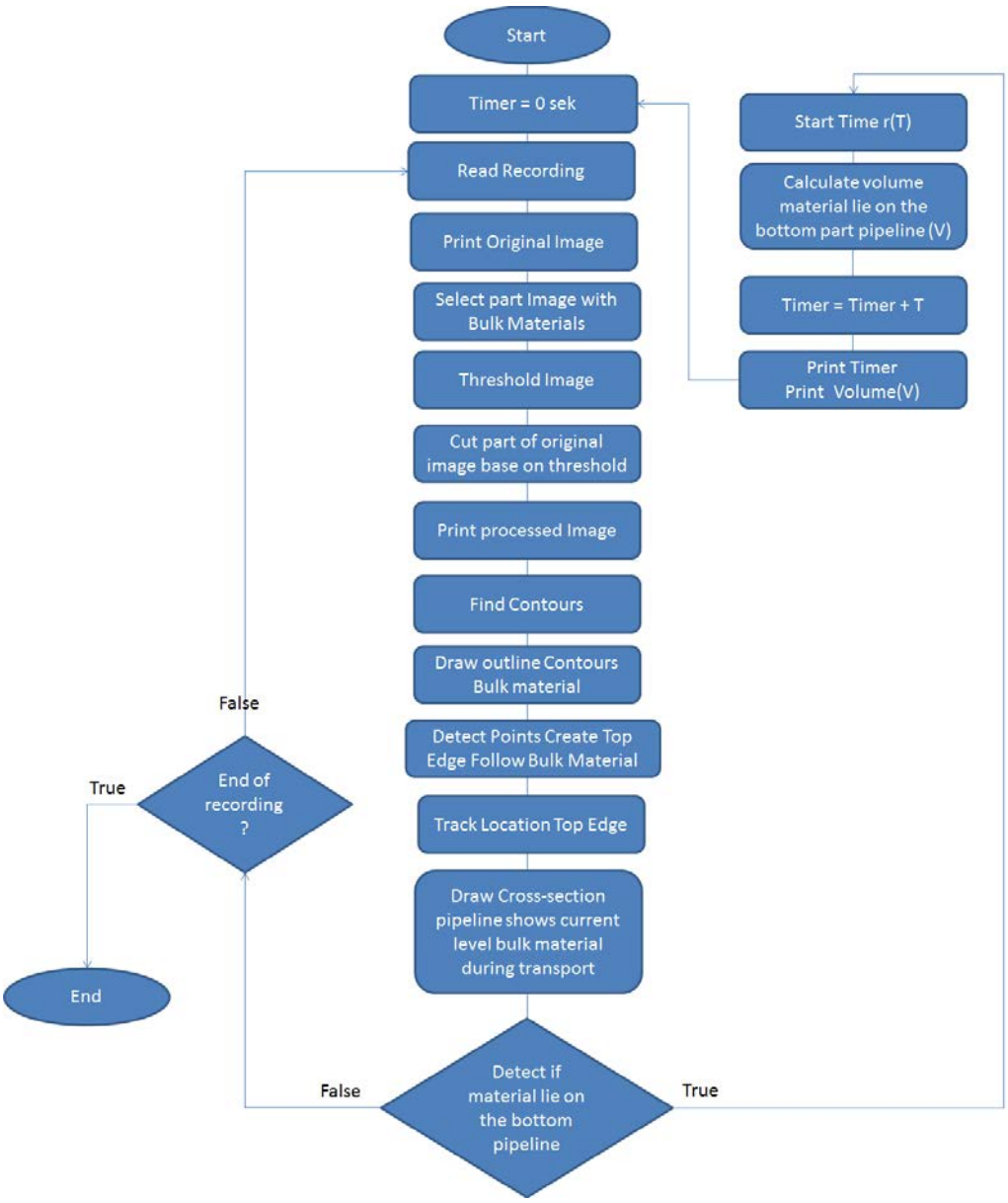


Figure 3: Algorithm flowchart

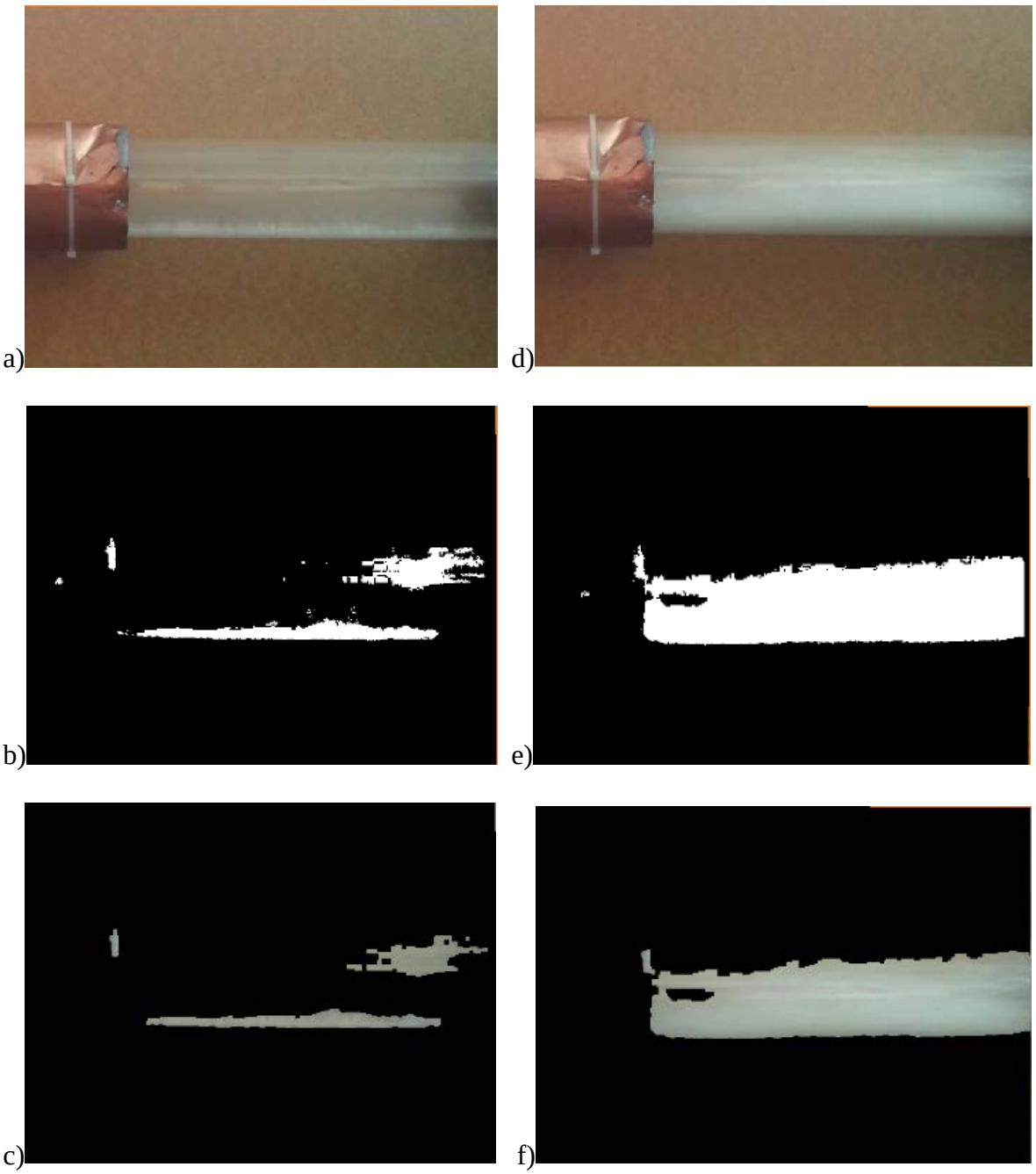


Figure 4: Examples of image processing results, a)image before threshold when part of transported material lie on the bottom pipeline, b) Image after threshold when part of transported material lie on the bottom pipeline, c) image after cut part of original image base on threshold from b), d) Image before threshold during transport material, e) Image after threshold during transport material, f) Image after cut part of original image base on threshold from e)

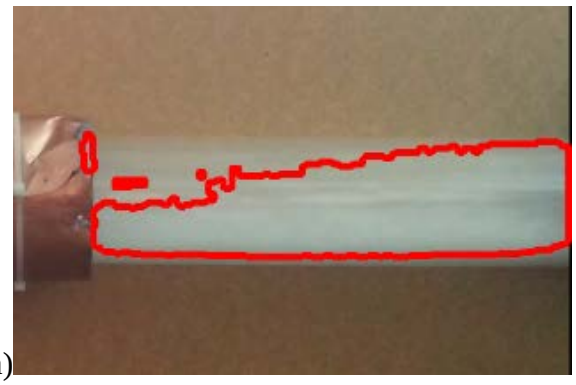
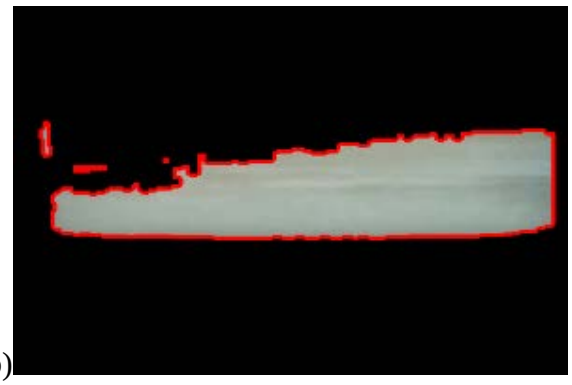
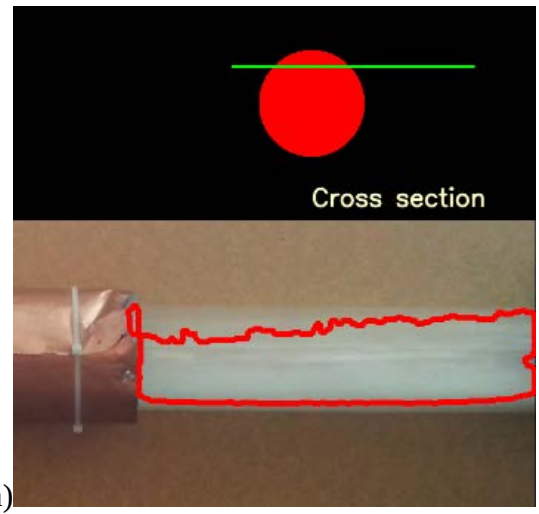
a)  b) 

Figure 5: Examples of counter detection results a) Original image with draw contour, b) Filtered image with draw contour

a) 

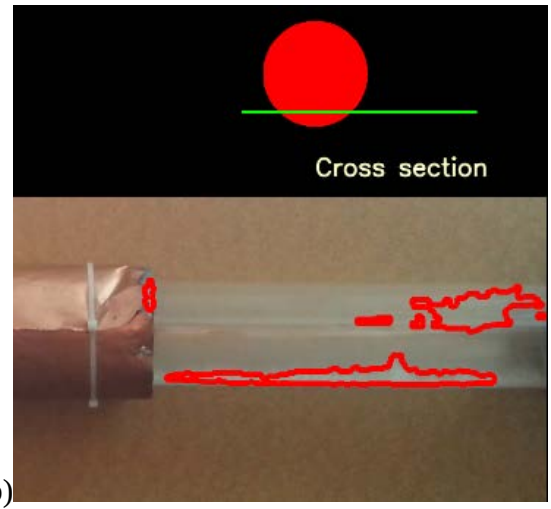
b) 

Figure 6: Visualization of level of material occupied analysed part of pipeline, a) image with cross section represent current volume during transport material, b) Image with cross section represent current volume when part of material lie on bottom pipeline

3. Results and discussion

The results obtained with the proposed image processing algorithm allow to estimate the level of pipe cross-section occupied by the granular material. Fig. 7 shows examples of flow material images with surrounding contours. Based on this information it is possible to estimate the percentage of pipeline occupied by material. Presented examples are taken for time points, when material creates stationary level (Fig. 7a and Fig. 7b) and when material is starting to form a slug (Fig. 7c and Fig 7d).

As it can be noticed, proposed algorithm generates better results when material is not moving and forms only the stationary layer (Fig 7a and Fig 7b). In case of dynamic fragment of flow, material is moving with high velocity and the calculation of the percentage of material occupied pipeline is subject to a large error. Flowing material causes glare and illumination resulted from light reflections in the acquired images, which are incorrectly detectable by algorithm as part of flowing material. In order to eliminate these erroneous phenomena it is necessary to at first improve the scene lighting and secondly reduce the impact of external factors by uniform illumination of the scene [19]. Additionally it is necessary to use higher data acquisition speed than currently used during measurements (30fps); as too long exposure time causes blurring the image, when recorded object is moving with high velocity. Another way to improve results is to develop algorithm with background removing techniques based on images acquired before capturing the flow process itself.

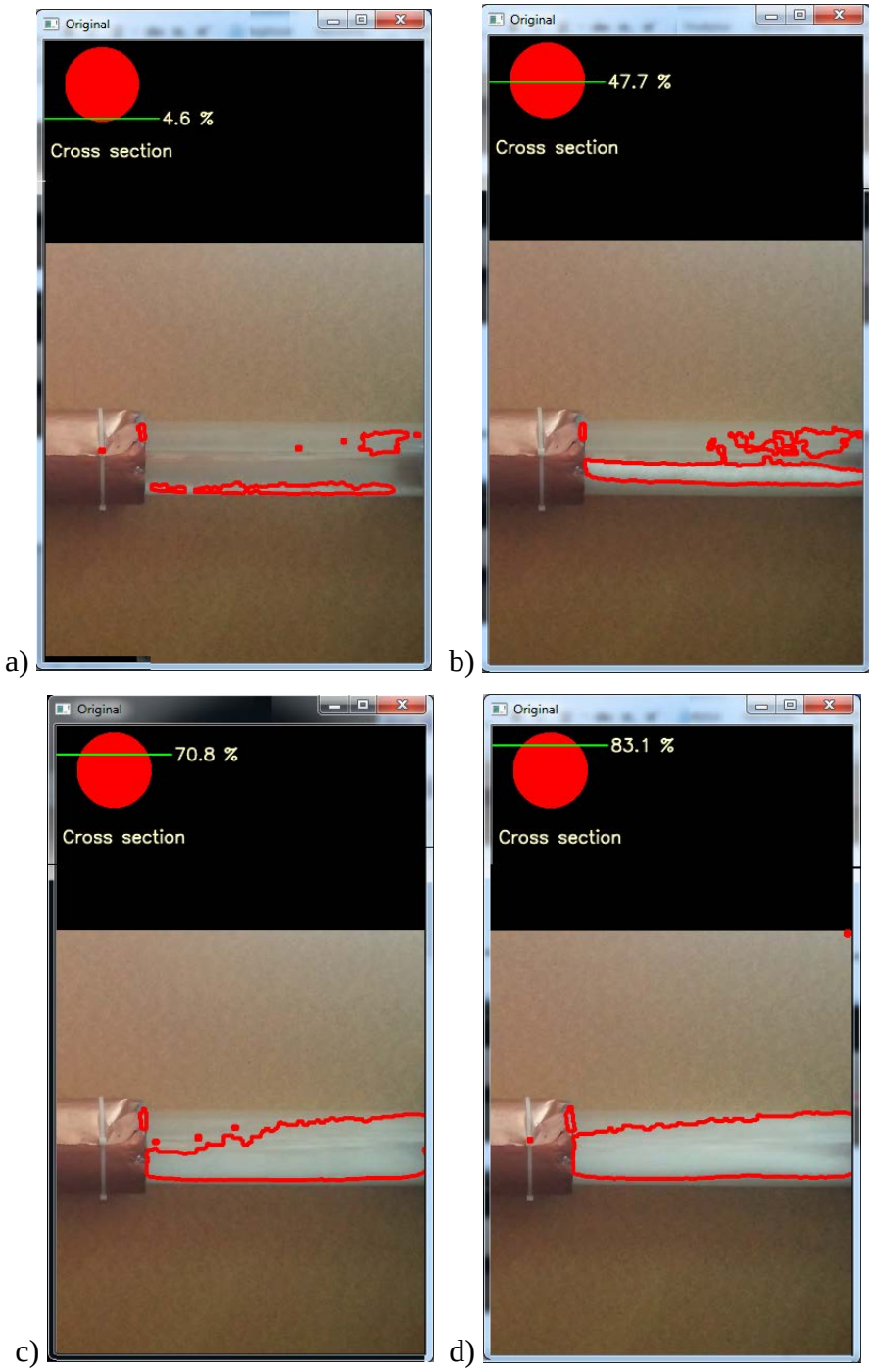


Figure 7: Examples of results for slug flow for material occupied pipeline with level of a) 4.6 %, b) 47.7% c) 70.8% d)83.1%.

Conclusion

In this paper we presented methods of image processing for measurement of the parameters of the slug phenomena during pneumatic conveying of solid with use of CCD camera monitoring. The preliminary results shown that proposed solution can be effective for determining the material percentage occupying the pipe volume for a given time moment. Additionally it is possible to compute the length of each slug as well as frequency and time break between the slugs appearing in a fixed time range. Yet, both the algorithm needs further

development as well as the images acquisition system requires additionally improvement of a lighting module and higher image acquisition speed.

References

1. Song D., Peng L., Lu G., Yang S., Yan Y.: *Velocity measurement of pneumatically conveyed particles through digital imaging*, Sensors and Actuators A 149 (2009) 180–188.
2. Rahmat M. F, Isa M.D., Hussin T.A, Sheikh U. U.: *Image Concentration Profile Generation Using Digital Imaging Technique*, International Conference on Intelligent Network and Computing (ICINC 2010), pp. 74-78, 2010.
3. Gao L., Yan Y., Lu G.: *Contour-based Image Segmentation for On-line Size distribution Measurement of Pneumatically Conveyed Par*ticles, IEEE International Instrumentation and Measurement Technology Conference, Binjiang, pp. 1-5, 2011, doi: 10.1109/IMTC.2011.5944318.
4. Fiderek P., Wajman R., Kucharski J.: *Fuzzy clustering based algorithm for determination for the two-phase gas-liquid flows similarity level*, Przegląd Elektrotechniczny, 90 (2), 2014, 52–55.
5. Saoud A., Mosorov V., Grudzien K.: *Measurement of velocity of gas/solid swirl flow using Electrical Capacitance Tomography and cross correlation technique*, Flow Measurement and Instrumentation, vol. 53, pp. 133-140, 2017, http://dx.doi.org/10.1016/j.flowmeasinst.2016.08.003.
6. Zhang W., Dai C.: *Experimental research of flow structure behind a circular cylinder with wave surface by PIV*, Journal of Hydrodynamics, 18, 97–103, 2006.
7. Cater T.M., Yan S.D.: *On-Line Measurement Of Particle Size Distribution and Mass Flow Rate Of Particles In Pneumatic Suspension Using Combined Imaging and Electrostatics Sensor*, Flow measurement and instrumentation, 16, pp. 309- 314, 2005.
8. Niederreiter G., Strauss M., Sommer K., Herrmann H.: *Investigations on the formation and stability of plugs at dense-phase pneumatic conveying*, International Congress for Particle Technology Nürnberg, PARTEC, 2004.
9. Cao X., Liu J., Jiang N., Chen Q.: *Particle image velocimetry measurement of indoor airflow field: A review of the technologies and applications*, Energy and Buildings, 69, pp. 367-380, 2014, doi: http://dx.doi.org/10.1016/j.enbuild.2013.11.012.
10. Westerweel J., Elsinga G.E., Adrian R.J: *Particle Image Velocimetry for Complex and Turbulent Flows*, Annual Review of Fluid Mechanics, 45:1, pp. 409-436, 2013.
11. Zhang H., Liu M., Li T., Huang Z., Bo H., Dong Y.: *Experimental study on plug formation characteristics of a novel draft tube type feeder for vertical pneumatic conveying of coarse particles*, Powder Technology, 301, pp. 730-736, 2016, http://dx.doi.org/10.1016/j.powtec.2016.06.049.
12. Jaworski A.J., Dyakowski T.: *Application of electrical capacitance tomography for measurement of gas-solids flow characteristics in a pneumatic conveying system*, Meas. Sci. Technol., 12, pp. 1109-1119, 2001.
13. Romanowski A., Grudzien K., Chaniecki Z. and Wozniak P.: *Contextual processing of ECT measurement information towards detection of process emergency states*, 13th

International Conference on Hybrid Intelligent Systems (HIS 2013), pp. 291-297, 2013, DOI: 10.1109/HIS.2013.6920448.

14. Caicedo G.R., Marqués J.J.P., Ruíz M.G.,Soler J.G.: *A study on the behaviour of bubbles of a 2D gas–solid fluidized bed using digital image analysis*, Chem. Eng. And Process: Process Intensif., 42, 1, pp. 9-14, 2003.
15. Grudzien K.: *Visualization system for large scale silo flow monitoring based on ECT technique*, IEEE Sensors Journal, PP, 99, pp.1-1, 2017, doi: 10.1109/JSEN.2017.2707538
16. Abdolahzare Z., Mehdizadeh S.A.: *Nonlinear mathematical modeling of seed spacing uniformity of a pneumatic planter using genetic programming and image processing*, Neural Computing and Applications, pp 1–13, 2016, doi:10.1007/s00521-016-2450-1,
17. Soleimani M., Mitchell CN., Banasiak R., Wajman R., Adler A.: *Four-dimensional electrical capacitance tomography imaging using experimental data*, Progress in Electromagnetics Research, 90, pp. 171-18, 2009.
18. Banasiak R., Wajman R., Jaworski T., Fiderek P., Fidos H., Nowakowski J.: *Study on two-phase flow regime visualization and identification using 3D electrical capacitance tomography and fuzzy-logic classification*, International Journal of Multiphase Flow, 58, pp. 1-14, 2014.
19. Fiderek P., Kucharski J.: *A Computer Vision System for On-Line Two-Phase Gas-Liquid Flows Recognition Using Fuzzy Methods*, Automatyka, 15/3, pp. 505-519, 2011.
20. Rymarczyk T., Szumowski J., Adamkiewicz P., Duda K., Sikora J.: *ECT Measurement System with Optical Detection for Quality Control of Flow Process*, Przegląd Elektrotechniczny, 92, 12, pp. 157-160, 2016.
21. Grudzien K., Chaniecki Z., Romanowski A., Niedostatkiewicz M., Sankowski D.: *ECT Image Analysis Methods for Shear Zone Measurements during Silo Discharging Process*, Chinese Journal of Chemical Engineering, Volume 20, Issue 2, 2012, Pages 337-345, ISSN 1004-9541, http://dx.doi.org/10.1016/S1004-9541(12)60396-6.

Piotr KRUPSKI[1], Michał AFTYKA[1],

The influence of commutation overvoltage on the gate voltage in the Push-Pull plasma power supply

The paper presents the impact of commutation overvoltages occurring at the transistor's collector-emitter juncture on the distortion of gate voltage of IGBT converter with push-pull topology, powering a plasma reactor. The voltage distortions occurring in the transistor gate circuit have a great significance, causing the possibility of weakening the transistor's dynamic parameters, as well as its partial switching-on and increasing its thermal exposure.

1. Introduction

Highly-efficient AC/DC/AC converters are more and more commonly applied in numerous applications of the plasma-chemical process. They enable adjustment of the voltage waveform, as well as the current, voltage, power and frequency regulation, for the purpose of a given plasma process. These capabilities of appropriately designed energy and electronic circuits used as power supplies of discharge units are characterised by numerous advantages and properties which are unavailable in the devices with a matching transformer [3]. A push-pull converter can be utilised for powering a gliding arc discharge plasma reactor. A converter current control circuit is located at the primary winding of the transformer. According to this topology, the matching transformer comprises a primary winding with a central point. Overvoltages occurring in the converters necessitate devising a method which will protect the semi-conductor switching elements against them. The protection of transistors against overvoltages, regardless of its type, is essential for the longevity and reliability of a power supply, as well as due to the power losses. High amplitude of voltage occurring at unprotected gate connector may exceed the limit value of voltage and cause permanent damage [1,2,4,5]. What is more, they may lead to a partial uncontrolled turn-on of a transistor switch and cause undesirable thermal losses. The level of overvoltages is determined by such factors as the inductance of transformer windings and the dynamics of current changes within. Powering plasma reactors usually requires high voltage; thus, the matching transformer must have a high turn ratio [3]. Both these properties significantly influence the overvoltages' value. The literature contains numerous descriptions of methods and systems for mitigating overvoltages [1,2,4,5,6,8,10,13]. In this paper, investigation of the commutation overvoltages occurring at the transistor and their impact on the gate voltage waveform distortion is presented. Mitigation of short-term overvoltages occurring in the form of high-energy impulses is of special interest. The converter was manufactured in the push-pull topology and used for powering a two-electrode GAD (Gliding Arc Discharge) reactor. The resistor capacitors, in

[1]Piotr Krupski, Ph.D. student at the Faculty of Electrical Engineering and Computer Science, Lublin, Poland, Nadbystrzycka 38A, 20-506 Lublin, tel. +48 81 5384 289, fax: 48 81 5384 643, e-mail:piotr.krupski@pollub.pl

which overvoltages undergo oscillatory suppression, are extremely important because without them the energy of a single impulse would be too high to disperse. Appropriately selected resistor capacitors (RCs) provide the balance between the protection and efficiency of a power supply. A resistor capacitor, found in the shunting branch of the converter, is capable of reducing the value of overvoltage or decreasing its energy to the level which can be handled by transil diodes. The IGBT have the minimum threshold voltage $U_{GE(th)}$ with the approximate value of 4 V [16, 18]. The oscillating current in the collector circuit, resulting from oscillatory suppression of commutation overvoltages, leads to a bi-stable gate voltage in the transistor gate, due to magnetic coupling in the transistor gate circuit. A partial turn-on of a transistor may cause a discharge of the collector current I_c when a fully turned-off state is set by the driver. Such a transient state is unfavorable because the transistor operates in a resistive state and it leads to an increased temperature in the switch, possibly altering the main stream in the transformer. This paper presents a study on the impact of oscillatory suppression of overvoltages in the collector-emitter circuit on the amplitude of the gate voltage of a transistor.

2. Origin of Overvoltages

Overvoltages are defined as spontaneous increases in the voltage, which exceeds the nominal level [1, 2, 10, 14]. The most frequent causes for overvoltages include: external switching operations and internal commutation processes. The occurrence of a commutation overvoltage usually results from a high current slope di_t/dt in a branch of a circuit, comprising an element accumulating energy in a magnetic field. During the winding commutation, the energy is accumulated in all elements of the electric circuit, increasing the voltage.

The amplitude of an overvoltage depends on the value of inductance L and the current in the first derivative in time t [1, 2, 4, 5]. The semi-conductor switching elements are especially vulnerable parts of energy and electric circuits. Exceeding the voltage of a transistor over the limit value U_{GES} may lead to the ground fault of a coupling and its irreversible damage; hence, the need for its suppression. The internal overvoltages are caused by switching operations in the branches of circuits containing inductance. Because these processes are connected with the commutation, i.e. switching or changing the direction of the current – the term *commutation overvoltages* is commonly employed [1, 2, 5, 9]. Figure 1a presents an exemplary circuit diagram in which overvoltages occur during switching operations. Figure 1b shows exemplary current and voltage waveforms in the course of commutation, i.e. transferring the current from one branch to another. The slope of IGBT collector current I_C depends on voltage E_d and the circuit inductance of the commutated circuit L_O. While passing through zero, the current I_C rapidly starts to increase in the opposite direction with an unaltered slope di_t/dt. When the excess charge in the connector is exhausted, then a rapid limitation to the reverse transient current occurs from the I_{rrM} to a near-zero value. Afterwards, the voltage exceeding the maximum value U_{CE} occurs in the transistor; its value being equal to $L_k(di_{rr}/dt)$. The producers of bipolar transistors with an isolated gate introduce new models, with increasing values of maximum value U_{CES} reaching 6.5 kV, and $U_{GE(ON)}$ from 15 to 18V. At the end of the switching operation, the IGBT has a “tail current”, which does not exist in the case of the MOSFET structure. This tail is caused by minority carriers trapped in the “base” of the bipolar output section of the IGBT, which causes the device to remain turned on. Unlike a bipolar transistor, it is not possible to extract these carriers to speed up the switching, since there is no external connection to the base.

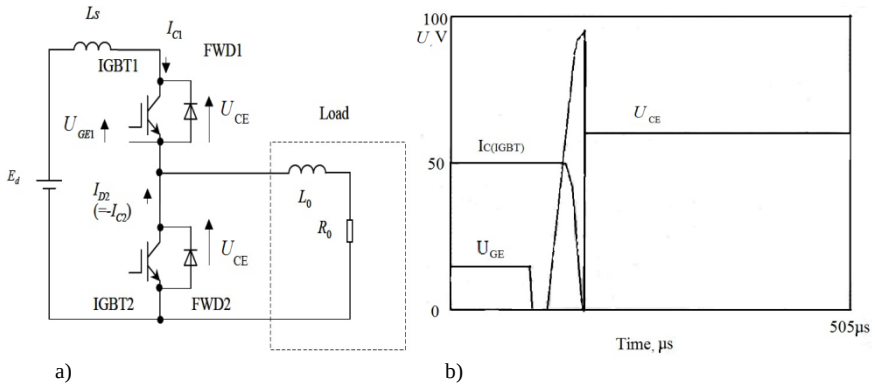


Fig. 1a) Diagram of a circuit with incident overvoltages, 1b) the waveform of overvoltages and current in the course of commutation [10]; $I_{C(IGBT)}$ – collector transistor IGBT current, U_{CE} – voltage of collector emitter transistor IGBT, U_{GE} – gate voltage,

Therefore, the device remains turned on until the carriers recombine. This "tail current" increases the turn-off losses and requires an increase in the dead-time between the conduction of two devices in a half-bridge circuit. In order to reduce the turn-off time, it is helpful to have a negative voltage (–5 V to –10 V) at the gate. When an IGBT is turned on, some overvoltages are generated on the gate terminal, due to the high d*u*/d*t* and parasitic capacitance between the gate and emitter. The overvoltage can cause a false turn-on for the bottom IGBT. A negative voltage at the gate enables to avoid this false turn-on trigger. Usually, 16 V is applied to the gate for turn-on and –8 V for turn-off. It is important to decide on the power requirement to drive the IGBT. When the suppressive oscillatory current flows in the circuits, the magnetic coupling occurs, resulting from the electric field between the circuit inductances. In order to prevent overvoltages, a number of solutions are employed, including: separation of circuits by increasing the distance between tracks, shielding, and shortening the circuits controlling the switches. In the gate circuits, i.e. the ones especially vulnerable to the exceeding of the maximum voltage amplitude, non-linear elements of semi-conductors are employed. When the characteristic voltage is exceeded, they rapidly attain the short circuit ability. Specialised drivers provide protection for the logic controlling the transistors and adjust the current of the gate overload. These drivers galvanically isolate the circuit from the transistor gate, usually beyond the optoelectronic part, and monitor the gate voltage. When an inappropriate control voltage occurs, the circuit adjusts the transistor gate potential, thus protecting it [1, 2, 4, 5, 10, 13].

3. Investigating the impact of commutation overvoltages on the transistor gate voltage.

An attempt to evaluate the correlation between the overvoltage within the collector-emitter coupling and distortion in the voltage of IGBT gate operating in a push-pull converter was made in the experiment. Owing to the application of two transistors controlling the primary winding current, very high efficiency can be achieved with relatively small dimensions of the converter. The topology of the circuit is presented schematically in Figure 2.

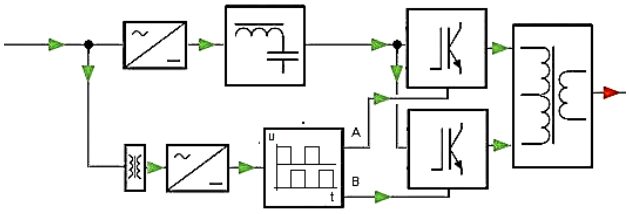


Fig. 2. Simplified diagram of a push-pull converter

A converter consists of two transistors T1 and T2 as well as a transformer in which the primary winding is powered by the rectified alternating current, filtered by means of the capacity. The transistor switches are controlled with two symmetrical square-wave currents, phased with appropriate dead-time. The discrimination of a simultaneous turn-on of both transistors is essential, because it may lead to the occurrence of two push-pull magnetic fluxes in the core. The maximum voltage for these transistors amounts to 1200 V. Additionally, the transistors are secured with two transil diodes, connected in a series, with the characteristic voltage equal to 410 V for each of them. The measurements are conducted for a push-pull converter with the apparent power over 1 kVA. The DC voltage with the value of 330 V is filtered by means of a choke and the capacity. The main switching frequency amounts to 14820 Hz. The converter operated with a transformer made on a 3C90 ferrite core with a single secondary winding that remains unloaded over the course of measurements. During the experiment, a digital oscilloscope is employed for the recording of the voltage and gate current waveform over the commutation period as well as the voltage waveform on the collector-emitter coupling in relation to the gate voltage of one of the IGBTs in symmetrical circuits. The waveforms in the function of time are superimposed on the charts containing a common Y-axis and two independent X-axes, due to different amplitudes of gate-emitter voltage U_{GE} as well as gate current I_G. The oscillogram in Figure 3 presents the charging waveform of gate transistor capacity. An increase in the current is visible at the moment of charging (t = 1-2 µs), occurring with an increasing U_{GE} voltage and significant negative current impulse at the moment of gate discharge when the U_{GE} voltage reaches a near-zero value (t = 7 - 8 µs). A clear lack of U_{GE} voltage distortion can also be observed.

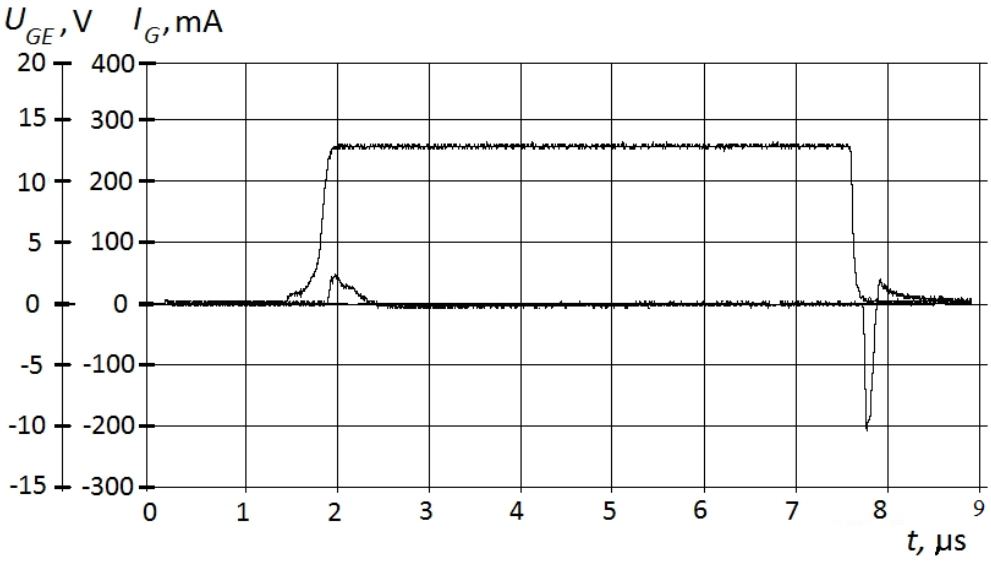


Fig. 3: Gate voltage U_{GE} and current I_G during switch processes.

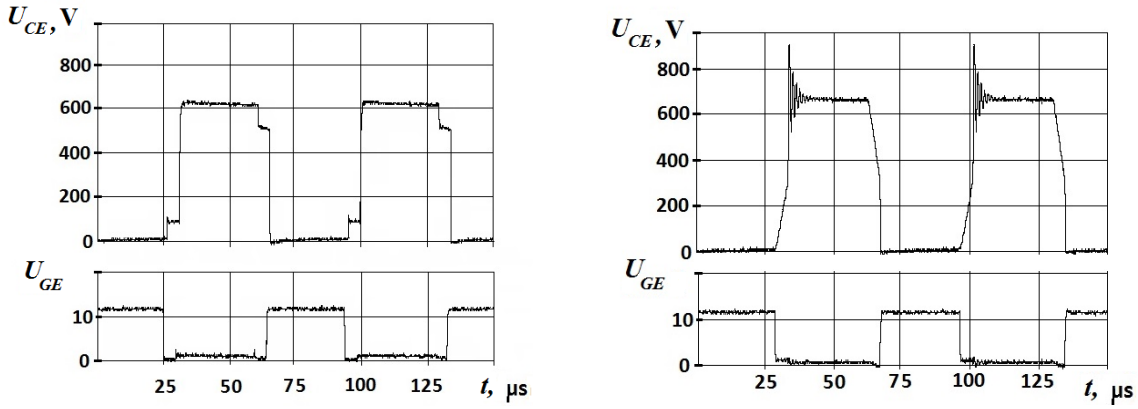


Fig. 4: U_{CE} voltage spike waveforms in relation to the control voltage U_{GE}: a) with suppressed overvoltage values at the collector-emitter, b) with oscillatory suppression of overvoltage at the collector-emitter to the level which is safe for transistors

The oscillograms in Figures 4a) and b) present the U_{CE} voltage waveforms in relation to the U_{GE} voltage. Two subsequent cases were presented: in the first one, the overvoltage on collector-emitter is completely suppressed by a system of RCs and on-load switches (Figure 4a); the second one shows an overvoltage suppressed to the level, which is safe for the transistors with a matched RC (Figure 4b). The waveform in Figure 4b) shows that the U_{CE} voltage is characterized by the oscillatory overvoltages, which are absent in Figure 4a). The oscillatory nature of the U_{CE} voltage distortions occurs at the same moment as in the case of U_{GE} voltage (Fig. 5, t = 14 -25 µs). Figure 5 shows that the amplitude of periodic oscillations of the U_{GE} control voltage approximates 4 V.

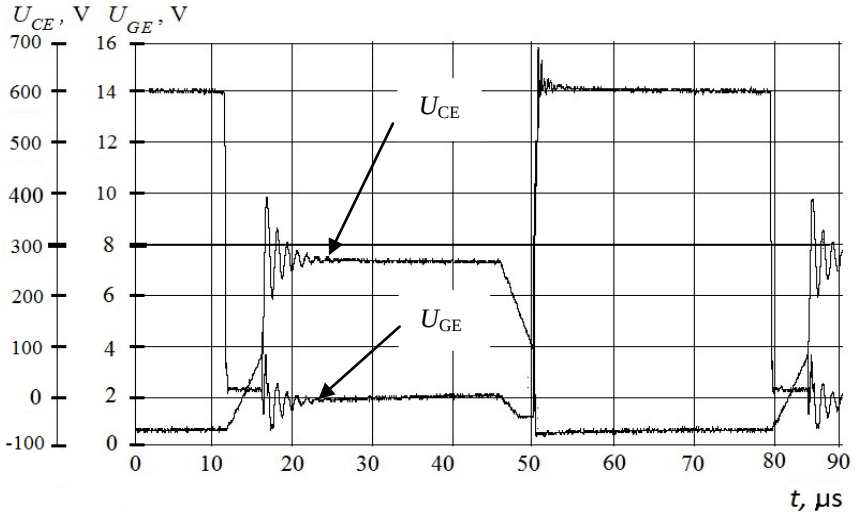


Fig. 5: Increased U_{CE} overvoltage waveforms on the converter transistor in relation to the control voltage with oscillatory overvoltage suppression in the collector-emitter to the level which is safe for transistors.

4. Remarks and conclusion

The aim of this paper is to indicate a correlation between the oscillatory suppression of an overvoltage in the collector-emitter of IGBT and distortion in the gate voltage. The threats connected with overvoltages and distortions in the gate potential during the operation of the transistor in a push-pull converter are presented as well. The observation of the current charging and discharging the IGBT is carried out. The acquisition of U_{GE} and U_{CE} voltage signals, which occurred at the transistor, is performed. The waveforms showing complete overvoltage suppression are compared with the ones subjected to oscillatory suppression. The waveforms 4a) and 4b) indicate that the oscillation suppressing the gate voltage occurs when the overvoltage amplitude in the collector-emitter is reduced by means of oscillatory suppression of RC. In a converter with hard commutation, the overvoltages occurring in the circuit of transistor gate are important, because their amplitude may exceed the turn-on voltage of a transistor. When there is no oscillatory suppression of high-energy overvoltages occurring at the collector-emitter, the gate-emitter voltage remains unchanged. Overvoltages may lead to a partial turn-on of a transistor and the occurrence of uncontrolled collector currents at the time when the transistor should remain completely blocked. Therefore, unintentional thermal losses may occur, thus lowering the efficiency of a plasma power supply. When the overvoltages last for a long time, it may lead to an extended period of transistor gate discharge and, hence, to other unintentional power losses. A longer period of discharge also prolongs the phenomenon of “tail current”. Thus, it is justified to apply the protection against overvoltages both in the collector emitter circuit of a transistor and in the gate. Due to the parasitic inductance in the gate circuits, an optoelectronic separation and shielding are recommended as well.

References

1. Horowitz P.: Hill W., *Sztuka elektroniki*, WKiŁ, Warszawa 2001, część 1 i 2.
2. Boksa J.: *Analogowe układy elektroniczne*, Wydawnictwo BTC, Warszawa 2007.
3. Stryczewska, H. D.: *Technologie plazmowe w energetyce i inżynierii środowiska*, Lublin: Wydaw. Politechniki Lubelskiej, 2009.
4. Wawrzyński W.: *Podstawy współczesnej elektroniki*, Oficyna Wyd. Pol. Warszawskiej, Warszawa 2003.
5. Barlik R., Nowak M.: *Poradnik inżyniera energoelektronika*, WNT, Warszawa, 2, 2013.
6. Toshiba.: *Power MOSFET Selecting MOSFFETs and Consideration for Circuit Design*., Toshiba Corporation 2016.
7. Borkowski A.: *Zasilanie urządzeń elektronicznych*, WKŁ Warszawa 1990.
8. Fuji Electric Co. Ltd.: *Protection Circuit Design chapter 5*, Fuji Electric Co., User’s manual May 2011.
9. Barlik R., Nowak M.: *Poradnik inżyniera energoelektronika*, WNT Warszawa 1998r. 359-393.
10. Mysiński W., Zabezpieczenia napięciowe i prądowe w obwodach automatyki przemysłowej, wykład 15.01.2014.
11. Abbate C., Busatto G., Iannuzzo F., Member, IEEE.: *High-Voltage, High-Performance Switch Using Series-Connected IGBTs*, IEEE Transactions on Power Electronics, vol. 25, no. 9, September 2010, 1-9.

12. Šoja M., Lubura S., Jokic D., Milan D. Radmanovic.: *Design and Realisation of Over-voltageProtection in Push-pull Inverters, Electronics*, Vol. 13, No. 2, December 2009.
13. Agilent Technologies,: *2.0 Amp Gate Drive Optocoupler with Integrated (VCE) Desaturation Detection and Fault Status Feedback*, Copyright © 1999 Agilent Technologies.
14. Szczerbowski R.: *Instalacje fotowoltaiczne – aspekty techniczno-ekonomiczne*, Przegląd Elektrotechniczny, ISSN 0033-2097, r. 90 nr 10/2014.
15. Mahesh Kumar A., Anantha Reddy B., Ramu B.: *Design of fast switching IGBT gate driver for high rated modules*, AIJRSTEM 13-148; © 2013, AIJRSTEM.
16. Hermwille M, Semikron.: *Gate Resistor – Principles and Applications*, Semikron Application Note International Gmbh 11/2017.
17. Lobsiger Y., StudentMember, IEEE, and Johann W. Kolar, Fellow, IEEE.: *Closed-Loop di/dt and dv/dt IGBT Gate Driver*, IEEE Transactions on Power Electronics, Vol. 30, No. 6, June 2015.
18. STMicroelectronic.: *AN4544 Application note IGBT datasheet tutorial*, © 2014 STMicroelectronics.

Adam ILNICKI[1], Mariusz RZĄSA[2]

Selected constructions of low-speed pneumatic motors

The work has pointed to low-speed pneumatic motors since constructions of such motors is much less than high-speed motors. The disadvantage of pneumatic motors is their relatively low efficiency, hence it is worthwhile to examine recognized constructions taking into account their efficiency improvement. The work presents design of several selected and patented pneumatic motors. Their constructions and working principles have been described as well as changes in structures have been proposed in order to improve efficiency of these motors.

1. Introduction

One of the methods of storaging and collecting energy is gas compression, usually air. Energy collected in this way can be used for pneumatic motor drive.

In this work a few modern solutions of low-speed pneumatic motors are described. The work presents some selected modern design solutions of low-speed pneumatic motors. The aim of this research is to draw attention to innovative pneumatic motor constructions as well as suggest a research method that enables quantitative and qualitative comparison of tested motors with power motor up to 3kW. Pneumatic motors are devices utilizing energy that comes from compressed air. The basic advantage of pneumatic engines is their reduced weight in relation to produced power with relatively high permissible overloads. Many pneumatic motor constructions are known, however, low-speed motors are definitely less in numbers than high-speed constructions. The use of high-speed pneumatic engines for powering machines and devices many a time requires applying speed reduction drives. Therefore, it is advisable to draw attention to low-speed pneumatic motors. The history of pneumatic engine design dates back to 17th century. The world's first pneumatic railway was invented in 1687. The most advanced development of mentioned engines started with the invention of a Polish engineer Ludwik Mękarski. In 1870, he created a pneumatic piston engine, which had been used in powering locomotives in mines as well as in trams of French cities such as Vichy, Nantes and Saint-Quentin [1]. Its principle was to use a check valve, which was actuated by the movement of the piston. (Fig.1).

[1] Ph.D. student, The Opole University of Technology, Faculty of Mechanical Engineering, ul. St. Mikołajczyka 5, 45-271 Opole, POLAND,
e-mail: a.ilnicki@doktorant.po.edu.pl
[2] Ph.D., The Opole University of Technology, Faculty of Mechanical Engineering, ul. St. Mikołajczyka 5, 45-271 Opole, POLAND,
e-mail: m.rzasa@po.opole.pl

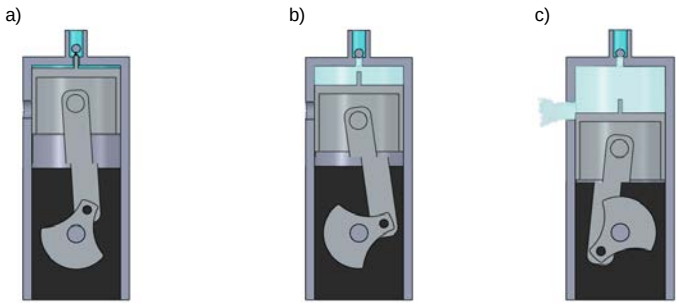


Figure. 1: Pneumatic engine operation strokes in a motor constructed by Ludwik Mękarski. a) suction valve open, b) power stroke, c) exhaust stroke.

A significant discovery, made by Robert Hardy in 1892, in the history of pneumatic motors was the invention of fuel pumping system which was activated during the braking process [2]. The idea was to use the engine as a pumping device during the braking process, which allowed some of the decompressed air to be compressed back during the deceleration of the vehicle. One of the latest pre-war achievements in this field of science was Johannes Wardenier's project, which according to reports, was the first engine not to be fuelled by liquid propellant. In 1991 Guy Nègre invented the dual power supply engine. That engine could be fuelled by compressed air as well as regular medium.

Nowadays, pneumatic engines are used for running-up combustion engines [3]. Small-sized pneumatic engines are commonly employed in power manual tools such as drills, jackhammers, sanders, wrenches [4], [5] etc.

Characteristic classification of pneumatic engine design solutions is found in fig. 2.

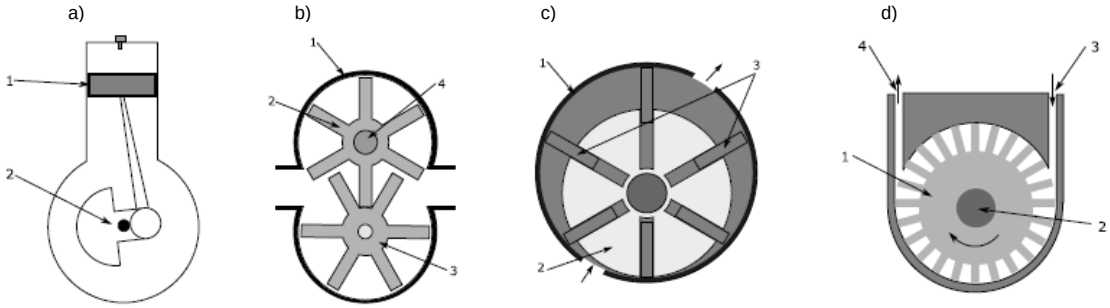


Figure. 2: Characteristic classification of pneumatic engine design solutions: a) piston, b) vane, c) paddle, d) turbine.

Piston-type motors contain working element which is a piston 1 (fig. 2a). An operation principle is based on an invention introduced by Mękarski. Piston-type motors are used with an operational power from 1,5 to 20 kW, and engine rotational speed amounts to 5000 rpm with supply pressure up to 0,6 MPa. Fig. 2b shows a paddle motor. These motors are applied in high power drives above 40 kW. Typical paddle motors (fig. 2c) work at speed of several thousand to several tens of thousands of rotations per minute, however, have low efficiency at low rotations due to the difficulty in sealing paddles. Turbine motors (fig. 2d) are applied at low loads but they can reach high rotation speed (of the order 500 000 rpm), for instance in dental drill drives.

2. Low-speed pneumatic motors

Nowadays there are few known low-speed pneumatic motors [6]. The simple design principle of pneumatic motors make them desirable structures as it significantly reduces cost of engine production and service. It is advisable to search for such structures because of ease with which compressed air can be collected and stored safely and the ability to store significant amounts of energy in this way. What is more, pneumatic motors are resistant to overload and they can be started promptly after stopping. They have high torque levels as well as produce no environmental pollutants. One of the advantages of pneumatic engines is that they can operate horizontally or vertically. Their operation mode can easily be reversed from motor into compressor. Pneumatic engines do not cause significant temperature rise, therefore, they can be applied in explosive environments such as: mines or wood, agriculture or food processing industries. The main priority in mining industry is safety so all drives need to be intrinsically safe. This requirement is completely fulfilled by pneumatic motors, and availability of compressed air in mine corridors is an economic justification for their use in running mining machines and devices. Presence of dust in food industry sometimes results in an explosive mixture occurrence for instance while threshing, transporting and sieving flour. The use of a pneumatic motor, both as a motor and a compressor, can take place in energy recovery systems such as vehicle braking process.

Broad spectrum of possibilities for pneumatic motors applications motivates searching for new construction solutions of such engines. They should guarantee achieving higher efficiencies than current solutions. It is advisable to quest for low speed pneumatic motors with high level torque, as this solution will not require the use of multi-step drive, which will significantly reduce costs.

The work presents three innovative and patented solutions of low-speed pneumatic motors based on patents: 214371, 218357 and 216801. Such solutions have been selected as they can operate at speeds of less than 1000 rpm and therefore they are classified in low-speed machine group.

A pneumatic motor in accordance with patent no. 214371

A pneumatic engine construction described under patent no.214371[7] has been shown in fig.3 The presented motor consists of a cylinder 1, inside of which there are air inlets 5 and air outlets 4. The operating element is a piston in a shape of cylinder 2 with three apexes arranged symmetrically on its perimeter. Additionally, in the housing there are two rotating sealants 3 of an adjusted shape so that they fit with the piston. The sealants are fixed coaxially in the housing, and their rotations are synchronized with those of the piston through the external gearing system, so that the angular velocity's ratio of a piston to the sealants equals 1/3 [8]. Four changing chambers can be distinguished in engine structure. Compressed air provided to the chamber A increases air pressure because of which a force causes the piston to rotate. At the same time in the chamber B decompression occurs through the outlet nozzle 3. Chambers C and D operate similarly. Work is performed simultaneously on opposite sides of drive shaft. Such solution provides regular torque during full working cycle as well as eliminates bending forces on the shaft. An appropriate sealant between piston apexes and the cylinder is a crucial element that has a direct effect on engine performance of such design. Greater matching of these elements (smaller amount of looseness) causes a significant increase in motion resistance, which in turn requires more intensive lubrication.

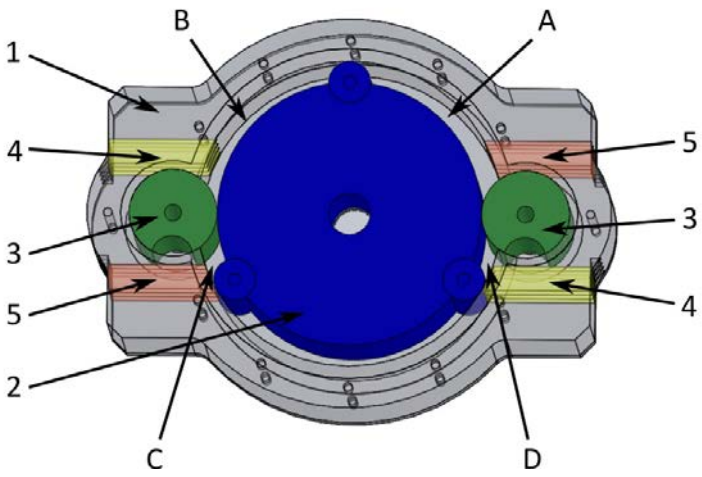


Figure. 3: Engine construction according to patent 214371

A pneumatic motor in accordance with patent no. 218357

Fig. 4 shows a construction of a pneumatic motor with the structure described in patent no.218357 [9]. It is the construction similar to the one shown in fig. 3, however, some parts have been modified. Air inlets 5 as well as air outlets 6 can be distinguished in the engine but slightly different fixed. In this solution the sealants 2 are element of a piston. The sealant between the cylinder and the piston constitutes rotating rollers 4. Due to the use of more rolling elements less resistance to motion has been obtained, which will decrease engine lubrication requirements and increase its durability.

The solutions (according to patent 214371 i 218357) are based on the principle of rotating piston, as it takes place in flow engines, however, sealing technique of rotors with the surface of the cylinder is an innovative solution. Sealing is based on a solution similar to that applied in vane motors but rotary thrust force is not applied to seal the system, therefore engines of this kind can operate at low rotational speeds.

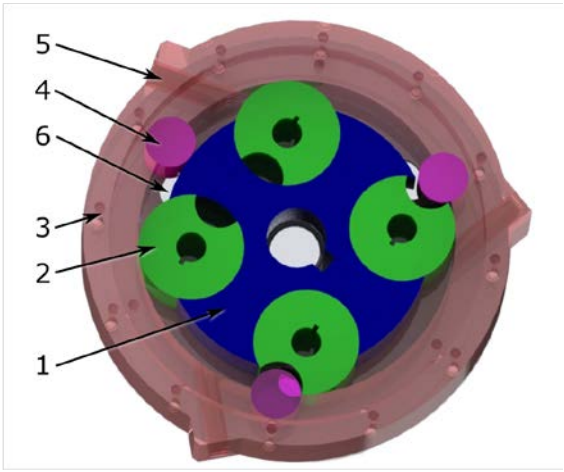


Figure. 4: Engine construction according to patent 218357

A pneumatic motor in accordance with patent no. 216801

Other solution is the engine described in patent no. 216801 [10], (fig. 5). Its operation principle is similar to the operation of a 2-stroke engine. Swing pistons have been applied in this engine. The main piston 1 and auxiliary piston 2. The pistons operate in pairs, performing a push-pull movement in toroidal segments of cylinders. In the cylinders, a variable pressure compartment is formed for the two pistons. For two pistons shared compression variable volume chamber is created in cylinders. The auxiliary piston 2 is synchronized with its counterparts in the subsequent cylinders through the mechanism 7. The structure consists of a shaft, two bushings, and pin chains.

The main pistons 1 and the auxiliary pistons 2 are attached directly to the rings 4 through connector 3.The cylinders are closed with a head 5 equipped with a suitable valve arrangement. The main pistons 1 are connected with crankshaft 6 through a wrist pin and a connecting rod.

The operation principle of a three-cylinder engine consists in push-pull operating of two pistons: the main piston and the auxiliary piston in one cylinder. Pistons coming closer together and moving away from each other create chambers of a variable volume. Once the main piston 1 in cylinder is placed in TDC-position, the auxiliary piston 2 approaches the main one as close as possible what results in minimum volume between pistons.

Next both pistons simultaneously move in this arrangement towards BDC-position for the main piston 1. At the time when the auxiliary piston reaches TDC-position, intake valve is opened and the cylinder is filled with working gas that exercise pressure on bottoms of both pistons. The power stroke starts. The main piston 1

at this moment is in BDC-position, and a crank of a shaft 6 is in most beneficial position to generate the torque. The torque from this piston is generated since the beginning of power stroke. The auxiliary piston 2 is in TDC-position and remains there motionless for some time, several degrees of crankshaft rotation 6. Then the torque is not generated from the auxiliary piston 2. After this time the torque is also generated by the auxiliary piston 2. Since then, both pistons generate the torque. Just as the main piston 1 reaches BDC-position, it remains there motionless for some time and the torque is generated only from the auxiliary piston 2. The outlet valve is opened just as the auxiliary piston 2 ends power stroke.

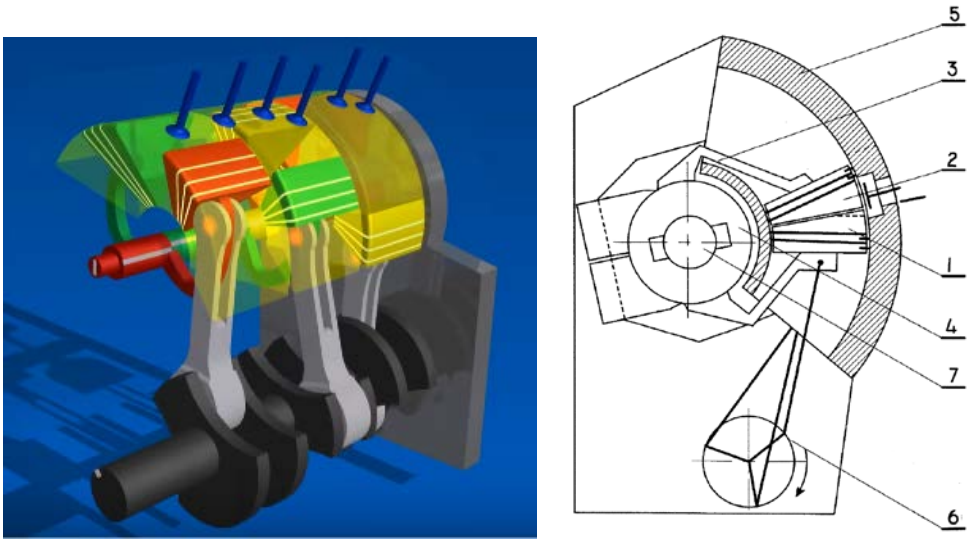


Figure. 5: Engine construction according to patent 216801

3. Results preliminary research

The laboratory tests of the pneumatic engine according to patent 214371 has been conducted and described in detail in paper [6]. The results are provided on the figure 6.

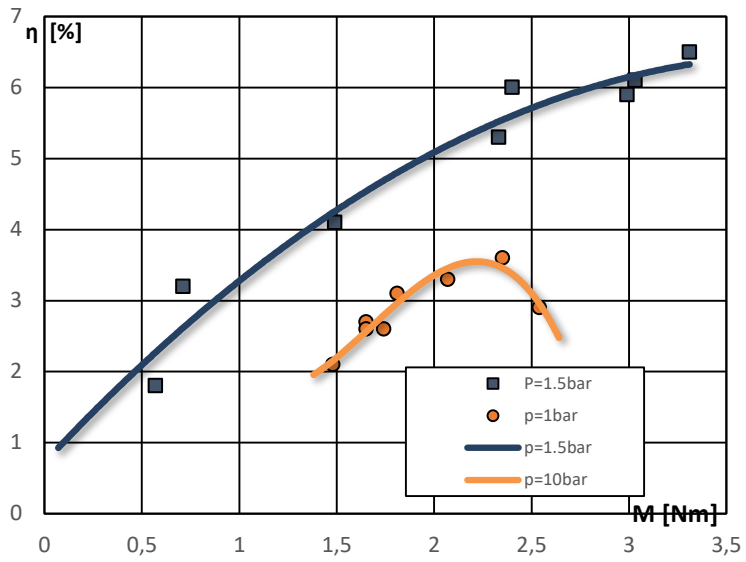


Figure. 6: Characteristic of efficiency relative to the torque on motor shaft in engine according to patent 214371

4. Summary

Relatively low efficiency is a characteristic of pneumatic motors that is why it is a parameter that often determines applicability of a given motor. In terms of engine efficiency, low-speed motors are highly influenced by sealing system and resulting from it resistance to motion. The assumption has been made that the basic parameters to be determined for a given engine are: total efficiency, power losses due to leaks and power lost due to friction of the sealants inside the engine. It has been assumed that the engine test stand should enable these parameters to be determined and be universal enough to examine described structures at the same stand. In order to meet these requirements, it is necessary to measure: power supplied to the engine in the form of compressed air, the power on the engine shaft, the engine pressure drop, engine rotational speed and the air flow delivered to the engine. Additionally, the engine test stand should allow measurement of power losses related with the leak and resistance to motion. The power range of tested motors should vary from 100 W to 3 kW. Research results should be used to identify space of largest power loss. It will be used to formulate structural changes of selected components in order to improve engine efficiency. Research should emphasize advantages and disadvantages of selected constructions. Although the main research aim is to evaluate the use of low-speed pneumatic motors for machine operating, while coming to conclusion it is necessary to take into consideration economic aspects of implemented changes.

In modern constructions of pneumatic motors constructors' attention is paid to increase their efficiency.

If there is an intention to apply these motors in devices being used nowadays, it is necessary to determine their actual operating parameters. Therefore, experimental research on engine prototypes is necessary. The method of testing pneumatic motors presented in this work allow to determine basic motor parameters essential for the design of machines operated by these motors. Such tests are necessary to assess the impact of selected engine construction elements on the engine's performance. The conclusions of such research will contribute to the design changes that improve operational data of pneumatic motors.

References

1. Wołoszczuk D. *Historia komunikacji miejskiej w Szczecinie (History of public transport in Szczecin http://www.swiatowy.org/bilety/ts_arch/zd_arch.html).* 2009

2. Schlager N., J. Weisblatt. *Alternative Energy. Volume I*. 2006

3. Śliwiński P. *Influence of Water and Mineral Oil on the Leaks in Satellite Motor Commutation Unit Clearances.* Polish Maritime Research 3 (95) 2017 Vol. 24; pp. 58-67

4. Trond Inge Eide . *Modelling and control of a pneumatic starting system for medium-speed gas engines* Norwegian University of Science and Technology, Department of Marine Technology, June 2011

5. Devang Marvaniaa, Sudhakar Subudhi. *A comprehensive review on compressed air powered engine*. Renewable and Sustainable Energy Reviews No 70. 2017 p. 1119–1130

6. Mishra KR, Sugandh G. *Study about engine operated by compressed air (C.A.E.): a pneumatic power source*. IOSR J Mech Civ Eng 2014, p. 99-03.

7. Ilnicki A. *Maszyna z obrotowym wirnikiem (Rotary machine - EP 2192265 A2 https://www.google.com/patents/EP2192265A2?cl=en).* Patent nr PL 214371.

8. Ilnicki A., M. Rząsa. *Nowe rozwiązanie konstrukcyjne silnika pneumatycznego*. *(New design solutions for a pneumatic motor)* Napędy i Sterowanie nr 4/2016

9. Ilnicki A. *Maszyna z obrotowym wirnikiem* (*A machine with a rotating rotor https://grab.uprp.pl/sites/Opisy/Miniaturki/Opisy/2306970.pdf).* Patent nr PL 218357.

10. Wójtowicz D., *Silnik wielocylindrowy zwłaszcza na sprężone gazy lub spalinowy wewnętrznego spalania ze zmiennym stopniem sprężania (A multi-cylinder engine especially for compressed gases or internal combustion with variable compression https://grab.uprp.pl/sites/Opisy/Miniaturki/Opisy/2436557.pdf).* Patent nr PL 216801

Konrad MIŚKIEWICZ[1], Robert BANASIAK[2], Maciej NIEDOSTATKIEWICZ[3], Krzysztof GRUDZIEŃ[4], Laurent BABOUT[5]

Automatic packing approach of mixed particles for DEM simulation of silo discharge process

Most of particle flow simulations use the sphere shape for particle boundary definition. Such solution is not effective enough for particles that have an ellipsoid shape. The paper describes new method for preparing the initial packing of different particle shapes inside a predefined space dedicated for DEM flow simulation. The developed algorithm allows to distribute ellipsoidal and spherical particle shapes. The implementation of the algorithm is accelerated using GPU parallel computation and multiprocessor environment.

1. Introduction

In many cases the discrete element method (DEM) is applied to prepare simulation of granular flow during discharge process in silo, as well as to compute the mechanical properties when the object is filled. For instance, the method may be used to investigate how particles may have influence on the design of different silo constructions. One key step of DEM is the implementation of the initial packing of object such like silo or box. Traditionally, particles are represented by spheres, while just few algorithms consider the packing of ellipsoid, a more general shape of particle representation.

As mentioned above, few types of algorithm have been invented to generate packing of spheres. One of them is dedicated to large scale simulations [2, 9]. It uses tetrahedral meshes, which can be generated by various software:Ghs3D, Gmsh or Netgen. This algorithm considers putting spheres at vertices and the middle of the segments connecting them. Then in the center of the tetrahedral meshes both: the position and the size of spheres are randomly calculated while making sure that all spheres have at most one contact point with adjacent spheres. Another known solution is to generate randomly sphere packing in the space [6]. When the particles fill the space enough, a so-called compression step starts. It consists of moving each particle to contact with others within given space boundaries. Yet another method is the of division the modeled shape to smallest problems and hence decrease areas, which will be filled with material [3]. When this step ended then started generating particle. Finally this process ends when the smallest problems are cloned to the original shape and hence all the modeled body is filled with the material. In addition to the above algorithms

[1] MsC., Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, tel: (+48) 42 631 24 43, e-mail: kmiskiewicz@iis.p.lodz.pl
[2] D.Sc., Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, tel: (+48) 42 631 26 89, e-mail: robert.banasiak@p.lodz.pl
[3] D.Sc., Faculty of Civil and Environmental Engineering, Gdansk University of Technology, ul. Stefanowskiego 18/22, 90-924 Gdańsk, POLAND, tel: (+48) 58 347 20 50, e-mail: mniedost@pg.edu.pl
[4] Ph.D., Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, tel: (+48) 42 631 27 50, e-mail: k.grudzien@kis.p.lodz.pl
[5] D.Sc., Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, tel: (+48) 42 631 27 50, e-mail: lbabout@iis.p.lodz.pl

there arethe basic methods [12] of packing shapes with particles as well, They are called Face Centered Cubic (FCC) and Hexagonal Close Packing (HCP).

The challenge of generating elliptical particle packing is different from known methods, because it uses computation of the contact between each object. However some algorithms use the tetrahedral meshes to set the position of ellipsoid and others have packed spheres in the standard FCC configurations before replacing the spheres with ellipses [1]. What is more, there is an evidence that DEM software can manage the packing of both types of shapes in the same space. For instance, the software packages which are considered in this paper to prepare the simulations, i.e.Yade [3] or WooDEM, do not include any algorithm to generate mixed materials with different geometrical properties (for example mixing spheres with ellipses).

The two software packages mentioned above use DEM to compute the motion and physical properties of each particle and their behaviors in different environments. They also integrate finite element method (FEM) to compute the structural strength or surface tension of object, as well as other physical properties. These DEM environments also consider the spherical mathematical model of object connection as a standard. In order to process objects with different shapes, this standard needs to be changed. Therefore, the purpose of this paper is to present a new mathematical approach that treats the contact of ellipsoid shapes with different aspect ratios in arbitrary packing. This extends the spherical model to be a more general case. Because this extension requires significant computational effort, the paper also discusses the strategy to accelerate data processing by using multiprocessor or GPU programming.

2. Mathematical equations

2.1. Detecting collision

An ellipsoid is an object which represents a geometric shape. This particle shape could be mathematically described by the following equation:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1 \tag{1}$$

where *a*,*b* and *c* are the principal semi-axes, and *a*>*b*>*c*.

Before checking that two ellipsoids have a contact point or are overlapped, one should prepare the equations for these two shapes in their canonical forms and define the matrices of positions and rotation for these two particles [10, 11, 8]. To detect the collision and intersection points between the two particles, one must solve the monic form of the characteristic polynomial [11] of two matrices which describe ellipsoids A and B defined by:

$$f(\lambda) = \det(\lambda A + B) \tag{2}$$

This equation has degree 4 and after normalization the $f(\lambda)$ *looks like this*:

$$f(\lambda) = \lambda^4 + a\lambda^3 + b\lambda^2 + c\lambda + d \tag{3}$$

where $a, b, c, d \in \mathbb{R}$.

The roots never have accepted 0 as a result. The vector have had values of roots and inform us about position between two particles. Two shapes for example ellipsoid and sphere are overlapped when in results are not had a positive roots. Otherwise this two particles are separated or was in touch.

2.2. Moveing particle

When it is detected that two ellipsoids are overlapped, the next step searches for the point on the surface of the other adjacent ellipsoids A(x_a,y_a,z_a) and B(x_b,y_b,z_b) which are connected with the centroid of each shape [4, 5, 7], as it is illustrated in Figure 1. To find the new position of point A the following expression is used.

$$\begin{bmatrix} X_{A'} \\ Y_{A'} \\ Z_{A'} \end{bmatrix} = \begin{bmatrix} X_{C_E} \\ Y_{C_E} \\ Z_{C_E} \end{bmatrix} + R \begin{bmatrix} X_A \\ Y_A \\ Z_A \end{bmatrix} \tag{4}$$

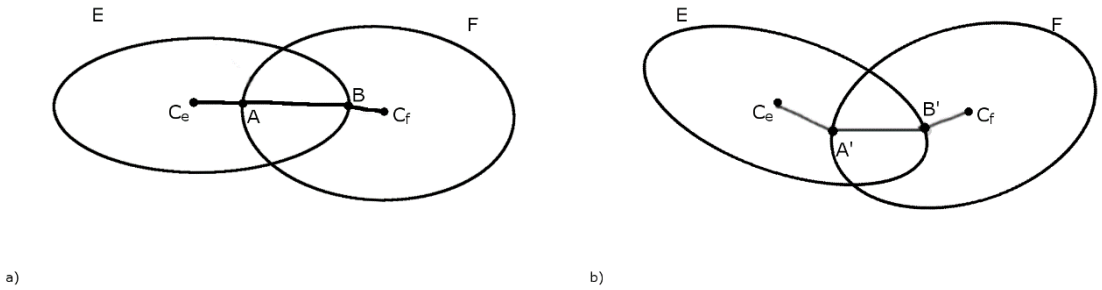


Figure 1: Simple schema of moving of two ellipsoids a) articles before rotation, b) and after rotation

where R is the rotation matrix and and C_E. is the centroid of ellipse E (see Figure 1). The same method is used to calculate coordinates of point B. Vector $\overrightarrow{\boldsymbol{AB}}$ is then calculated and used to move generated particle. It helped to separate objects and to assure that two material is notoverlapping in the simulation.

3. Algorithm

Some algorithms only generate spherical or ellipsoidal shapes of particle. The proposed algorithm computes a mixed packing that consists of these two types of shape. Granular materials are generating randomly, which causes this unique arrangement.

The particles are generated by the algorithm when users given basic parameters such as: the object to be filled (e.g. silo structure), the radius of the mono dispersed sphere and the semi axes of mono dispersed ellipsoid. Additional parameters such as: the percentage of the ratio between this two shapes, material properties and colour for recognize two different types of particle are accepted.

The flow chart of the algorithm is proposed in Figure 2. In the first step the main prism are calculated by the algorithm. This shape define the size of object which was being filled with material. . The subproblems are created from the orginal size of the main prism by the division to smallest one. Because this process are working on the three dimensional object, usually is corresponding to horizontal layers.. In this case the dimension of subproblem are calculated and is related to height the main object and the radius of sphere. Final of this value is two times of sphere radius. . When the problem is too small and there is no possibility to split it, the algorithm computes the original state. Two subproblems, which are connected together,have had the same boundary. On this plane the particle that in the end of process was saved in database can be generate. This process will used multiprocessors or GPU, because

the multitasking is the best way to compute the same problem in parallel. This method gives faster results, and the calculation platform uses all the computing power than the traditional way of serializable computation on one processor. In standard packages for example WooDem have had the methods to parallel computing. However this problem could be divided to subproblems and each boundary in the first step was computed by the same parallel process. In results are got the layers that was filled with materials. After that the parameters of each particle were saved on the same database . Naturally, the main process controls all subprocesses and waits for signal to carry out calculation tasks. When particles are generated at the top or bottom boundary, then the signal is sent to the main process to start computing – generating a sub process for top or bottom area. These two tasks can be started,

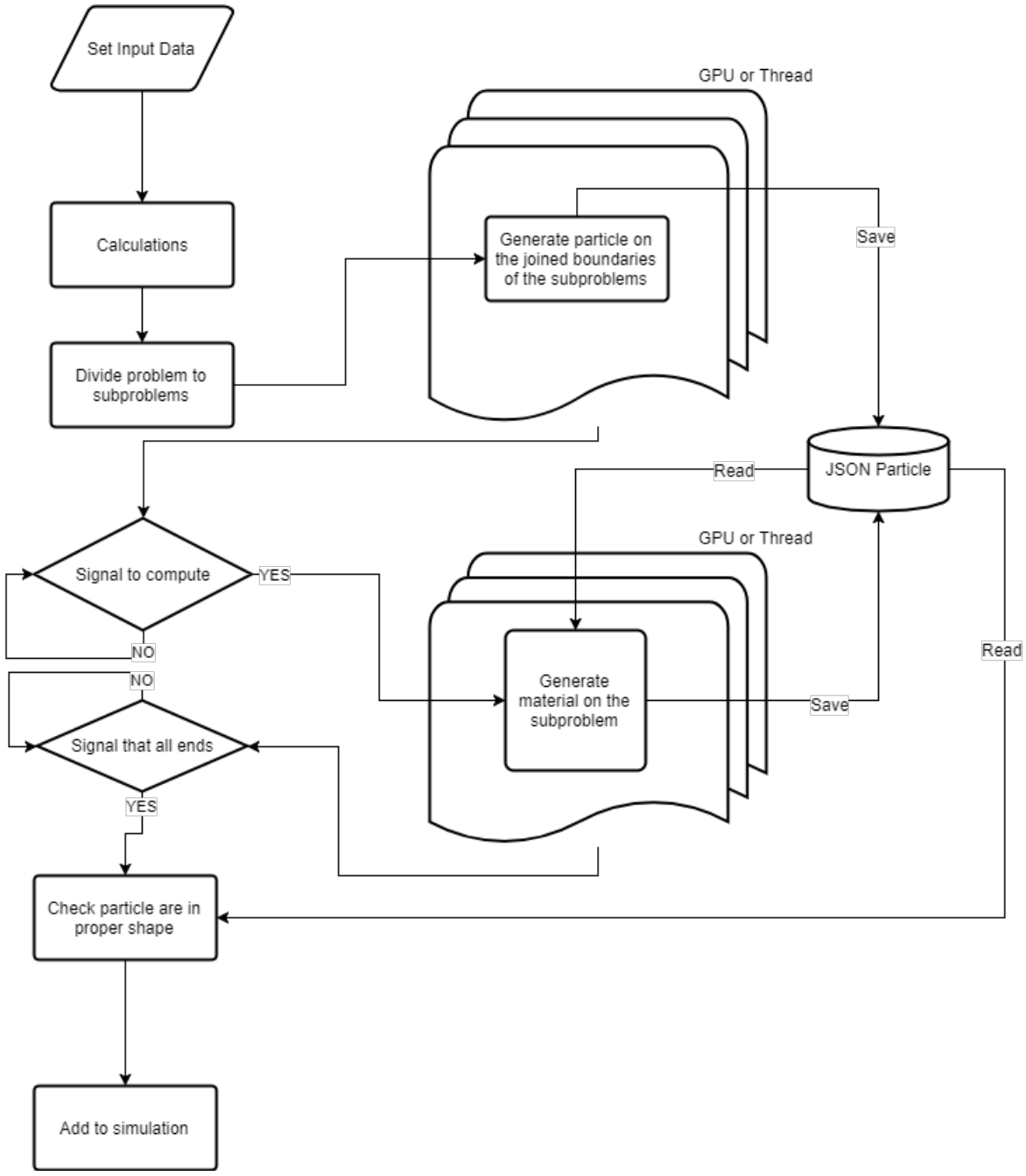


Figure 2: Schema of the algorithm to generate material on a shape

because they have one boundary in common with other subproblems (horizontal layers). Other subprocesses have two boundaries and to start the entire calculation the algorithm must know that all data have been saved on the database. The ellipsoids and spheres that was generated by algorithm, are saved in database. This place where was stored the information about particles is been one for two steps – generating on boundary and in area. The main process are received all signals from tasks. After that the last stage was started. It consists in reading all particles from database and checking which spheres or ellipsoids are located in the processed area. The object to be filled can be of any shape, e.g. hyperboloid, a square or a prism. In this research work, the prism shape has been used. It corresponds to a silo with flat edges. Naturaly, an algorithm checks if each sphere or ellipsoid stored in database has Cartesian coordinates that define in the object domain. When the computational process ends, the next step adds the matched materials to the simulation.

The process of filling the material consists in randomly generating the centroids of the spheres and ellipsoids, and the rotation matrix for ellipsoids. The rotation of each ellipsoid is described by the quaternion equation. This method is more natural to describe rotation in simulation. Compared to Euler equations, this solution gives simple composition and good efficiency. The percentage of ratio between two different types of particle is accepted by the algorithm. It is important for the user to know in advance which shapes are in larger amount. The materials are generated in a loop and for each iteration is calculated the ratio. This parameters is decided which type of particles must be created. When all data are generated by the algorithm, next step is to checked if no collision between particles occurs. The ellipsoid or sphere are moved or rotated in such a way that the problem is resolved, after that the algorithm is detected the overlap between this shape to other.

4. Conclusion

The state-of-the-art methods and algorithms for generating particle packing model do not give capabilities to mix particles with different shapes. The proposed algorithm allows to generate distribution of particles with ellipsoidal and spherical shapes and also to accelerate the computation. In order to prepare the particle shape distribution it is necessary to solve the collisions and particle overlapping problems. This task needs high computing power. The implementation of the proposed algorithm bases on parallel computing and significantly reduces computation time. This is crucial for further effective investigation of flow simulations. The proposed algorithm needs improvements in order to prepare simulations not only for different particle shapes, but also for different materials properties. Moreover, a next step of the algorithm should allow to prepare the packing of particles with irregular shape in defined space. Such solution will lead to modelling of the real shape of particles in DEM calculations.

References

1. Donev A., Stillinger F. H., Chaikin P. M., Torquato S.: "Unusually dense crystal packings of ellipsoids", Physical Review Letters, Vol. 92(25 I), 2004, pp. 255506–1.
2. Jerier J. F., Richefeu V., Imbault D., Donzé F. V.: "Packing spherical discrete elements for large scale simulations", Computer Methods in Applied Mechanics and Engineering, Vol. 199(25–28), 2010, pp. 1668–1676.
3. Kozicki J., Donzé F. V.: "A new open-source software developed for numerical simulations using discrete modeling methods", Computer Methods in Applied Mechanics and Engineering, Vol. 197(49–50), 2008, pp. 4429–4443.

4. Ligas M.: "Cartesian to geodetic coordinates conversion on a triaxial ellipsoid", J. Geod.,Vol. 86, 2011, pp. 249-256.
5. Ligas M.,Banasik P.: "Conversion between Cartesian and geodetic coordinates on a rotational ellipsoid by solving a system of nonlinear equations", Geodesy and Cartography, Vol. 60(2), 2011, pp. 145–159.
6. Lubachevsky B. D., Stillinger F. H.: "Geometric properties of random disk packings", Journal of Statistical Physics, Vol. 60(5–6), 1990, pp. 561–583.
7. Feltens J.: "Vector method to compute the Cartesian (X, Y, Z) to geodetic (φ, λ, h) transformation on a triaxial ellipsoid", J. Geod., Vol. 83(2), 2009, pp. 129–137.
8. Rubio-Largo S. M., Lind P. G., Maza D., Hidalgo R. C.: "Granular gas of ellipsoids: analytical collision detection implemented on GPUs", Computational Particle Mechanics, Vol. 2(2), 2015, pp. 127–138.
9. Jerier, J. F., Imbault, D., Donze, F. V., Doremus, P.: "A geometric algorithm based on tetrahedral meshes to generate a dense polydisperse sphere packing", Granular Matter, Vol. 11(1), 2009, pp. 43–52.
10. Wang W., Wang J., Kim M.-S.: "An algebraic condition for the separation of two ellipsoids", Computer Aided Geometric Design, Vol. 18, 2001, pp. 531-539
11. Jia X., Choi Y.-K., Mourrain B., Wang W.: "An algebraic approach to continuous collision detection for ellipsoids", Computer Aided Geometric Design, Vol. 28, 2011, pp. 164-176
12. Conway J. H., Sloan N. J. A.: “Sphere packings, lattices and groups”, Springer Science+Buisnes Media LLC, 1991

Adrian CHOJECKI[1]

Fuzzy logic toolkit for MCU development boards and Arduino-like programming environments: design and evaluation

The paper describes a new toolkit for implementation of fuzzy controllers and inferencing systems in Arduino-like boards and programming environments. An original solution has been included to accelerate some of the computation steps. The toolkit was tested with Arduino Uno (AVR) and TI Launpad (MSP430) boards. Prototyping boards with 8-bits and 16-bits, small, cheap MCUs have become very popular. The current, rapid development of modern computational intelligence and soft computing methods leads to attempts to implement them in a small MCU. One of the method is fuzzy logic, which is nowadays well known and validated in many applications, where traditional algorithms and control methods are less effective.

1. Introduction

At present, there are many microcontrollers (MCUs) prototyping boards available. The best known, *Arduino* [10] is popular among hobby electronics, but it also enters into academic applications. An easy and affordable programming environment and a large number of additional libraries available make that solution often chosen as a controller of a variety of devices, and even laboratory stations at universities [1].

The current, rapid development of modern computational intelligence methods leads to attempts to implement them in a small MCUs. Possible methods include: inference systems and controllers based on fuzzy logic, simple classifiers based on neural networks, regression-based learning algorithms, etc. The computing power seems sufficient enough for those applications. Unfortunately, there are currently no tools and libraries to support and accelerate this process. Each time is needed to write a specific code for the particular application. It would be very beneficial to use a library facilitating the transfer of a mathematical model based on the aforementioned methods into the source code. It would be useful to be able to use a description using one of the languages used for fuzzy logic, such as the FCL language described in the standard IEC 61131-7 [24] or the FIS file type used by the *Matlab's Fuzzy Toolkit* [14]. Such a solution is described in the next parts of this paper.

2. Problem definition

As mentioned above, fuzzy logic is quite well known, bringing measurable effects. Today, there are many tools available to enable its practical use in industrial and scientific applications. To simulate a system model and creating a rule base, the most popular *Matlab/Simulink* with the *Fuzzy Toolkit* [14] can be used. Other software are also available, including programs under open source license: *Octave* with *Fuzzy Toolkit* [15], *R* with *FuzzyR* [16], *FISPRO* [19] and

[1]M.Sc., PhD student at Lodz University of Technology, Department of Electrical Apparatus, Stefanowskiego St. 18/22, 90-924 Lodz, POLAND, tel. +42 631 27 72, e-mail: adrian.chojecki@p.lodz.pl

FISDET [20], described in [2] with interesting case of use. Also a *fis2fcl* [21] tool may be useful.

The implementation of the developed fuzzy systems can be done e. g. by the automatic source code generation with *Matlab/Simulink*, also a number of libraries are available to facilitate this task: *AwiFuzz* [11], *fuzzyTech* [18], *FuzzyLite* [17], *jFuzzyLogic* [25]. However, most of these libraries are intended for use on x86 architecture, with Windows or Linux. Admittedly, some of these tools are able to generate the code in C/C++ on the basis of the FCL description, but it is not suitable for direct deployment to MCUs. The author managed to find a *eFLL* [12] library for built-in systems, such as *Arduino*, but this tool is unable to load rules from the FCL file. So far, only min as t-norm, max as s-norm operator and only one method for defuzzification: center of area have been implemented. The utility is available under CC BY-ND 3.0 license, which unfortunately, despite the availability of source code, limits the possibility of its modification and further code sharing.

This toolkit was created during the author's work on a dissertation related to smart buildings. The fuzzy logic based system has proved itself very well when controlling the energy flow between the renewable energy source and the energy storage, taking into account price conditions and internal climate control. Unfortunately, initially the code was implemented directly in *Arduino IDE*, which made the changes difficult and very laborious. A large part had to be rewritten each time.

3. Fuzzy Logic Principles

Fuzzy logic is a multi-valued logic, which is a generalization of the classical Boolean logic. The concept of fuzzy sets was introduced by L. Zadeh [9], although as a precursor of a similar approach we may consider a Pole, Jan Łukasiewicz, and his three-valued logic [6].

Application of the fuzzy logic gives an ability to easily describe the processes and phenomena for which there are no strict mathematical models or on the basis of survey data it is impossible to accurately determine their status. As well, many systems can be easily described using natural language.

The practical use of the fuzzy logic principles to build FLCs (Fuzzy Logic Controllers) and decision-making blocks, as shown in Figure 1 consists of three basic steps [5, 7]: fuzzification, inference process and defuzzification.

The FLC can be connected in the same way as the standard controller in the feedback loop [7], as shown in Figure 2. The FLC can have MIMO, MISO, SIMO, SISO structure, so the number of inputs and outputs is not indicated in that figure.

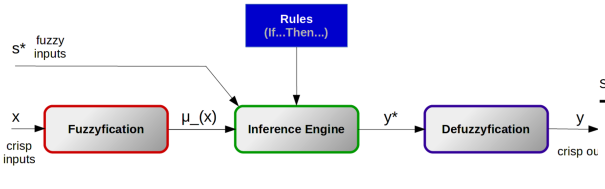


Figure 1: Block diagram of FLS (Fuzzy Logic System)

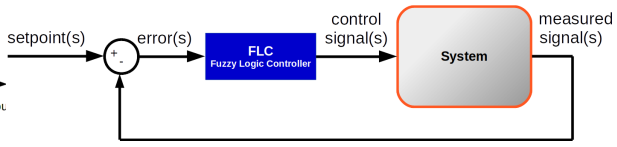


Figure 2: FLS as a closed loop controller

4. The toolkit

4.1. Embedded part

Programming environment for tested MCUs boards (Arduino [10], Energia [13]) use open source *gcc* compiler [22]: *avr-gcc* for *Arduino* and *msp430-gcc* for *Launchpad*. Those compiler forks support C and C++ languages, unfortunately not all functionalities are fully supported (only up to the C++ 11), but basic features for OOP (Object Oriented Programming) are possible to use. Therefore, was decided on an OOP approach. The class diagram was shown in Figure 3. The *Fuzzy* class is the main class that should be created in the Arduino sketch. It provides a basic interface consisting of three methods, directly corresponding to the steps shown in Figure 1. In the remaining classes, all parameters are initialized by default in source code based on the FCL description. Due to the limitations, it was decided to use arrays and basic data types. Fixed point types are used. A dedicated solution has been used to accelerate the multiplication based on a look-up table and use in some cases bit shifts instead of multiplication, which using much less MCU cycles.

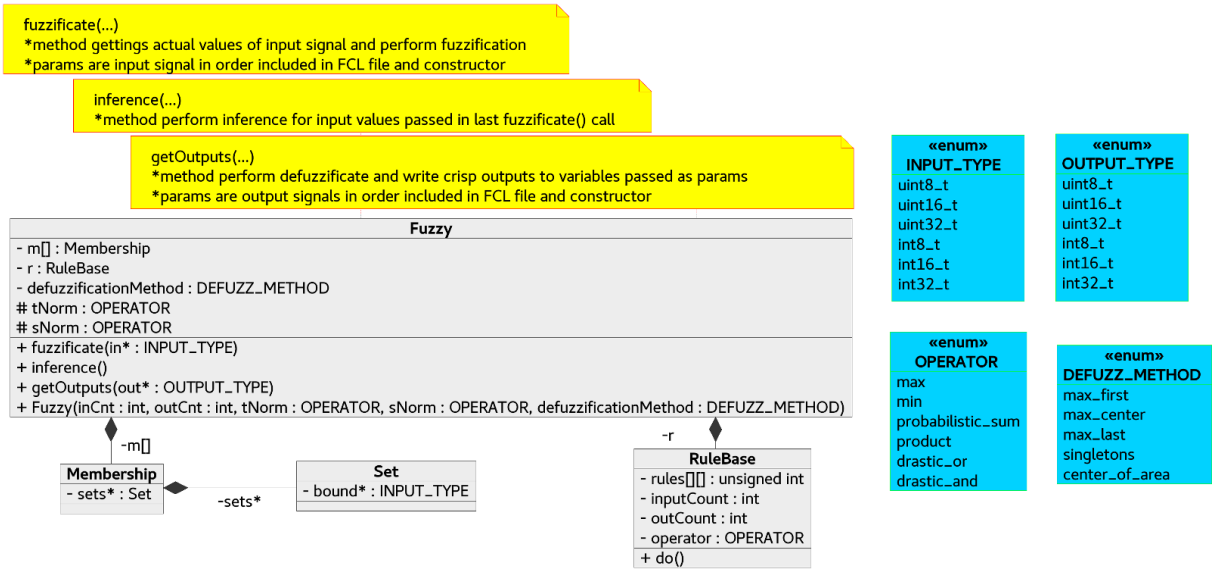


Figure 3: Class diagram

Some sample code pieces demonstrating the way of converting the FCL description (for the pendulum model described in the section 5.2.) into C/C++ code was shown in Figure 4. In this example, the variables are mapped to fixed point types, by default with bit size in accordance with microcontroller architecture (Arduino AVR – 8 bit, Launchpad MSP430 – 16 bit). There is a possibility of change the bit size by the user, maximum supported variables are 32-bit integers. Field names in classes available from the Arduino IDE level (*Fuzzy* – see Figure 3) are substituted in accordance with the FCL description. Not all names are substituted, it is assumed that the user will not edit the generated C/C++ code in order to make changes, but the FCL file and re-configured and re-generated the code with *fuzzyCFG.py*. The degree of membership during fuzzification process is determined with conditional statements and a set of segments of linear functions composing the shape of a given fuzzy set. Additional details can be also found in Figure 4. By default, a multiplication operation is used, but as mentioned previously, the author has also tested the use of bit shifting operations and look-up tables. For the Arduino and a controller for time-critical applications, with a small number of inputs/outputs (memory

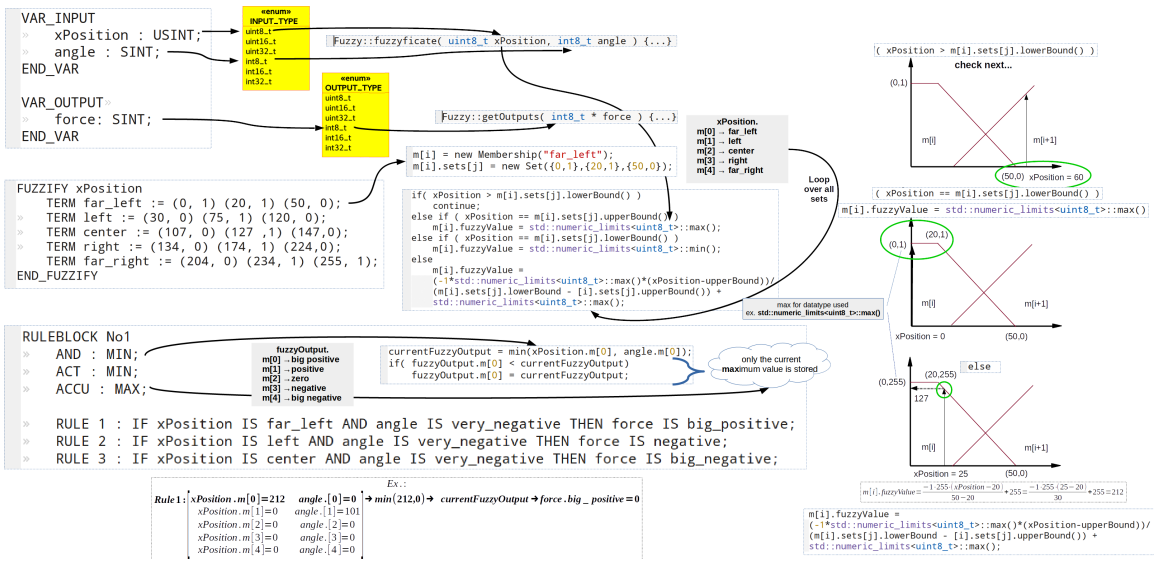


Figure 4: Example of C++ code parts generated based on FCL description and examples of computation methods

limitations for look-up tables case), these techniques allow for a significant acceleration of computation time.

4.2. Python part

The part responsible for automatic code generation is the script *fuzzyCFG* written in Python. After loading the FCL system description, it allows configurations of some parameters. Parameters such as the shapes of the membership functions or data types are included according to the FCL [24] file standard, but configurability allows the use of FCLs, e. g. prepared for PLCs, without the need for editing. The script works in console mode, has a simple menu created using the *ncurses* [23] library. Flowchart of *fuzzyCFG.py* is shown in Figure 5. In the first step (see left part of the Figure 5 - Steps 1,2), the user, having a prepared FCL file, loads it. The integrity of the file is verified and some parameters can be set. After selecting a board type, the source code generation begins. The developed script loads the subsequent parts of the FCL file corresponding to each stage of fuzzy control process (see right part of the Figure 5 - Step 3) and fills in the prepared templates of classes (Figure 3). For Python, a Codegen [26] package is available to facilitate automatic code generation, but in this solution it was not used because it did not have all the required functionality and a large part of the code was already prepared by author and rewriting it to use CodeGen was unnecessary. The result of this process is set of four files that should be copied to the Arduino sketch folder and *Fuzzy.h* should be included in main project file, the remaining files are linked together.

5. Verification and Results

5.1. Tip amount problem

The amount of tip depending on the quality of service and food is one example where the principles of fuzzy logic can be applied. The authors of the jFuzzyLogic [4, 25] have prepared a FCL file to solve this problem. This was considered an appropriate test case to help with the

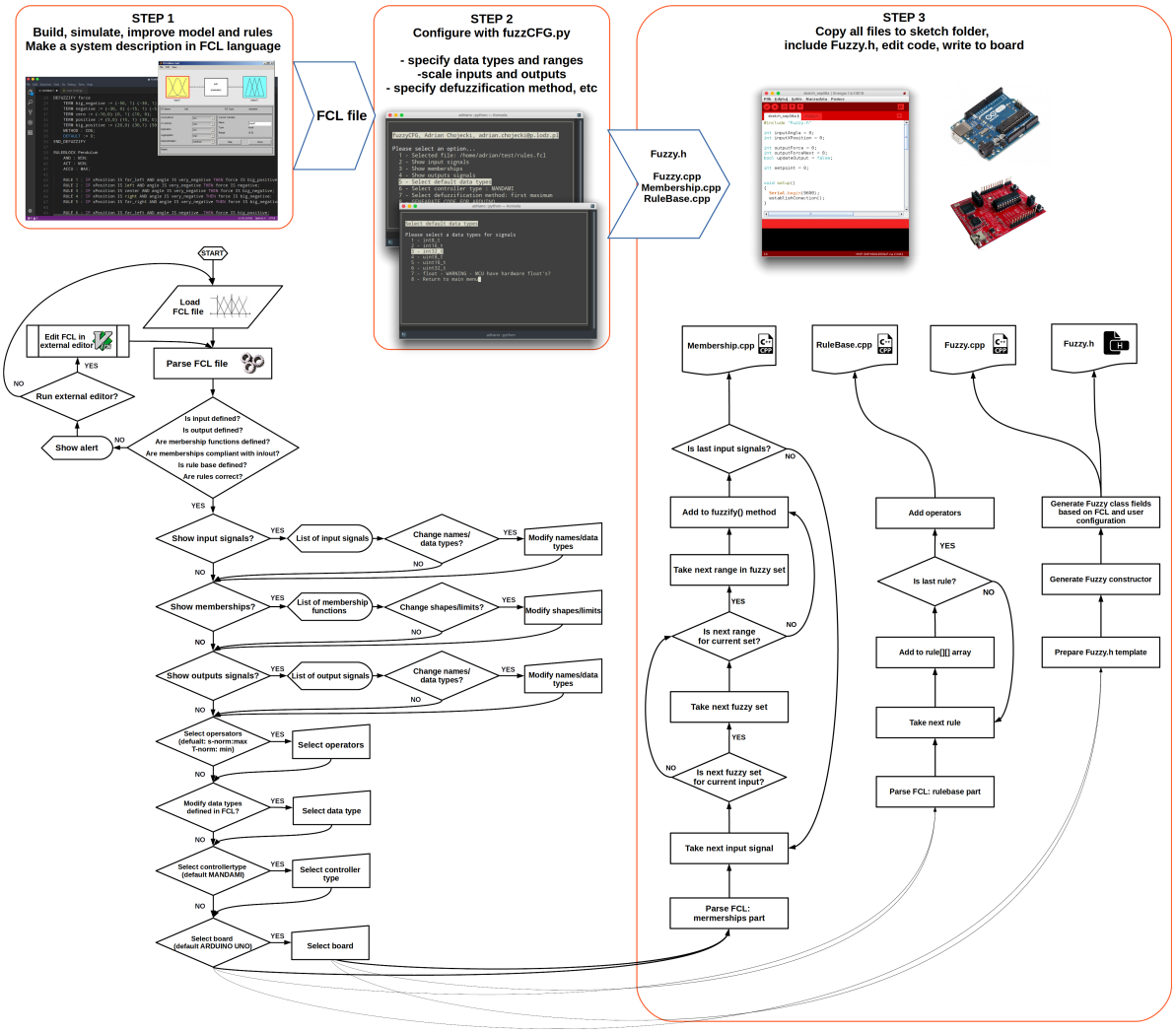


Figure 5: The workflow during the use of the toolkit

first verification. Results were convergent, average error for the same input data was 2.8 %. This discrepancy was probably due to the use of a integer types, where in *jFuzzyLogic* floating point types were used.

5.2. Inverted pendulum

Using the state space method, inverted pendulum simulations were prepared in the Processing [27]. For easy verification of the result, model and parameters presented in [8] has been used. The developed FLC had a MISO topology (Figure 6): position and angle were used as input signals, while the output signal was the force applied to the pendulum's cart. The rule base contains 25 rules. The communication between the controller and the object took place through a serial port, both for the MCUs boards and the PLC. The PLC Allen Bradley CompactLogix L3x (1769) with PID algorithm was used as reference. The view of the test stand is shown in Figure 7. An example step responses (angle) was presented in Figure 8.

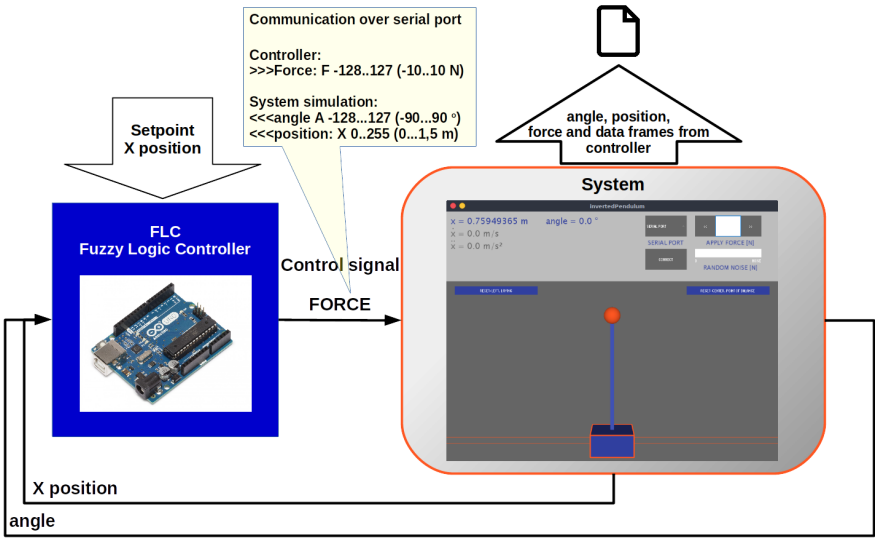


Figure 6: A schematic of the test system

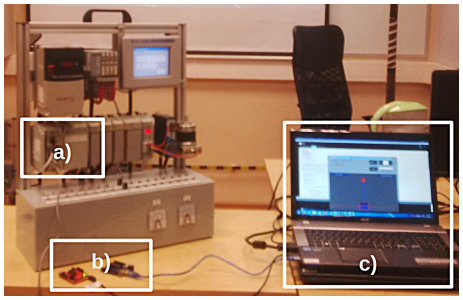


Figure 7: View of the test stand:
a) reference PLC controller
b) tested boards
c) inverted pendulum simulation on PC

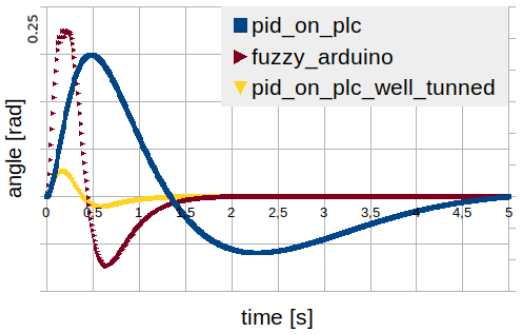


Figure 8: Step responses for inverted pendulum system

5.3. Energy management in smart buildings

It was the primary issue for whose work this toolkit was created. The system has 6 analog inputs, including: power consumption, state of charge, temperatures, etc. and 8 binary inputs for presence sensors and buttons. The rule base contains 31 rules, uses min, max operators. The membership functions are triangular and trapezoidal and defuzzification are performed using the first maximum method. It was also possible to create a cascade controller by creating more than one *Fuzzy* class instances. This system is the topic of author's possible future publication.

6. Conclusions and future work

Three test cases were used to check the results. The first one was used to validate the correctness of the implementation, the second one was well known in control engineering, time critical object which allows verification of the possibility of building a FLC, while the third one was a fuzzy inference system designed for energy management in smart buildings. In tested cases, the described library was working properly. It is suitable for both the inferencing sys-

Table 1: Features at current stages of development

Input file format:	FCL
Output file format:	CPP, H
Membership function's shapes:	triangle, trapezoidal, singleton
T-norm operators:	min, product, drastic
S-norm operators:	max, probabilistic sum, drastic
Defuzzification methods:	maximum(first, center, last); singletons; CoA (Center of Area)

tems and fuzzy controllers, where computation time is critical. The advantage of this solution is the possibility of quick implementation of the FCL system description. A potential limitation is only the memory size of the used MCU and the computation time, which is dependent on the number of rules and shapes of membership function.

For the aforementioned application in smart buildings, due to the large number of input–outputs signals and longer computation times, attention should be paid to reliability and robustness. A hardware watchdog clock [28], available in both tested boards can be used for this purpose. However, such functionalities will not be added to the described library, because it focuses only on fuzzy logic, but it can be implemented in the microcontroller program using the library.

Unfortunately, at the current stage of the development only two boards were tested: *Arduino Uno* and *TI Lauchpad G2*. The automatically generated code differs by some optimization aspects depending on the MCU utilized.

A further development of the library is planned. After resolving some legal issues and selection of the license, the author will make it available under open source (probably by the end of 2017). Presently, fine-tuning of the source code is underway and investigations are being performed to determine the time of execution in relation to the data types used, the membership functions shapes, the number of rules. A comparative analysis with an industrial fuzzy controller, taking into account the methods developed for accelerating calculations, will be the topic of another publication. The current features (Table 1) support type-1 fuzzy logic, and perhaps in the future an attempt will also be made to implement type 2 fuzzy logic [3] or simple neuro-fuzzy [5] algorithms.

References

1. F. A. Candelas et al., *Experiences on using arduino for laboratory experiments of automatic control and robotics*, IFAC-PapersOnLine, 48(29), pp. 105-110, 2015

2. G. Castellano et al., *The FISDeT software: Application to beer style classification*, 2017 IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE) 2017, pp. 1-6

3. O. Castillo and P. Melin, *Type-2 Fuzzy Logic: Theory and Applications*, Springer Science & Business Media, 2008

4. P. Cingolani, and A-F. Jesus, *jFuzzyLogic: a robust and flexible Fuzzy-Logic inference system language implementation*, IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE) 2012

5. C.T. Leondes, *Fuzzy Logic and Expert Systems Applications*, Academic Press, 1999

6. J. Lukasiewicz, *O logice trojwartosciowej*, Ruch Filozoficzny, (5), pp. 169-171, 1920
7. A. Piegat, *Modelowanie i sterowanie rozmyte*, Akadem. Oficyna Wydawn. EXIT, 2015
8. H. A. Suthar, K. Pandya, *Designing Fuzzy Logic Controller for Inverted Pendulum*, Second IFIP Conf. on Artificial Intelligence Applications and Innovations, 2005, pp. 61-68
9. L. A. Zadeh, *Fuzzy sets*, Information and Control, 8(3), pp. 338-353, 1965
10. Arduino. An open-source electronic prototyping platform https://www.arduino.cc/
11. AwiFuzz - control system library https://sourceforge.net/projects/awifuzz/
12. eFLL (Embedded Fuzzy Logic Library) https://github.com/zerokol/eFLL
13. Energia. Prototyping Software to Make Things Easy http://energia.nu/
14. Fuzzy Logic Toolbox https://www.mathworks.com/products/fuzzy-logic.html
15. The fuzzy-logic-toolkit https://octave.sourceforge.io/fuzzy-logic-toolkit/
16. Fuzzy logic for R https://cran.r-project.org/web/packages/FuzzyR/
17. The FuzzyLite libraries for fuzzy logic control http://www.fuzzylite.com/
18. *fuzzyTech* the software development tools for fuzzy logic http://www.fuzzytech.com/
19. FisPro 2013 https://www7.inra.fr/mia/M/fispro/fispro2013_en.html
20. FISDeT: Fuzzy inference system development tool https://github.com/Fisdet
21. fis2fcl converter https://sourceforge.net/projects/fis2fcl/
22. GCC the GNU compiler collection https://gcc.gnu.org/
23. ncurses 6.0 - GNU project https://www.gnu.org/software/ncurses/
24. *Programmable controllers: Part 7: Fuzzy control programming*, IEC 61131-7:2000 Standard
25. jFuzzyLogic http://jfuzzylogic.sourceforge.net/html/index.html
26. CodeGen - python code generation https://github.com/andreif/codegen/
27. Processing. Computer programming in a visual context https://processing.org/
28. Tushev S., Arduino and watchdog timer https://tushev.org/articles/arduino/5/arduino-and-watchdog-timer

Paweł KAPUSTA[1], Piotr DUCH, Tomasz JAWORSKI, Jacek KUCHARSKI, Krzysztof ŚLOT

Generative network input shaping for controlling visual object rendition in Adversarial Networks

To generate instances of learned classes by means of Generative Adversarial Networks - one of the recent developments in deep learning, one needs to identify appropriate distributions of input noise that drive the rendition process. The paper proposes a method for derivation of such distributions. We propose to build the considered vectors as a result of minimizing the Kullback-Leibler divergence over a set of inputs that correspond to manually selected class representatives. It has been shown that the proposed approach is superior with respect to other possible strategies in terms of not only subjective assessment, but also in terms of quantitative measures (Hausdorff distance was used as an assessment criterion).

1. Introduction

Generative Adversarial Networks (GANs) [1] are capable of learning raw representations of concepts (e.g. of visual object categories), so they can be used to generate class instances, which are distinct from any of known training examples. Feasibility of creating artificial objects can be extremely attractive from the point of view of several application fields, such as Virtual Reality (to populate virtual environments), movies and computer gaming (to render virtual scenes or actors) or forensics (e.g. to render portraits of wanted criminals). Concept learning by GAN is stimulated by various random distributions that are fed as inputs to the generative component of the network. As a result, a specific learned category can be retrieved by applying a specific distribution of inputs to a trained GAN. A simplistic approach to enable control over generation of output categories would be to analyze a large volume of input-output mappings and manually select an input that produces the 'best' outcome. However, since only a single, subjectively chosen instance is used to represent a class, one cannot assure reasonable rendition results for distorted input distributions, which have to be applied if an ultimate goals of generative models - generation of diverse class instances - is to be attained.

The presented paper proposes another approach for constructing input distributions that induce generation of specific category objects and that enable concept generalization, i.e. provide generation of plausible class instances for distorted outputs. This is accomplished by considering sets of manually selected input-output mappings for each category. Having several input distributions that map to some particular output category, a prototype for this category can be derived. We consider two different methods for aggregating information embedded in input vector distributions. The first one is simple averaging, whereas the second one is derivation of a distribution that minimizes Kullback-Leibler (KL) [3] divergence over the considered input set. We show that the latter approach is superior both in terms of subjective assessment and in terms of the adopted quantitative criterion. The criterion we use is maximization of the Hausdorff distance between the prototype and input set members, which quantifies how large is the margin between the adopted solution and boundaries of a region populated by samples from the considered category.

[1]Ph.D., Institute of Applied Computer Science, Lodz University of Technology, ul. Stefanowskiego 18/22, 90-924 Łódź, POLAND, tel: (+48) 42 631 27 50, fax (+48) 42 631 27 55, e-mail: pkapust@kis.p.lodz.pl

A structure of the paper is the following: after a brief review of the related concepts, we present details of the proposed method and provide its experimental evaluation, based on MNIST digits database [4].

2. Generative Adversarial Networks

The primary objective of introducing Generative Adversarial Networks was to deliver a tool for visual object classification that is superior with respect to Deep Convolutional Neural Networks [5] [6]. The idea of the concept was to introduce a mechanism of competition between genuine and fake inputs that gradually refines network's intrinsic category models in response to increasing plausibility of artificially rendered objects.

GANs, of general architecture depicted in Fig.1, are composed of two functional units: a discriminative network, which attempts to learn to distinguish between genuine and fake inputs, and a generative network, which attempts to improve in mimicking the genuine inputs.

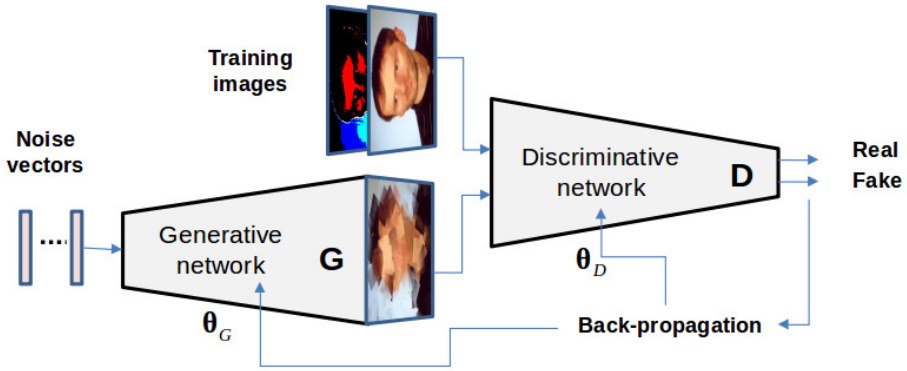


Figure 1: An architecture of Generative Adversarial Networks.

GAN training attempts to find parameters of both of its components, so that they satisfy the following min-max criterion:

$$J(\theta_G, \theta_D) = \min_{\theta_G} \max_{\theta_D} (E_{p(I)}(log(D(I)) + E_{p(n)}(log(1 - D(G(n))))$$

Parameters of the discriminative network (θ_D) are to provide maximization of the cross-entropy over outputs, generated by a discriminative component (D) fed with input images I of some distribution $p(I)$, and cross-entropy over fake inputs generated by a generative network (G), fed with input noise n of some distribution $p(n)$. Parameters of the generative network (θ_G) are supposed to minimize this quantity.

The original GAN concept assumed that the fully-connected feedforward-only, multiple-layer neural nets were used both as the discriminative and the generative components. This has been extended to the Deep Convolutional GAN (DCGAN) [6] paradigm, where significantly more complex architectures for both components were introduced.

3. Derivation of class prototypes for GAN

In the course of GAN training, the generative network learns to transform its random input distributions into objects that attempt to match instances of genuine categories that are provided

to the discriminator. After completion of this process, the generator can render objects of different appearance, some of which belong to the learned categories, depending on a particular form of input. Distributions that result in generation of some desired outcome need to be determined experimentally, by feeding the generator with diverse input vectors.

To enable generation of objects of a specific category one needs to identify a distribution, which provides a required input-output mapping. In addition to that, this distribution should become a basis for generating also different instances of the same class. To accomplish this goal, the distribution should be derived in a careful way, rather than selected randomly from a set of acceptable candidates. A procedure that is proposed, adopts this strategy: we derive a distribution that would become a class prototype from a set of input vectors, which provide correct input-output mappings. Two different methods for input data aggregation have been considered. The first one is simple averaging of distributions, yielding for a given class i, represented by a set of input vectors $\mathbf{x}^i_j$, $j = 1...n_i$, the mean class prototype ($\mathbf{z}^M_i$):

$$\mathbf{z}^M_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{x}^i_j \tag{1}$$

The second method that we examine is to derive a vector that minimizes an average Kullback-Leibler (KL) divergence over training examples for a given class. Divergence between a d-element vector $\mathbf{z}_{d\times 1}$ that is to be determined, and some $\mathbf{x}^i_j$-th element of a set, which comprises input vectors that produce i-th class objects, is given by:

$$KL^j(\mathbf{z},\mathbf{x}^i_j) = \frac{1}{d}\sum_{k=1}^{d} p(z_k) log\left(\frac{p(z_k)}{p(x^i_{j,k})}\right), \tag{2}$$

A Kullback-Leibler prototype for representing objects of some i-th class is a vector that minimizes average KL divergence over all input vectors for that class:

$$\mathbf{z}^{KL}_i = arg\min_{\mathbf{z}}\frac{1}{n_i}\sum_{j=1}^{n_i} KL^j(\mathbf{z},\mathbf{x}^i_j) \tag{3}$$

To compare the two derived prototypes, a Hausdorff distance d_H is used. Hausdorff distance from a set $\mathscr{A}$ to a set $\mathscr{B}$ is defined as:

$$d_H(\mathscr{A},\mathscr{B}) = \max\min_{\mathbf{z}\in\mathscr{A},\mathbf{x}\in\mathscr{B}} d(\mathbf{z},\mathbf{x}) \tag{4}$$

where $d(..)$ is the Euclidean distance between points. In the considered case, we evaluate Hausdorff distance between a set of all input vectors that generate objects of a given class ($\mathscr{A} = \{\mathbf{x}^i_1,...\mathbf{x}^i_{n_i}\}$) and a single-element set, containing the derived input prototype vector (either the mean prototype $\mathbf{z}^M_i$ or the KL prototype $\mathbf{z}^{KL}_i$). As the prototype vectors reside within a region occupied by input vectors $\mathbf{x}^i_j$, Hausdorff distance measures the maximum margin from the prototype to the most distance point on a class boundary. Therefore, calculating Hausdorff distance to k farthest class points evaluates how far the prototype is from the boundary, which can be viewed as a measure of its robustness.

4. Experimental verification of the concept

Handwritten digits MNIST database [4] has been used as a source of data for evaluation of the proposed concept. A DCGAN implemented using TensorFlow library [2] has been used for learning ten different categories of data (digits). We assumed short, five-element input noise vectors that drive the generative network ($d = 5$). After DCGAN have been trained, a set of input-output mappings have been generated to enable manual selection of 'correct' mappings, which would form a basis for derivation of ten different input prototype distributions. Sample sets illustrating mixed, correct and incorrect renditions made by the trained DCGAN network, have been shown in Fig. 2.

Figure 2: Sample images generated by the DCGAN: random mappings (left), mappings considered to be correct (middle) and incorrect (right).

Having collected input vector sets for all ten digits, the aggregated prototypes defined by equations (1) and (3) were derived. As it can be seen from Fig. 3, a general composition of vectors is similar, although exact values of components are different.

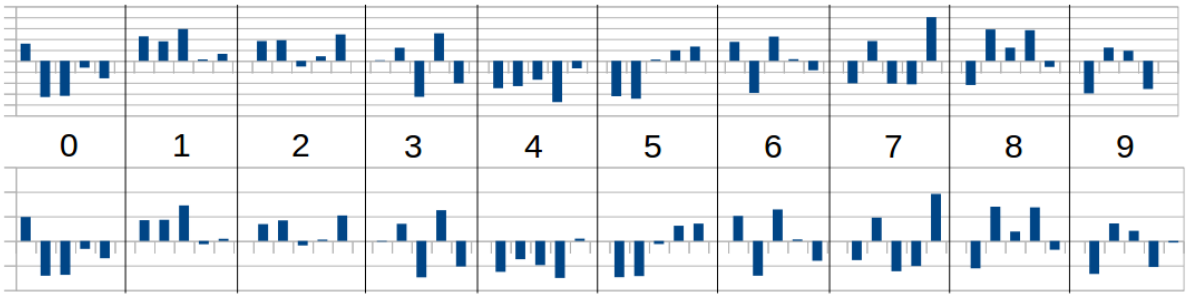


Figure 3: Prototypes derived for all ten digit classes using simple averaging (top) and using mean KL distance minimization (bottom).

Digits generated by the derived prototypes have been shown in Fig.4. One can observe that differences between mean and KL prototype components, albeit small, are sufficient to result in generation of diverse outputs. The main observation from the presented rendition results is that digits produced using KL-derived vectors are qualitatively better than the ones produced using simple averaging. In particular, 'zero', 'three' and 'four' produced by the KL-derived prototypes are of clearly better quality than their Mean-prototype generated counterparts. These qualitative observations are supported with Hausdorff distance measurements, presented in Table 1. Hausdorff distances were calculated for up to rank-3 (to the third most distant sample) and were summarized by a mean of the corresponding scores. It can be seen that prototypes generated by KL divergence minimization are in all cases superior to the prototypes produced

Table 1: Hausdorff distances for the prototypes derived based on KL-distance minimization (KL) and based on simple averaging (M). R1, R2 and R3 denote rank-one, rank-two and rank-three distances, respectively.

Digit	KL				M			
	R1	R2	R3	Average	R1	R2	R3	Average
0	1.5533	1.4408	1.2270	1.4053	1.4716	1.3319	1.2269	1.3435
1	1.1260	1.0116	0.9388	1.0459	1.1030	1.0715	0.9388	1.0378
2	1.0309	0.5832	•	0.8070	0.9921	0.5270	•	0.7595
3	1.0947	1.0061	0.8790	1.0072	1.0610	1.0399	0.8790	0.9933
4	1.2630	1.1732	0.9234	1.1193	1.2055	1.0242	0.9234	1.0510
5	1.0584	0.8541	0.7994	0.8871	0.9921	0.8637	0.7994	0.8850
6	1.2529	1.0183	0.9016	1.0511	1.0014	0.9764	0.9016	0.9598
7	1.2549	0.5557	•	0.9053	1.3136	0.6546	•	0.9841
8	0.8527	0.8187	0.6729	0.7675	0.8467	0.7600	0.6729	0.7599
9	1.2799	0.9807	0.9051	1.0710	1.3860	0.9056	0.9051	1.0656

by averaging. It follows that these prototypes should be more robust to distortions, enabling generation of a wide variety of correct class instances.

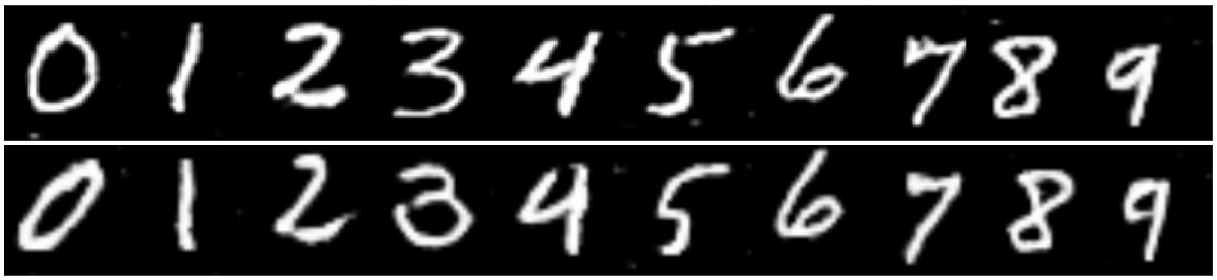


Figure 4: Digits produced by the generative network fed with KL-derived prototype input vectors (top) and with mean prototype vectors (bottom).

To verify robustness of the two considered prototypes to random distortions, a set of digit generation experiments was made. Random noise of magnitude increasing from 1% to 50% of the input dynamic range was added to both prototypes and the corresponding generation outcome was recorded. Results are summarized in Fig. 5.

5. Remarks and conclusion

The proposed methodology for derivation of input distributions that are used by generative component of Deep Convolutional Adversarial Generative Networks for generating specific category images proved to produce correct results. Of the two particular strategies for the prototype vector derivation: simple averaging of samples and optimization with Kullback-Leibler divergence criterion, the latter one proved to be superior. KL-derived prototypes have been shown to provide better representation of a class both in terms of subjective criteria and in terms of objective evaluation. Moreover, it has been shown that the derived KL-based prototypes are actually robust to severe random distortions, enabling correct rendition of underlying concepts.

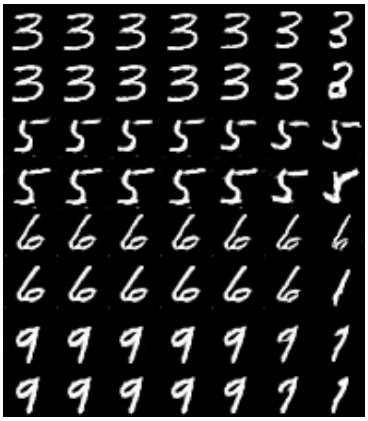


Figure 5: Generation of selected digits from distorted prototype vectors: derived using KL-procedure (odd rows) and using Mean-values (even rows). Noise level increases rightwards.

One needs to note though, that the results are preliminary and they need to be thoroughly verified using other datasets. Also, several details of the procedure, such as e.g. its performance on multimodal categories, need further careful evaluation.

References

1. Goodfellow I. J., Pouget-Abadie, J.,Mirza, M.,Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: "Generative Adversarial Networks", *arXiv:1406.2661 Freely accessible [stat.ML]*, 2014

2. Kim, T.: "Tensorflow implementation of Deep Convolutional Generative Adversarial Networks", *https://github.com/carpedm20/DCGAN-tensorflow*, 2017

3. Kullback, S.; Leibler, R.A.: "On information and sufficiency", *Annals of Mathematical Statistics. 22 (1), MR 39968*", pp. 79–86, 1951

4. Lecun, Y., Cortes, C.: "The MNIST database of handwritten digits", *http://yann.lecun.com/exdb/mnist/*, 1998.

5. Li, J., Liang, X., Wei, Y., Xu, T., Feng, J., Yan, S .: "Perceptual Generative Adversarial Networks for Small Object Detection", *arXiv:1706.05274*, 2017

6. Radford, A, Metz, L., Chintala, S.: "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks", *arXiv:1511.06434*, 2015