AIS 2018

13th International Symposium on Applied Informatics and Related Areas organized in the frame of Hungarian Science Festival 2018 by Óbuda University

PROCEEDINGS

November 8, 2018
Székesfehérvár, Hungary

ISBN 978-963-449-086-9

Big Data Overview and Connected Research at Óbuda University Alba Regia Technical Faculty

Éva Hajnal
Óbuda University Alba Regia Technical Faculty, Székesfehérvár, Hungary
hajnal.eva@amk.uni-obuda.hu

***Abstract*—21st century is called the data era. In this paper a wide overview is drawn about big data phenomenon. It includes the definition and main characteristics of big data, as well as the conceptual changes that is caused in the thinking and in the way of information gathering, processing and storing. Finally a research example illustrates the big data strategy and as a part of it, the data science methods. The intelligent data analysis of expert database systems is in the focus of this example namely the high education database research. A SOM based prediction model was developed, and its accuracy was investigated on five years real data.**

I. INTRODUCTION

The 21st century is called the data era. Big data means that during the industrial and social processes, in the sensor networks etc. large volume of data is generated with a very large velocity. The processing of these data gives large opportunities and creates interesting challenge to the data scientists as well as to the other sectors.

The computer technics development caused that the cost of the storage of one byte is millionth part less today as it was 15 years ago. The consequence of it the stored data volume started a rapid increasing [1], so by the predictions it will be 41 zettabyte stored data/year till 2020 (Figure 1). By this data explosion started the new data driven paradigm, the name of which is big data. This big data phenomenon has 3, 5 or 7 characteristic Vs [1]–[3](Figure 2). The volume means the amount of data. It can only be handled with help of special technological solutions from the special distributed file storage system (HDFS) to modern NoSQL database types via cloud solutions.

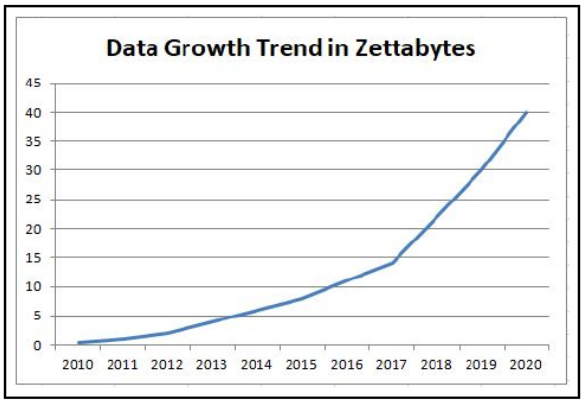


Figure 1 Trend and estimation of stored data volume per year [1].

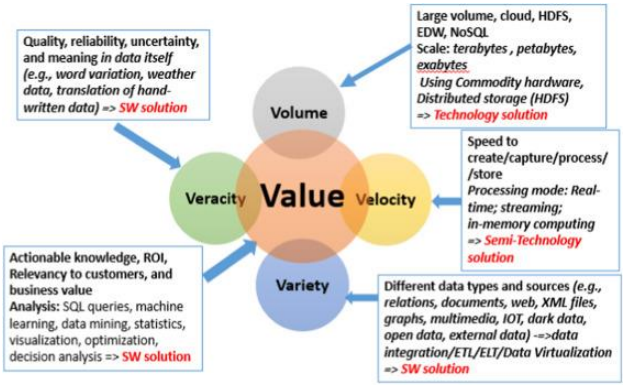


Figure 2 The 5 Vs of Big Data: Volume, Velocity, Variety, Veracity and Value [2]

The velocity of the data creation is large, that means a necessity of the adequate processing and storage. This can be achieved by technological solutions, but additionally needs careful design and appropriate software solutions. The data can be originated from different sources, as well as from internet applications, industrial processes, IoT devices, etc., and later the storage method also can differ and results in the variety of the stored data. In these cases the data integrating software solutions have large importance. Moreover the veracity of data is questionable and has to take care of attention to avoid the misinterpretations. The main feature is the value which can be yielded out from the data by data mining methods. This gives the added value of the big data.

The traditional relational databases after normalization have a performance limit [4], which caused the development of other database formats in this century (Figure 3).

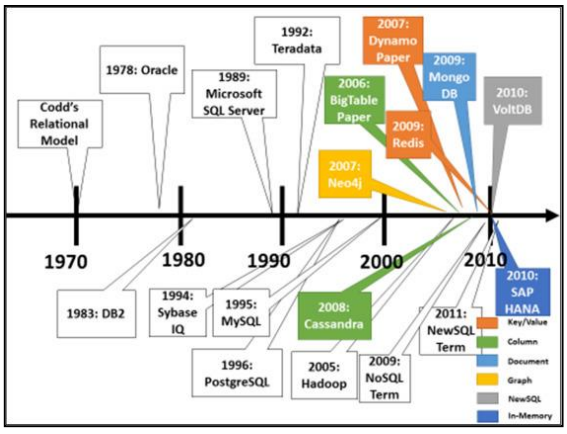


Figure 3 History and evolution of different database systems

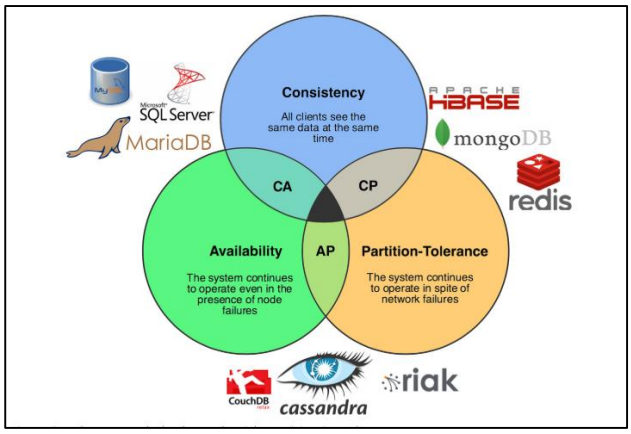


Figure 4 Database types grouped by the Brewer theorem [4].

The large amount of data requires partitioned storage for easy scaling, however traditionally we need to keep our data consistent and available in every moment. To reach this hoped aim is impossible as it was first drew up by Eric Brewer. According to his theorem we have to be satisfied with two from these three requirements and one point of view in choosing the right database system [1]–[4] is the place of it in this system (Figure 4). The lately developed NoSQL databases are in their childhood. According to the demand of them there is a need to develop them into a matured state. The main features of traditional relational databases and these NoSQL systems are compared in table 1. The schema less storage system gives the opportunity of NOSQL databases to process the large amount of continuously raising data. However the missing transaction tracking system makes the traditional relational databases the only acceptable solution in bank systems and other systems where the data tracking is extraordinary important.

The new data driven paradigm changes the point of view, the way of thinking and working of IT specialist, the decision-makers and finally of civils. This new point of view differs from traditional data management in several cases. One important new methodology is the whole dataset sampling, so we collect all data instead of the traditional representative sampling methods. After investigation of these big data we have to accept probability prediction without any reasons. In this new era the large amount of data makes impossible to a human expert to see the relationships between different things. In this situation big data supplemented with AI help can outperform experts [5]. The results of data mining is often in form of correlations, and decision makers have to accept them. This data driven paradigm means that all processes are supported with data analysis including the sentiment analysis as well. For this purpose either of the things can be the initiator of data creation, the name of which is datafication.

II. DATA MINING RESEARCH

During the last ten years a research group was created for data mining investigations. At first the name of the group was Mathematical modelling research team with the leadership of József Lakner and later Éva Hajnal, and this group was reorganized and renamed in 2018.

Table 1 Relational Versus Not Only SQL (NOSQL) Database Management Systems

Relational	NOSQL
Based on relations (normalized sheets)	Variable storage technics (column based, key-value pair, document based, graph databases)
Work with schema	Dynamically varying scheme
Vertically scalable	Horizontally scalable
Querying with standard SQL syntax	Querying with SQL like but not standardized syntax
Complex queries	Basic **C**reate **R**ead **U**pdate **D**elete operations
Transaction tracking	No transaction tracking
Before MySQL 8.0, classic Oracle, PostgreSQL	Apache Cassandra, MongoDB, HBase, Redis, Neo4j

Now it works as intelligent data analytics group. During the period a few research topics were investigated, as well as in basic and applied sciences. Our main partners are Hidrofilt Ltd., University of Pannonia and some multinational firms. The fields and the connected publications are shown in table 2.

Table 2 Scientific research and publications by field

Scientific research	Number of publications
Educational questions	5
Modelling, Cell automaton, ecological modelling	4
Phytoplankton and perifiton research	22
Data mining in economy	4
Handwriting recognition	4
Text classification	2

III. A SOM BASED MODEL

In this paper an intelligent data mining method and its usage is presented. Our purpose was to investigate the financial support and demographic data in Research and Development Sector including the high education. Based on the collected data set our purpose was a prediction with SOM about the expected number of students and the expected scientific results. These calculations were made and published five years ago [6], and now it is the time to investigate the accuracy of that prediction against the real data.

Methods

The Self Organizing Map is an unsupervised learning neural network, which was intended by Kohonen.

These types of neural networks do not get feedback about the output of the data processing. The aim is to find the optimal output based on the input. They consist of only one neuron layer and defines an input and an output space. The output space forms a map. It is important that the algorithm takes into account the places of neurons, consequently in successful data processing can preserve the original topology.

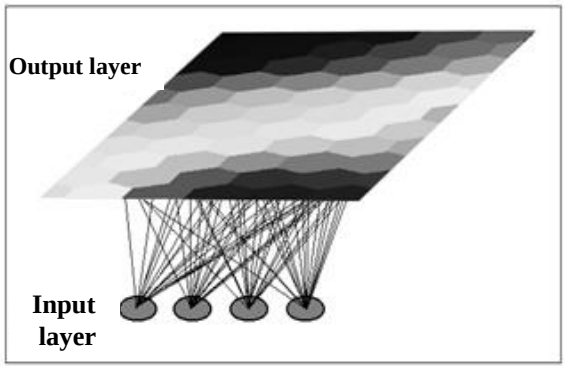


Figure 5 Topology of the Kohonen-type Self Organizing Map

The neighborhoods of neurons are respected during the data processing. The input neurons have weight vectors, which are responsible for the connection between the input and output data.

The steps of the learning algorithm are:

1. Initialization of neural network, number of neurons, weight vectors
2. An input data calculation, distance from every neurons.

$$(d_{ij} = \| x_k - w_{ij} \|), \quad (1)$$

where i and j are row and column indices of a neuron, k is the index of the input data d_{ij} the distance between the input neuron and the i^{th} and j^{th} neuron, x_k is the input data vector and w_{ij} is the weight vector of the i^{th} and j^{th} neuron.

3. The determination of the closest neuron, this is the Best Matching Unit (BMU)

$$w_{winner}(w_{ij} : d_{ij} = \min(d_{mn})), \quad (2)$$

where w_{winner} is the BMU.

4. Actualization of weight vector according to the distance and the learning rate.

$$w_{ij} = w_{ij} + a * h(w_{winner}, w_{ij}) \| x_k - w_{ij} \| \quad (3)$$

where a is the learning rate and h is the neighborhood function.

These 2.-4. steps are iterated till a previously determined criterium is reached.

The Self Organizing Map is an efficient clustering, correlation hunting and estimation method. It is widely used for optical character recognition, for face recognition [7] [8], for robotics to recognize barriers in the environment [9]. In the literature there are examples to scientific [10] and technical modelling [11] usage and financial analysis. However only few examples deal with data estimation or prediction with the help of SOM.

The theoretical basis of the estimation is to find the best matching units of an input data point, the calculation of the mean weight vector and denormalization.

During calculations standard setup (Euclidean distance, Gauss neighborhood functions, standard batch training algorithm, hexagonal lattice) and range type data normalization were used.

Data

Digitalized Data: Data were downloaded from the official website of the Hungarian Central Statistical Office: main ratios of R+D (1990-2012), staff number in R+D (1990-2012), financial sources of the R+D input (2000-2012), number of publications (1990-2012), R+D support (1990-2012), patent activities (2000-2012).
Paper based data: The R&D data before 1990 came from the Statistical Yearbooks.

Results and discussion

The collected dataset was validated and then some statistical calculations were made. Among them there were traditional statistical steps and the previously described SOM based model, which was used not only for correlation hunting, but moreover for prediction. Based on this model a 20 years long prediction was made to the total number of students in high education [6]. This prediction is presented on figure 5. The prediction seemed very pessimistic and thus got some critics. Additionally it is not based on widespread surveys, only on statistics. Consequently it had only statistical base and no detailed reasoning by each factors and effects.

After five years the first check of this prediction was executed against the Educational Office's data [12]. The absolute difference between the model and real data was calculated and presented in Table 3. The average accuracy of our model was -0,36%. The calculated prognostic accuracies that based on the predicted yearly changes and their difference from real yearly changes are presented in Table 4. In this system the average accuracy of prediction is 17% which is yet acceptable. However its changes are between -23% and 51% which in connection with the quite exact absolute accuracy means a bit of difference (1 year) in the temporal pattern.

Table 3 The accuracy of our five year old prediction [12][6], [13]–[15]

Years	Statistical number	Estimated number	Difference	Accuracy
2013	320 124	325 624	5500	1,7%
2014	306 524	304 362	-2161	-0,7%
2015	295 316	292 874	-2441	-0,8%
2016	287 018	286 122	-895	-0,3%
2017	283 350	278 617	-4733	-1,7%

It is concluded, that by the first five years the model is quite accurate and seems to be relevant in prediction. It is based on only few factors, so we interestingly wait its later accuracy. Generally the models are differing increasingly by the time. The social trends and social processes based on the behavior of a large multitude and their inertia cause a good predictability in social processes which was demonstrated by this example.

Table 4 Accuracy of the predicted yearly changes

Year	Real change	Estimated change	Difference	Accuracy
2013-14	-21262	-13600	-7662	36%
2014-15	-11488	-11208	-280	2%
2015-16	-6752	-8298	1546	-23%
2016-17	-7506	-3668	-3838	51%
				17%

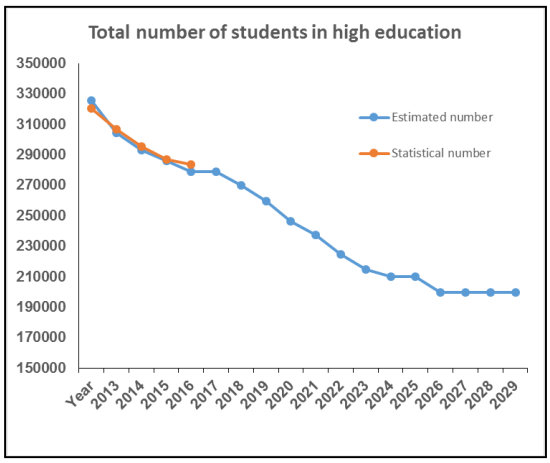


Figure 5 Accuracy of the prediction by five year statistics

IV. Conclusion

In this paper the big data principle was defined and its features were presented. The so called data driven paradigm changes our life, the way of working and thinking. It brings new technological solutions as well as in file storage and handling system, in database management and cloud computing. These solutions often are in their childhood, and we are waiting for new and better solutions to take the IT work easier and the systems safer. It is clear that conceptual thinking is generally behind the technology. It means the missing concepts in database design especially in design of hybrid SQL-NOSQL systems. More interesting are the changes of our thinking, the consequence of the data mining and statistical point of view. An interesting example demonstrated the acceptance of statistical analysis and correlation based prediction.

Acknowledgment

I would like to thank the help and the effort of colleagues and students. József Lakner, Valéria Varga, Péter Ivanics, Balázs Nagy, Csaba Ambruzs, Gaye Ediboglu Bartos and the colleagues of the Limnological Department of University of Pannónia helped our work.

References

[1] S. Muhammad and F. Akhtar, *Big Data Architect ' s Handbook*. Packt Publishing Ltd., 2018.

[2] V. C. Storey and I. Song, "Data & Knowledge Engineering Big data technologies and Management : What conceptual modeling can do," *Data Knowl. Eng.*, vol. 108, no. February, pp. 50–67, 2017.

[3] U. Sivarajah, M. M. Kamal, Z. Irani, and V. Weerakkody, "Critical analysis of Big Data challenges and analytical methods," *J. Bus. Res.*, vol. 70, pp. 263–286, 2017.

[4] J. R. Lourenço, "Choosing the right NoSQL database for the job : a quality attribute evaluation," *J. Big Data*, pp. 1–26, 2015.

[5] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, and L. Sifre, "Article Mastering the game of Go without human knowledge," *Nat. Publ. Gr.*, vol. 550, no. 7676, pp. 354–359, 2017.

[6] M. Majdán and B. Nagy, "Self Organizing Maps in Economical Analysis (Önszervező térkép (SOM) használata gazdasági elemzésekben)[in Hungarian]," Óbuda University, 2013.

[7] C. H. Lee, D. S. Seong, and K. H. Park, "Face recognition using self-organizing map," *J. Korea Inf. Sci. Soc.*, vol. 20, no. 11, pp. 1730–1738, Nov. 1993.

[8] J. L. Alba, a. Pujol, and J. J. Villanueva, "Separating geometry from texture to improve face analysis," in *Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205)*, 2001, vol. 2, pp. 673–676.

[9] K. Ishii, S. Nishida, and T. Ura, "A self-organizing map based navigation system for an underwater robot," in *2004 IEEE International Conference on Robotics and Automation IEEE Vol. 5*, 2004, p. 5726.

[10] A. Ultsch and R. Frank, "Self-organizing feature maps predicting sea levels," vol. 144, pp. 91–125, 2002.

[11] M. Cottrell, P. Gaubert, C. Eloy, D. François, G. Hallaux, J. Lacaille, and M. Verleysen, "Fault prediction in aircraft engines using Self-Organizing Maps," 2009.

[12] OH, "Statistics." [Online]. Available: https://www.oktatas.hu/felsooktatas/kozerdeku_adatok/felsooktatasi_adatok_kozzetetele/felsooktatasi_statisztikak. [Accessed: 15-Sep-2018].

[13] B. Nagy, M. Majdán, V. Varga, É. Hajnal, N. Balázs, M. Márk, V. Valéria, and N. H. Éva, "The Use of Self-organizing Maps (SOM) in Demographic Analysis," in *10th International Symposium on Applied Informatics and Related Areas (AIS 2015): AIS 2015*, Székesfehérvár: Óbudai Egyetem, 2015, pp. 22–26.

[14] E. Hajnal, B. Nagy, and V. Varga, "Intelligent estimation method for demographic effect onto the research and development sector in Hungary," in *SACI 2016 - 11th IEEE International Symposium on Applied Computational Intelligence and Informatics, Proceedings*, 2016.

[15] B. Nagy, M. Majdán, V. Varga, and É. Hajnal, "The Use of Self-organizing Maps (SOM) in Demographic Analysis," in *10th International Symposium on Applied Informatics and Related Areas (AIS 2015): AIS 2015*, O. Tamás, Ed. 2015.

Design of Software Modules for Serving DPWs

Tamas Jancso

Institute of Geoinformatics, Alba Regia Technical Faculty, Óbuda University
Pirosalma u. 1-3, Székesfehérvár, H-8000, Hungary
jancso.tamas@amk.uni-obuda.hu

Abstract— **the paper deals with the capabilities of digital photogrammetry including all the possible evaluation processes and products gained from digital images. Nowadays the UAV technology is more and more popular not only in everyday life but also among the professionals dealing with map production or various data source collection methods for GIS. New software applications can offer full automation of the image processing and mainly focusing on DTM and orthophoto production. The paper tries to overview all the system components serving a digital photogrammetric workstation (DPW).**

Keywords— **digital photogrammetry, orthophoto, DTM, DSM, aerial triangulation**

I. Introduction

Digital photogrammetry means of using digital images on digital photogrammetric workstations (DPWs). The classical photogrammetric evaluation instrument is changed into a computer, which means all the orientation and evaluation process of images is done in mathematical way. This fact offers the possibility of automation of essential tasks of photogrammetry like interior and exterior orientation, digital terrain extraction and orthophoto production. Some tasks remain manual work requiring the engineering planning and supervision. Typically, this task is the compilation of vector-based maps or in more general: vector-based collection of geographic information system (GIS) data. The UAS technology brought a new era simplifying the complicated evaluation processes.

The autocorrelation of common image points is a required module in every application since the number of tie points for aerial triangulation is too much for manual working [3]. On the other hand, the automatic processes hide any steps and the user has no possibility to control all the aspects in an evaluation process. Sometimes this black box effect makes harder to improve the results.

The other important and in the same time interesting topic for discussion is to find the borderline between photogrammetry and GIS. Photogrammetry is defined as a technology for primary data source collection. GIS is serving for analysis of these primary source data and deriving secondary source data using the built models. Some companies don't see sharp border between photogrammetry and GIS, and the software applications merge these two areas resulting very exciting and robust solutions.

In the following chapters we discuss the systems components of a DPW in narrow and in a wider sense.

II. System Components

If we decide to plan the structure of software modules for a DPW, probably we would follow the classical way of photogrammetric software packages. It means we need at least some basic modules like orientation, measurement of DTM points and contours, aerial triangulation, and vector-based mapping. Now, let's follow another way and plan a system by grouping the software modules into categories depending on the functions. On Fig. 1 we see also the hierarchy from top to down as the steps are following each other in the whole process.

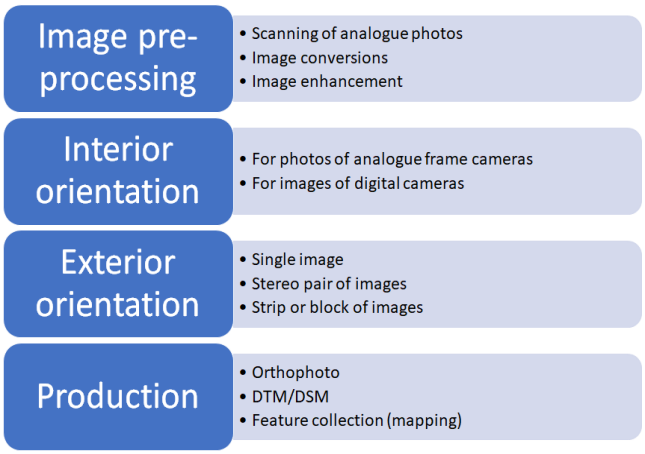


Figure 1. Modules of the evaluation process for a DPW

In the following chapters we discuss these modules in more details.

III. Image Pre-processing

This module is essential since the preparation of images has influence on the quality and accuracy of products at the end of the evaluation process. Here we should consider the following aspects:

- Geometrical resolution
- Image quality

A. Scanning of analogue photos

Before we start to process images, we need to prepare them for evaluation. If we must deal with analogue photos, first we must digitize them using an office scanner or a scanner designed especially for photogrammetric tasks. In both cases the scanning software should offer the scanner calibration as an extra tool to determine the systematic errors. These errors are

then considered when the final image is prepared after scanning the photos.

The scanner calibration is usually done by scanning a well-defined grid of points. The scanned grid points then measured by an automatic process using an area-based image matching technique (Fig. 2). Comparing the error-free grid coordinates with the measured positions we can calculate an error matrix applying a plane transformation procedure. The transformation itself can be an affine or a second polynomial transformation.

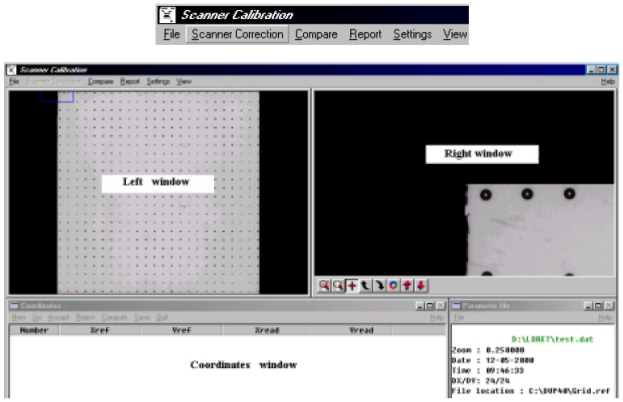


Figure 2. Scanner calibration using DVP workstation

It is important to note that when we decide on the scanning resolution we should consider several aspects like number of channels (grey or colour image), the radiometric resolution (number of bits for storing the grey values of pixels), and the pixel size. For practical reasons the pixel size is usually between 13-21 microns for conventional 23x23 cm aerial images. The experiments showed that pixel size under 13 microns doesn't give remarkable improvement in accuracy, but the image size will grow unnecessarily large as it seen in Table I for grey images.

TABLE I.
SCANNER RESOLUTION AND IMAGE SIZE

Scanning resolution		Image size	Ground resolution for some mapping scales (m)		
DPI	μm	MByte	1:40 000	1:20 000	1:5000
1000	25	81.0	1.0	0.5	0.13
1200	21	116.6	0.8	0.4	0.11
1500	17	182.3	0.7	0.3	0.09
2000	13	324.0	0.5	0.3	0.07
3000	9	729.0	0.4	0.2	0.05

B. *Image conversions*

It is essential to have image conversion module in our system since we need to convert images for different purposes. Usually the following formats are recognized: RAW, BMP, TIFF and JPEG. Here we also build image pyramids for later image matching and for acceleration of viewing images on the PC display. Conversions can be done by lossless compression, and some compressions, like JPEG are lossy.

C. *Image enhancement*

All the image enhancement techniques are necessary to enhance some properties of images. The content information is not increased just some aspects of the image are enhanced for specific goals [1], these goals can be grouped as it seen on Fig. 3. These functions are all important and useful for DPWs at visual interpretation, manual and automatic measurement of image points. Among the geometrical operations the georeferencing operation can be regarded as a separate evaluation method as well, since it produces a georeferenced image which can be used directly for mapping purposes with some limitations. especially it is a useful function for satellite images.

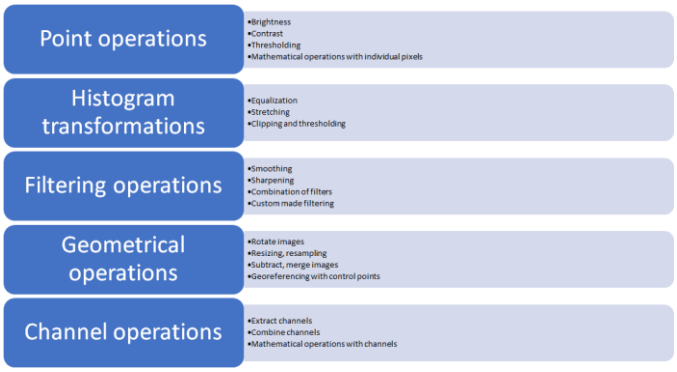


Figure 3. Image enhancement modules

IV. INTERIOR ORIENTATION

A. *Interior orientation of analogue photos*

After scanning of analogue photos, we must carry out the interior orientation which means we need to indicate the camera constant, the principal point coordinates, and we need to setup a photo coordinate system using the fiducial marks. The measurement of fiducial marks can be done manually or automatically using some area-based image matching algorithm.

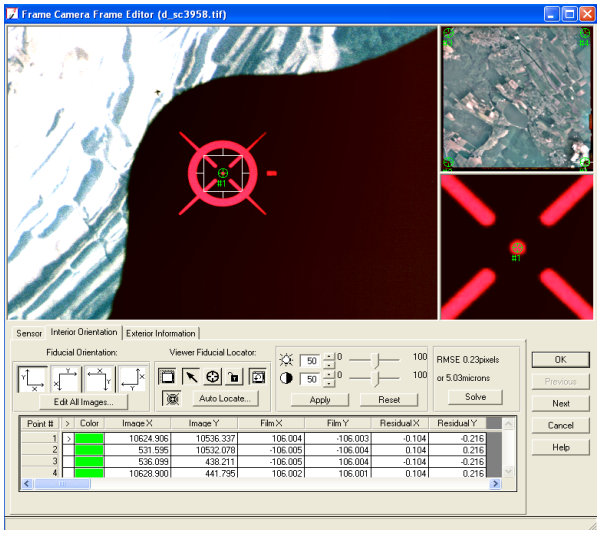


Figure 4. Measurement of fiducial marks in a Leica LPS system

After the measurement of fiducial marks usually a plane affine transformation is carried out, but the software should offer other transformations like plane similarity or second order polynomial transformation. transformation (Fig. 4). The interior orientation can be extended by indicating the lens distortion parameters or values.

B. Interior orientation of digital images

The interior orientation of digital images is a very simple process since here we just indicate the interior elements of the camera and the sensor data. The sensor is characterized by pixels size, width, length of the sensor. The lens distortion parameters are also indicated here. A sample screen is shown on Fig. 5.

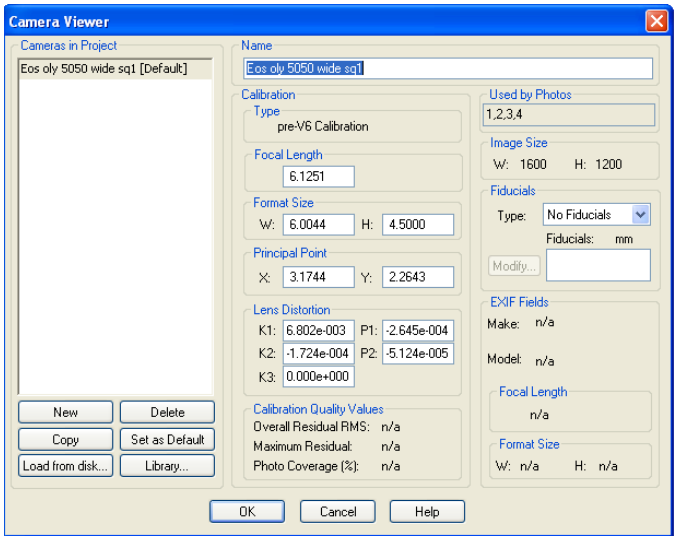


Figure 5. Interior orientation data for digital image sensors

Most of the sensor date can be read from the EXIF data, but these data can be inaccurate, therefore a camera calibration gives better parameters for a digital camera. The camera calibration is done usually during the exterior orientation by so called self-calibration procedure. It means that first we ca assign approximate interior data to the camera and later, during the exterior orientation these data are adjusted.

V. EXTERIOR ORIENTATION

A. Space resection for a single image

If we have only one image, then the exterior orientation becomes a space resection task. For this we need at least four control points to be measured in image. After that the software makes the calculation by least square adjustment solving the collinearity equations. Using only one image we can produce later an ortho-photo, or we can collect features for mapping by digital monoplotting. In both cases we need DTM or DSM for interpolation of heights, otherwise we can measure and calculate planar coordinates only. A sample screen is shown on Fig. 5 from AeroSys software.

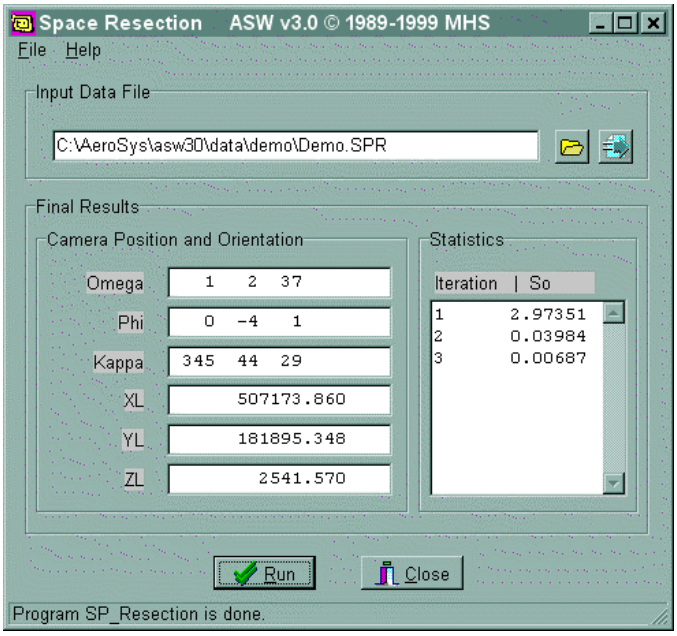


Figure 6. Results of space resection in AeroSys software

It's important to note that the measurement of control points can be automatized if we used well visible signalized terrain points. In this case the software should be capable to recognize these signs automatically. Especially it is a useful task if we need to measure large amount of control points, for example in camera calibration projects. The second remark is that if we have enough control points then the software should offer the calculation of lens distortion parameters as well, since the collinearity equations can be extended by these additional parameters.

B. Exterior orientation for a stereo-pair

The exterior orientation of a stereo-pair of images can be done in two ways. The first option is to measure first at least six tie points and build a model with relative orientation. After this the control points are measured manually or automatically. The automatic measurement of tie points is well known procedure. Here the software can offer not only the are-based matching techniques, but an object based, or a semi-global matching technique is also an option. The control points can be measured only if they are unified and signalized points.

The second option for exterior orientation is to use the bundle adjustment of all measured points regardless of their types, which means, the measurement of tie points and control points are put together to form a system of collinearity equations. In principal the bundle adjustment is more flexible and, in most cases, gives more accurate result thanks to additional parameters considering the camera and lens parameters. The bundle adjustment is also capable to carry out a camera calibration project. On the other hand, the relative orientation is useful for building a model without knowing the

control point positions. The stereo model is then used for stereo measurement of additional points including the control points. If the measurement is completely automatized or we discard the stereo measurement of points, then the bundle adjustment is preferred. A sample screen is shown on Fig. 7.

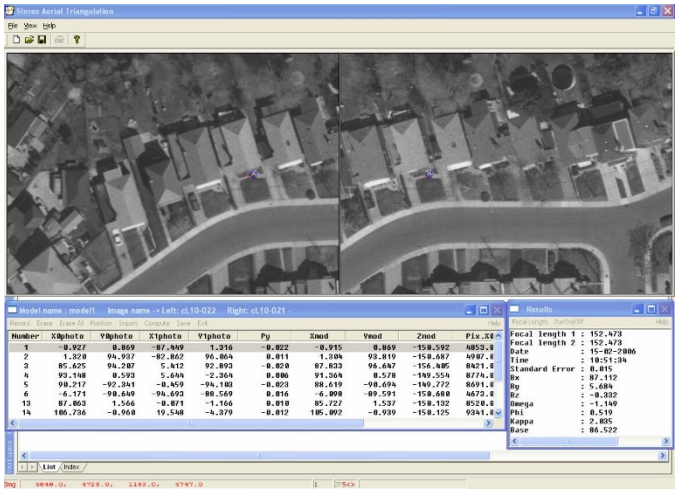


Figure 7. Relative orientation in DVP system

C. *Exterior orientation for a block of images*

The exterior orientation of a block of images is like the orientation of a stereo-pair, but here we must deal with multiple overlaps between the images and strips of images. Typically, this task can be more complicated for UAS images since there the strips of images are not always so regular. The whole procedure is called aerial triangulation. In many cases the measurement of tie points is the first step in order to build models and strips. after this the measurement of control points can be half automatic if we know the exterior orientation elements from GPS/IMU recordings. The half-automation means, the software tries to predict and locate the control points based on the approximate exterior orientation elements and the user should adjust it with manual measurement. Of course, if the points have well known signalized mark on the ground, then the measurement can be fully automatic, but this option is not offered by all systems. On Fig 8 we can see a sample screen from the Trimble UASMaster, which is specially designed for processing of UAS images.

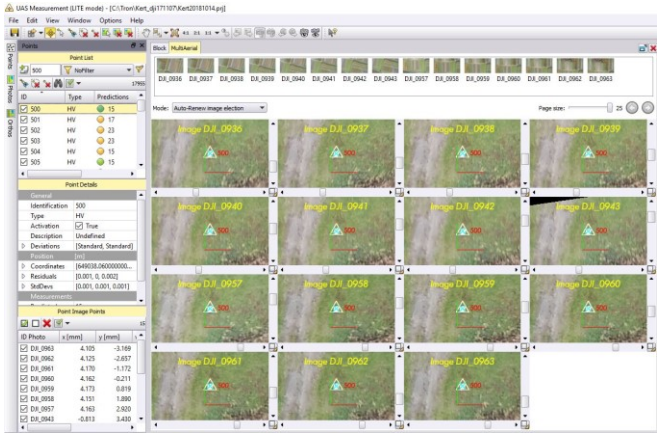


Figure 8. Measurement of control point in Trimble UASMaster

VI. PRODUCTION

A. *DTM/DSM extraction*

The production of DTM or DSM points is essential for producing orthophotos. The extraction of DTM/DSM points should be done by autocorrelating techniques, and the result is usually edited by manual measurement and corrections including the filtering, smoothing and object-based grouping of points. This module should be the most interactive and sophisticated part of the whole software package since here the resulting DTM/DSM effects the accuracy and quality of the orthophoto and other terrain-based derived products. The DTM/DSM file formats should include well-known formats including the LAS format. From the resulting point cloud, the software should be capable to derive secondary products like contour map, slope map, section lines and so on. The DTM/DSM points can serve area and volume calculations as well. On Fig. 9 we see a resulting DTM generated by PHOTOMOD software.

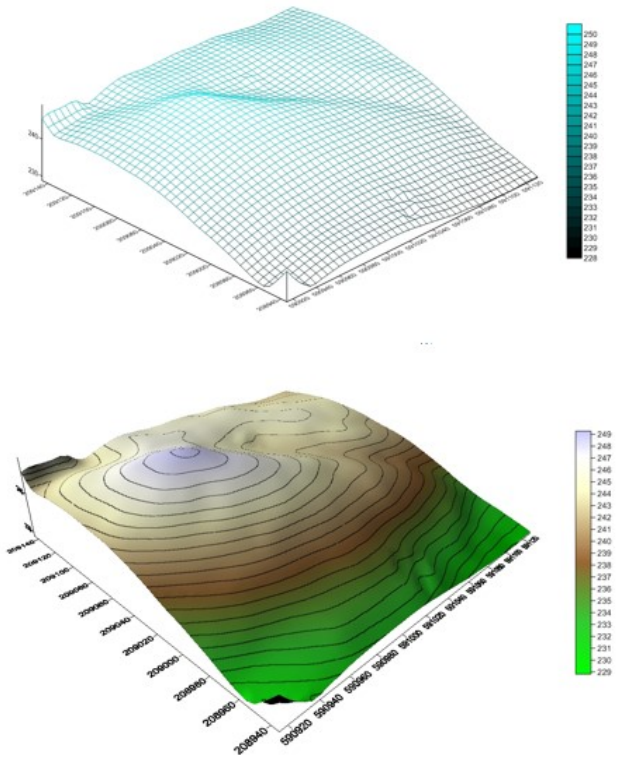


Figure 9. DTM generated by PHOTOMOD software

It's important to note that this process needs a lot of PC memory and multi-core processor. The software should recognize the characteristics of the PC and utilize all the possible sources for calculations including the graphic card processor as well. The multi-threaded and parallel processing can accelerate the calculation dramatically. Today the photogrammetric DTM generation can compete with the LAS technology in many aspects.

B. Orthophoto production

The software should support all the steps of orthophoto production. These steps are

- Orthophoto production separately for every image in the block
- Orthophoto mosaic by seamlines including the photo colour adjust and equalization
- Orthophoto map production

The process can be fully automatic if we don't insist on the manual drawing of seamlines. The well-known resampling methods (nearest neighbour, bilinear interpolation, third order convolution) should be applied for all pixels, although the software should be flexible to setup the resolution, the working area and the projection system of the produced orthophoto. If the DSM is known than we ca produce a true orthophoto. If only the DTM is known than a classical orthophoto is produces. Most simple case when we know only an average height on the terrain. in this case the accuracy of the produced orthophoto can be very low.

The resulting orthophoto can be used direct mapping of planar points (X,Y coordinates), or linking the orthophoto with the applied DTM or DSM, we can measure X,Y,Z coordinates and collect features for mapping in 3D. Although the digital monoplotting is very attractive way for evaluation of spatial points, we should admit that the measurement accuracy is directly depends on the used DTM or DSM. Therefore, in practice we use this evaluation method only in those cases where the high accuracy is not so essential. On Fig. 10 we see an orthophoto mosaic produced by UASMaster:

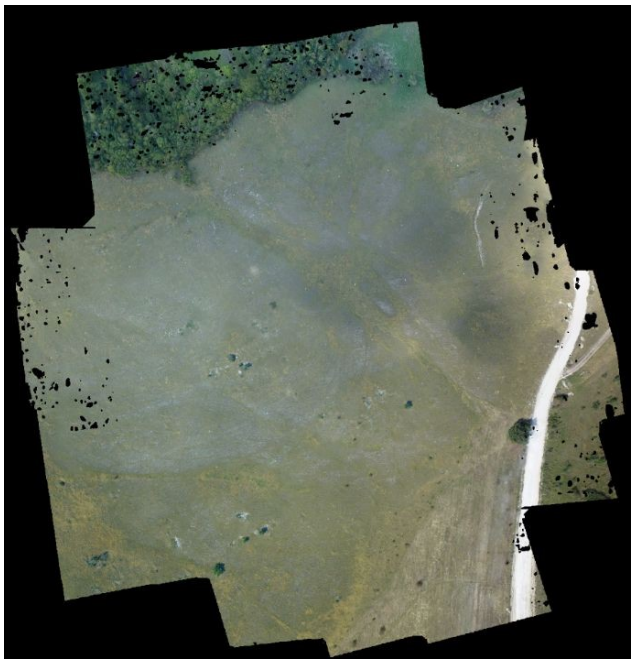


Figure 10. Orthophoto produced in UASMaster

C. Feature collection

For feature collection the DPW should offer at least the following functions in 2D or in 3D supporting the stereo-viewing and evaluation:

- Collection of points.
- Vector-based mapping including the support of features symbology with mapping legends.
- Interface for well-known mapping applications like AutoCAD, ArcGIS, and MicroStation.
- Import/export functions.

The feature collection is a very exciting area regarding the automation especially in city modelling. This a very intensive research area. Therefore, the DPW package should support new challenges like BIM (Building Information management) modelling as well [2]. On Fig. 11 we can see a 3D model produced by automatic photogrammetric feature collection.

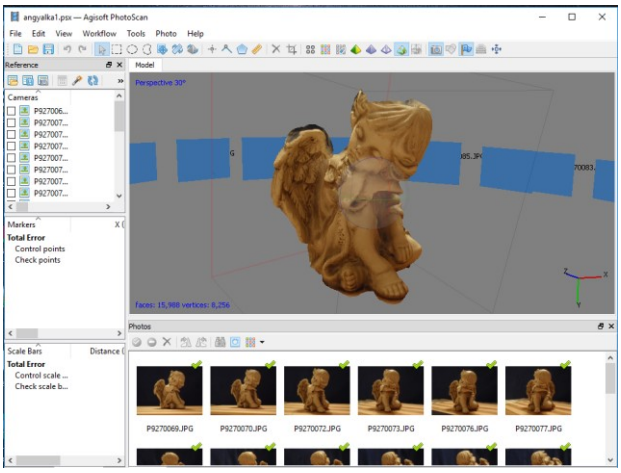


Figure 11. 3D model produced by Agisoft Photoscan

VII. CONCLUSIONS

As a summary we can say that a good DPW software supports the automation in all steps. A wide selection of import/export functions is a very important aspect and it enhances the portability. The UAS software market is developing very fast and hopefully it brings new, more sophisticated solution to the traditional aerial photogrammetry as well.

REFERENCES

[1] Galatsanos N. P, Segall C. A., Katsaggelos A. K., "Encyclopedia of Optical Engineering", *Marcel Dekker, Inc.*, DOI: 10.1081/E-EOE 120009510, pp. 388-402, 2003

[2] Volk R., Stengel J., Schultmann F., "Building Information Modeling (BIM) for existing buildings — Literature review and future needs", https://doi.org/10.1016/j.autcon.2013.10.023, *Elsevir, Automation in Construction, Volume 38*, pp. 109-127, March 2014

[3] Wu Y., "Autocorrelation techniques for soft photogrammetry", *Iowa State University Capstones, Theses and Dissertations*, Digital repository: https://lib.dr.iastate.edu/rtd, pp. 105-170., 1997

Improvement of Network in an Educational Institution According to Demands of Industry 4.0

Bertalan Beszédes[1], György Molnár[1], Attila Sáfár[1]
1 Óbuda University, Alba Regia Technical Faculty, Budai Str. 45, H-8000 Székesfehérvár
{beszedes.bertalan, molnar.gyorgy, safar.attila}@amk.uni-obuda.hu

***Abstract*— Gap between education and demand of industry is broadening as new and new technologies become viral every year. In order to sort out this phenomena educational institutes must be able to prepare its students for real life challenges, by imitating industrial environment as much as it is possible. This article describes the possibility and implementation of a network that fulfils requirements of modern educational and research work.**

I. INTRODUCTION

Neglecting cyber security may put industries, economies, countries and even private life in danger.

In informatics firewalls are such network security systems which monitor data traffic, both ingress and egress, control them according to rules set by network administrator, hence implementing a barrier between the reliable network inside and the not always reliable network outside the firewall.

Next Generation Firewalls (NGFW) (Fig. 1) are part of third generation firewalls, uniting the network filtering functions with traditional firewall roles, like Deep Packet Inspection (DPI), Intrusion Prevention System (IPS), application monitoring, website filtering, antivirus, SSH, TLS/SSL encrypted traffic monitoring, LDAP, RADIUS, QoS etc. Figure 2. demonstrates main features of Next Generation Firewalls. [1], [2]

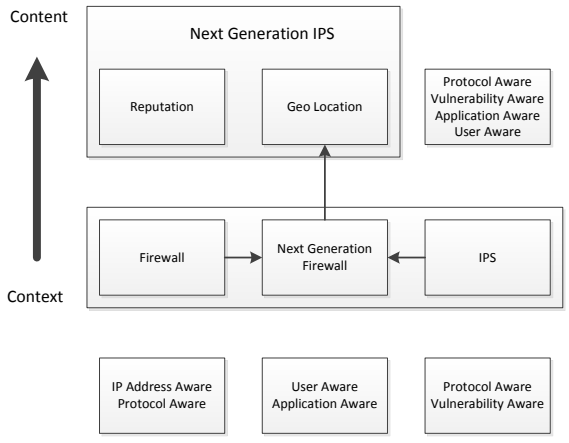


Figure 1. Services of Next Generation Firewalls

II. THEORY

Basic principles of NGFW is to monitor as many OSI model Layers as possible (maybe all 7) in order to extend classic firewall roles (packet filtering, NAT, VPN), implementing a more profound inspection in content of packets and matching of validations.

In order to make communication safer through the internet HTTPS, SSL, and TLS protocols were created. The most important feature of these protocols is that the data is passing through the channel in an encrypted way that is only understandable for the sender and the receiver, with the use of presaged keys. Nowadays more and more network attack is making use of this encrypted "camouflage" as in order to filter these contents deep inspection compatible instruments are needed. NGFWs are capable of monitoring encrypted data traffic, in the way that the gateway decrypts data before forwarding, and sends it to destination only if it is malware free. Administrators can decide if it is needed to inspect all the data traffic or only traffic based on predefined ACLs, or based on black/white list, data coming from or going towards not reliable destinations. By decryption of traffic, we are able to monitor traffic of our network, improve the efficiency of our filters, block unwanted data stream, control access based on predefined rules. [3]

The two main method of revealing network threats are Signature-based detection and Statistical anomaly detection. The basis of signature-based detection is that the system is seeking for typical patterns in code based on a continuously updated list. In case the list doesn't contain the particular pattern, it is added to it for successful preventions in future. The Statistical anomaly detection is looking for network traffic patterns that diverge from usual usage. In case the surveyed sample differs from the calculations of the system, prevention occurs. [4][5]

Centralized authentication, authorization, account management (AAA) is implemented by RADIUS (Remote Authentication Dial-In User Service) network protocol. RADIUS is a client/server protocol, running in application layer. The client is traditionally a NAS (Network Access Server), RADIUS server is usually a UNIX or Windows based daemon process running on the computer. Typically network gateways contain RADIUS client, which helps communication with central server, to identify remote access requesting users. Client forwards request to RADIUS server, and offers service according to the response. This process takes place with the help of UDP (User Datagram Protocol), therefore connection is not guaranteed, in return network traffic is less oppressed. RADIUS server authenticates by PAP, CHAP, EAP methods, and sends approval along with necessary parameters of session, which are valid during the

existence of the connection. There is a chance to collect accounting data using RADIUS services, with what security can be improved, moreover billing can be implemented. Transactions between client and server take place by the help of never sent secret keys, passwords are sent in encrypted format in the network. [6], [7], [8]

LDAP (Lightweight Directory Access Protocol) controls access of directory services. Database, on what it is based, optimized for rapid requests, tree structured (Directory Information Tree, DIT). Suitable for centralized monitoring individual users and groups, directory management. In network environment rapid access of data, management of its organization is inevitable duty, LDAP is perfect service for this role, which helps reaching data quicker, and search is well organized. Main idea while creation was that it is intended for more reading process than writing, therefore synchronization process is allowed. Applications access requested services simply and quickly.

LDAP record is a set of information assigned to the user. Each record comprises there main information: distinguished name (DN), group of attributes, group of object classes. DN explicitly defines the record and its place in the tree structure. Attributes contain the data of records, possess a type, null or more attribute options, set of values, which build up a data. Type defines, how the record must be managed by the LDAP clients and servers, while attribute options hold metadata.

Object classes define group of attribute types. Each record is assigned to group of objects, therefore defines, what kind of object it represents.

LDAP also applies client/server model, where the client accesses the server by TCP/IP protocol (in case of OpenLDAP slapped server and slurped server). Client communicates with slapped, which forwards requests to databases in the background, and response also forwarded to client through slapped. Function of slurped is the multiplication, therefore redundancy, and the sustainability of security. [9], [10], [11], [12]

III. CURRENT INFRASTRUCTURE

Local Area Networks (LAN) can be segmented by VLANs (Virtual Local Area Network), according to its function, project group, application or by any other aspect, independently from where the user or the device is physically in the institute. Devices in the same VLAN behave as they were in a totally separate network, despite the fact, that they might be using the same infrastructure as other devices in other VLANs. Each VLAN make up a logically independent network, domain, and all the packets sent to other VLANs must go through a device that is capable of routing. Different levels of security policy, and different access of users can be set to each VLAN .

Currently the network of the laboratory is a sub-network of VLAN set up for teachers (VLAN10), by the help of a router in-between (Fig. 2.). The PLCs, laptops, PCs are assigned static IP, while router is responsible for the dynamic IP assigned to mobile devices as DHCP server. Devices use router as default gateway, which exclusively communicates the other segments of the above network. For security reasons the connection to wireless network is only possible with usage of password, furthermore it is also the routers responsibility to filter devices connection to network according to their MAC address – based on a pre-defined client list -, broaden segregation in the lab's network.

The size and the manageability of current automation laboratory is a great advantage. Network configuration on the router can be implemented and modified easily in short time, by authorized person with the possession of passwords, and proper background of knowledge. Further filtering might be implemented, such type which might

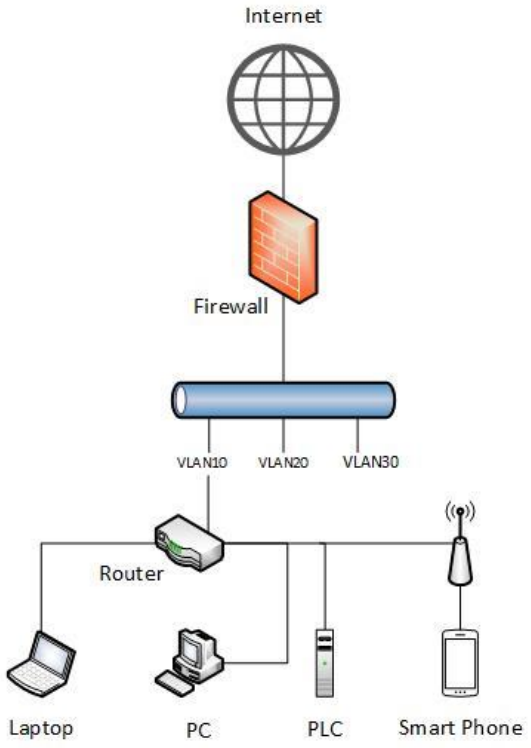


Figure 2. Current Infrastructure

not be used on institutional level, but this level of security still cannot be compared to the security level of a Next Generation Firewall, as the system is maximum as secure as its weakest point. Unfortunately, the previous firewalls are not armed with modern defensive functions, dedicated equipment is necessary in order to make use of these features.

IV. IMPROVED INFRASTRUCTURE

As soon as the improvement of network of the institute takes place, we will have the opportunity to restructure the network of the laboratory. The purchase of a Next Generation Firewall will bring about the main change in our institute. According to decision maker's plan, a Stormshield SN910 device will be purchased in near future. As the whole network of the institute will be reformed, new VLANs will be created, and as a result a dedicated VLAN will be created for the automation laboratory (Fig. 3), which will give the chance to a comprehensive configuration of a separate VLAN. Due to the separate VLAN, other laboratories, physically located at different parts of the faculty, can be connected easily, therefore ease shared projects and research work. Implementation of port filtering and other security

functions (IPS, deepscan) might seem exaggeration in an educational environment, nevertheless it helps creation of a network as similar to an industrial environment as it can be. Thanks to this option, developments from the beginning will take place in the end-usage imitating environment. We will have the opportunity to test the security of the system involving students as well, considering ethical hackers opinions – making use of knowledge students of Óbuda University ethical hacker faculty.

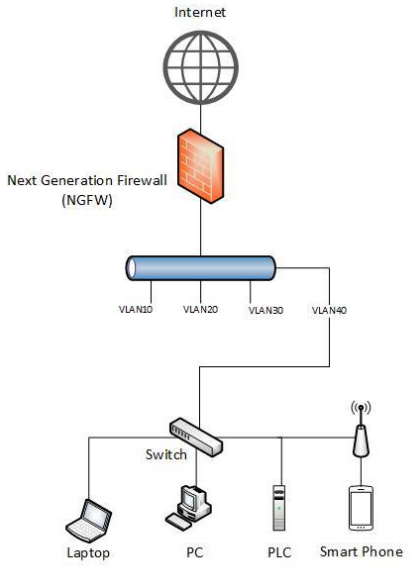


Figure 3. Improved Infrastructure

One of the most important service is the authentication of users and user groups, therefore assigning the appropriate network segment based on authorization. The system applies two methods for this task, which complete each other: LDAP and RADIUS. Implementation of LDAP by Microsoft is well known today as Active Directory. According to the expectations a programme running in a dedicated equipment will serve its purposes of the institute more stably and for longer time period, compared to a virtualized service. Further security is offered by RADIUS, also a service run by firewall. This completes and adds more functions to services of LDAP. LDAP manages user accounts by the help of RADIUS server, while providing encryption itself [13].

As an institute providing studies and research in informatics it is highly important to keep the network secured. Hopefully the technology purchased according to potentials of the university will help us reach the goal. Thanks to IPS/IDS system the NGFW will be able to filter the communication in encrypted channels, reveal anomalies in network traffic, alert network administrators, and even implement automatized steps to eliminate problem if necessary.

V. Practice Application

In order to keep up new trends, we cooperate and discuss with partners from industry, who describe their views on new developments. This attitude lead to the test of SIEMENS Simatic IoT 2020 device. This equipment serves as a gateway between various devices, which not necessarily use the same communication protocol nor data format, but with the help of this device it is easy to use a common platform for communication. The operation system running on Simatic IoT 2020 is the Linux based Node-RED. The device can be easily expanded by Arduino shields and miniPCIe cards. Beside programming in high level programming languages Simatic IoT also offers great a user friendly graphical interface accessible by built in webserver. This interface also offers a great opportunity for programming and controlling through mobile devices, and also ease access of numerous equipment it has connection with. The main reason of application of the device in industry is the reliability, longevity and robust, compared to widely used IoT equipment which are less appropriate (let's say useless) choice in industrial environment [14].

In our laboratory various devices are connected by this method, from different vendors from different years. The main concept is to make use of OPC Unified architecture (OPC UA) which is a machine to machine communication protocol developed particularly for industrial automation [15]. With the Simatic IoT it is even possible to make older series of PLC-s to communicate with the IoT device, therefore it is possible to connect to server and to other clients. This can result financial savings in industry, as by this solution it is not necessary to replace immediately devices from older series, in order to make them compatible with demands of Industry 4.0 and Internet of Things.

Conclusion

In this article, after presenting the current network of our institute, we considered the most up to date solutions of network development in harmony with potentials of the university.

Next Generation Firewalls provide the most effective response to virtual attacks coming from cyber world, and as privacy and data security is the most important issue recently worldwide, we believe that every institute must be prepared for these challenges, keeping themselves and their users secured. This paper gives the opportunity to take a glimpse of modern security systems, and consider further steps according to the needs of particular organizations.

Acknowledgment

The authors wish to thank the support to the Arconic Foundation.

References

[1] Next-Generation Firewall. [Online] November 2017. https://en.wikipedia.org/wiki/Next-Generation_Firewall

[2] Next-generation firewall (NGFW). [Online] November 2017. http://searchsecurity.techtarget.com/definition/next-generation-firewall-NGFW

[3] Managing SSL/TLS T. [Online] 2017. November. https://www.cisco.com/c/en/us/td/docs/security/asacx/9-2/user/guide/b_User_Guide_for_ASA_CX_and_PRSM_9_2/prsm-ug-cx-decryption.pdf

[4] The basics of network intrusion prevention systems [Online] August 2018 https://searchsecurity.techtarget.com/feature/The-basics-of-network-intrusion-prevention-systems

[5] What is an intrusion prevention system? [Online] 2017. November. https://www.paloaltonetworks.com/cyberpedia/what-is-an-intrusion-prevention-system-ips

[6] How Does RADIUS Work? [Online] 2017. November. https://www.cisco.com/c/en/us/support/docs/security-vpn/remote-authentication-dial-user-service-radius/12433-32.html

[7] RADIUS. [Online] 2017. November. https://en.wikipedia.org/wiki/RADIUS

[8] RADIUS (Remote Authentication Dial-In User Service). [Online] 2017. November. http://searchsecurity.techtarget.com/definition/RADIUS

[9] Introduction To LDAP. [Online] November 2017. http://quark.humbug.org.au/publications/ldap/introldap.html

[10] LDAP-ról pár szóban. [Online] 2017. November. http://padre.web.elte.hu/ldap.html

[11] Lightweight Directory Access Protocol. [Online] 2017. November. https://hu.wikipedia.org/wiki/Lightweight_Directory_Access_Protocol

[12] Basic LDAP Concepts. [Online] November 2017. https://www.ldap.com/basic-ldap-concepts

[13] Firewalls play by new rules. [Online] November 2017. http://searchsecurity.techtarget.com/feature/Firewalls-play-by-new-rules

[14] Széll Károly, Manhertz Gábor, Development of a web-based training module in robotics, 2016 International Symposium on Small-scale Intelligent Manufacturing Systems (SIMS), Narvik, Norvégia, 2016.06.21-2016.06.24., IEEE, 2016. pp. 1-6.

[15] Széll Károly, Czmerk András, Korondi Péter, Friction with Hysteresis Loop Modeled by Tensor Product, AUTOMATIKA 55:(4) pp. 463-473. (2014)

Analysis of Crowdsourcing based data acquisition

Zs. Rassányi, Z. Tóth*
* University of Óbuda, Alba Regia Technical Faculty, Institute of Geoinformatics, Székesfehérvár, Hungary
toth.zoltan@amk.uni-obuda.hu

Abstract — The crowdsourcing based data collection is a relatively new and interesting part of the geoinformatics. In our article, we would like to present the application possibilities of a large number of images on different image and video sharing portals for an architectural survey. The article shows the possibilities of automation, the advantages and disadvantages of the technology, as well as the other potential uses of the crowdsourcing based data acquisition.

I. CROWDSOURCING

Since the end of the last decade crowdsourcing had become a fast spreading data collecting procedure due to the increasing of (mainly laic) users. We can find similar methods mainly in the field of GIS (for example community map editing), but due to the rapid spread of point cloud formation algorithms based on image matching, this field had been supplemented with an interesting application group build on large amount of pictures in various quality and resolution, originate from different photo and video sharing websites. According to the PhD dissertation of Somogyi Árpád [1] the first spatial reconstruction, based on pictures from social websites, was made of Rome with the use of approximately 150.000 pictures [2]. As a Hungarian related article, we can mention one from Dr. Gede Mátyás [3], who has made researches in connection with shared photos with geographical data. Based on the writings we can say this field counts as a newness, that is why the studying of it had become topical and reasonable. In our article we would like to present the steps and devices of the assessment, on the example of a statue in Székesfehérvár.

II. THE STUDY'S OBJECT AND TOOLS

As the concrete object of the study we chose Ohmann Béla's creation, named Fehérvári jog, in its more popular name the Országalma. One reason of our choice is symbolical. The statue, made in 1943, had been inspirited by the regalias and has become not only the city's but our faculty's emblem too. As it is the most well known monument in Székesfehérvár we expected sufficient amount of available photos therefore we did not check the number and quality of them in advance. In the interest of studying the geometry of the point cloud made by the help of crowdsourcing we needed a reference, what we get from the laser scanning measurement of the Országalma (Figure 1.).

Figure 1. Reference point cloud from the laser scanning measurement

Other than the laser scanning measurement we were taking own pictures too with simple amateur ground cameras within ideal measuring arrangement (Figure 2.). From these photos we created the dense point cloud of the statue with the help of Agisoft Photoscan software (Figure 3.). We also analyzed what was the maximal resolution and accuracy that we can reach with this measurement.

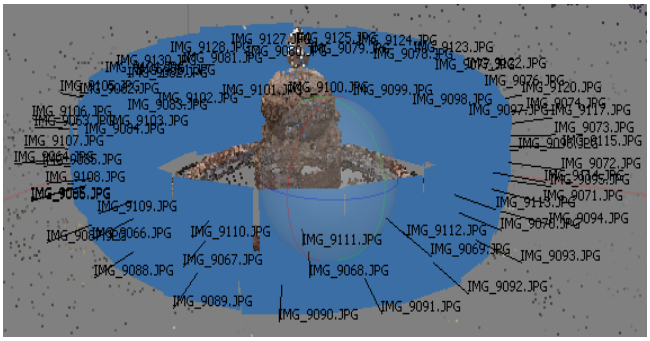


Figure *2*. Measurement layout.

Figure 3. Point cloud - made from our own photos.

During the investigation of crowdsourcing we used several sources of data. From the popular social websites of Flickr, YouTube, Vimeo, and Instagram we were exact photos and video recordings as well. Naturally crowdsourcing can be adapted the most efficiently if the filtering and downloading of pictures is automated. An example innovation with the same purpose is the API *(Application Programming Interface)* of Flickr. With calling the application's functions into our application or simply by using the web interface designed for this purpose, it is possible to filter photos by place capturing time, uploading time, attached meta data or keywords and then download them. In the course of the searching method we used the filters listed above. From our aspect it was significantly important to limit the capturing time of photos, because the statue had been resurrected and rebuilt in the past years, this way we could utilize only those photos which belong to the current position.

We obtained nearly 340 useful pictures from some of the sources. The obstacles that we have faced during the filtering, were mainly caused by the incorrect geo referencing or because the keywords were given wrong. We have done the filtration manually but further on we design to automate this process through some developments. The evaluation was also done with the above-mentioned Agisoft Photoscan commercial product. We present the result point cloud on the 4th figure. As we can see, as for the resolution, it has not as fine quality as the reference point cloud.

Figure 4. Point cloud produced from crowdsourced images

Figure 5. Manually selected ground control points

We made point clouds not only from automatic evaluation, but with the manual giving of tie points as well (Figure 5.). The comparison of the point cloud achieved with crowdsourcing and the reference one was done by the analogy with a former casework [4]. All of the point clouds were transformed, with an ICP algorithm based transformation (Figure 7-8.), into a common reference system, then we examined their average distance and deviation correlate with the reference point cloud. The study had been carried out separately on recordings with different sources, the outcome is exposed in the 1st table.

It is noticeable, that naturally, the best resolution belongs to the point cloud produced from our own pictures what ensures the highest overlap. The next one is the point cloud made from the numerous pictures extracted from videos. The model, calculated by the help of ~340 crowdsourced images, contains much less points. One of the reasons of it is although the number of pictures is relatively large, the spatial distribution of them is not nearly even.

Source of the reference point cloud	Source of the adjusted point cloud	Average distance [m]	Deviation [m]	Number of points in the point cloud (pcs)
Laser scanner	336 assorted pictures, without tie points	0,043	0,063	3740
Laser scanner	336 assorted pictures, with tie points placed on buildings	0,001	0,001	2486
Laser scanner	336 assorted pictures, with tie points placed on the statue	0,028	0,033	12059
Laser scanner	Pictures extracted from videos	0,023	0,024	61011
Laser scanner	Own pictures	0,011	0,010	204683

Table 1. Average distances from the reference point cloud in the cases of recordings with different sources

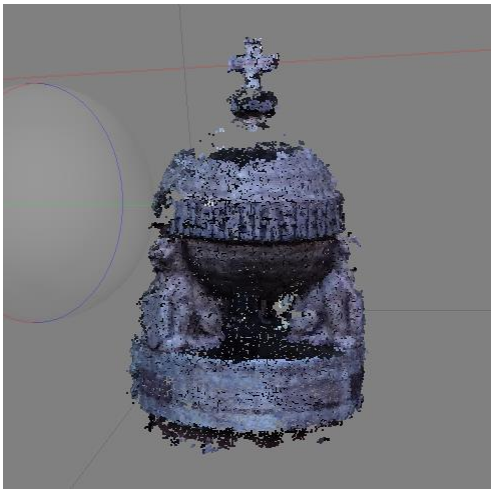


Figure 6. Point cloud from videos

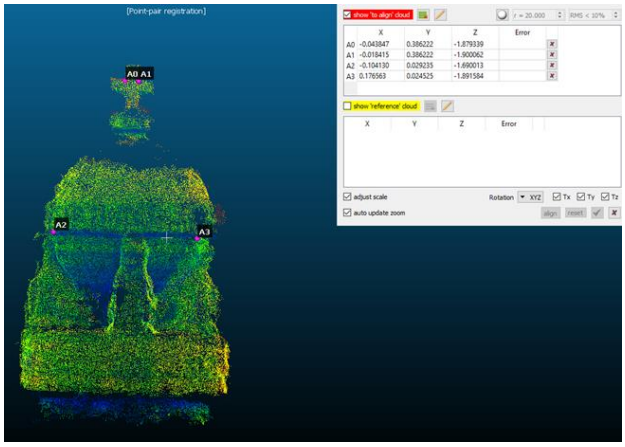


Figure 7. Point cloud (from Photos)

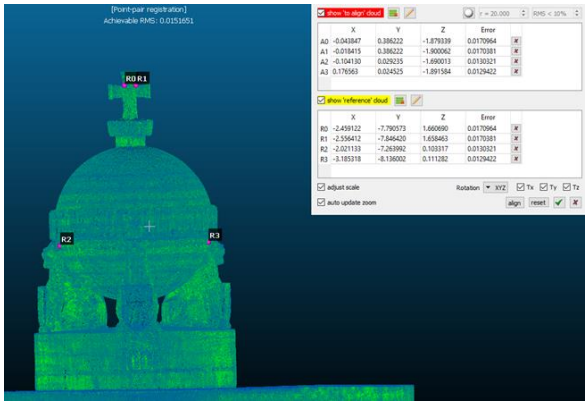


Figure 8. Point cloud (from scanner) and the result of ICP algorithm

III. CONCLUSIONS

All in all, on the score of the paper we can establish that the crowdsource based point cloud processing holds promising developmental possibilities. One of its main advantages lies in the opportunity of automation. Especially in those locations that are beloved by tourists we can expect the sufficient quantity (and quality) of imagery that lead to proper coverage and geometry. The final product can be suitable for architectural or cultural heritage conservational purposed assessment too [5, 6]. In further developments we set our sights on the automation of filtering the pictures originate from crowdsourcing.

REFERENCES

[1.] Somogyi József Árpád (2017): Pontfelhő alapú mérnöki célú eljárások optimalizálási eljárásai PhD disszertáció BME, Budapest.

[2.] Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M. Seitz, Richard Szeliski (2009): Building Rome in a Day. IEEE International Conference on CV.

[3.] Gede M. (2013): Fényképek térképei – geotaggelt fotók térbeli eloszlásának térképes vizsgálata. Geodézia és Kartográfia 65. pp. 17-19.

[4.] Gáti B, Tóth Z (2017): A tarnaszentmáriai templom felmérése XVIII. Földmérő találkozó kiadványkötete p 4.

[5.] Pokrovenszki K., Vágvölgyi B., Tóth Z. (2016): Practical Experience with the 3D Photogrammetric Methods used at the Excavation of Csókakő Castle MAGYAR RÉGÉSZET pp. 20-27.

[6.] Kiss A., Balázsik V., Tóth Z. (2018): Modern data acquisition tools in Archaeology, (THE NATIONAL TECHNICAL-SCIENTIFIC CONFERENCE kiadványkötete XVII.)

Remote Sensing Methods for Open Pit Mine Surveying

Z. Tóth*
* University of Óbuda, Alba Regia Technical Faculty, Institute of Geoinformatics, Székesfehérvár, Hungary
toth.zoltan@amk.uni-obuda.hu

Abstract — In the last few years, new data acquisition technologies and related processing techniques - typically based on laser scanning or image matching- have been dominant in surveying and mine surveying. This new technology often comes into the users with only a "difficult to parametrizable" commercial product; their scientific research - from an accuracy, or economical point of view -but at the current level of development even in the search for uses - is considered.

Photogrammetry, Remote Sensing, Laser scanning

I. NEW REMEOTE SENSING METHODS IN MINE SURVEYING

Technical solutions used in the mine surveying practice in the last ten years, experienced significant changes. So far almost exclusively used the traditional row of producing discrete polar survey by these new procedures, common to the traditional territory of the professional terminology lists these remote sensing. These new procedures as well as the practice of more important, generally provide a point cloud.

Several of these techniques in "chronological" order:

- Terrestrial Laser scanning
- LIDAR
- Mobil Laser Scanning
- Digital photogrammetry (Close Range and Aerial)

The practical applicability of *Terrestrial Laser scanning* was examined in our previous article [1]. In practice, the scanners with a smaller range of reliability, but longer range, based on ToF are used. This technology has a clear economic advantage over the aerial survey, but at the operating mines (for example stocks) it is unclear even with traditional surveys.

Applying for the use of M*obile Laser Scanners* as a domestic example, we suggest Csörgits [2] presentation. Its advantage is unquestionable compared to static measurement, but the continuous provision of external orientation -in addition to the obvious economic disadvantages- may be a technical problem that significantly affects its practical applicability.

Photogrammetry -if not for digital- is of course not a novel technical solution in mine surveying. There is an example of using this close-range photogrammetry in open pit mining survey from 1979 in [3]. The practical spread of the conceptual solution was a major obstacle to the relatively high added value work, which was strongly influenced by the global matching methods.

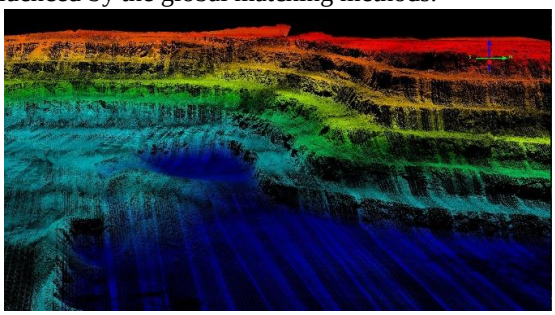


Figure 1. Lidar based Point cloud [4]

LiDAR. The appearance and spread of airborne laser scanners can be dated back to the 1990s, and their distribution is mainly related to the development of mobile telecommunications. There was a need to produce a fast surface model of large areas for the visibility test for the installation of mobile towers. For mining applications, see also international literature. Domestic relations are unlikely to allow a significant spread of the LIDAR surveys. This is primarily due to economics. On the other hand, the number of flying days is considerably limited mainly due to practical needs, weather and airspace usage. Practical approaches have not been helped by recently-based *photogrammetrical* methods.

II. LiDAR IN MINE SURVEYING

There are some practical advantages of using LiDAR in mine surveying in this cases:

- large-scale, fully homogeneous surfaces where purely photogrammetric procedures based on image-matching do not provide sufficient interface points (mining ponds)
- not entirely homogeneous, but possibly identical, repetitive texture areas are typically multi-layered faces, most of the wood-plated surface, which makes the aerial triangulation difficult
- By recording multiple reflections, it is possible to measure the vegetation-covered surface.
- This is particularly useful for abandoned mines and recultivated areas, as well as for the research surveys described in our example, where we need the DTM instead of the surface model.

An examination of the practical applicability of a pointcloud -produced by aerial laser scanning- is presented by using surface (terrain) model for mining research. For economic reasons that the technology is directly exposed under the current technical-economical conditions. But it is also a by-product of other Lidar surveys to have a practical need to deal with the subject. (e.g. abandoned mines and research surveys.)

III. LiDAR IN MINING RESEARCH

Our research is presented through a mining research example. For this purpose, almost exclusively state analogue (1:10000) topographic maps have reported geometric background content. The essence of our studies is to compare the DTM based on the topography map with the DTM based on the LiDAR survey.

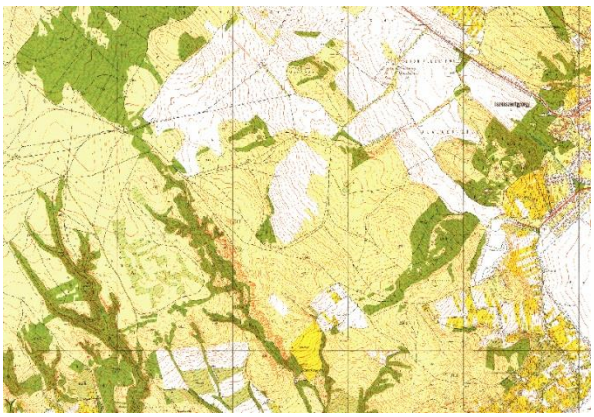


Figure 2. Test area on topographical map

For the purpose of the LiDAR survey, the DTM of the test area was produced using only the state topographic map, which was considered the base of the study. (Figure 3.)

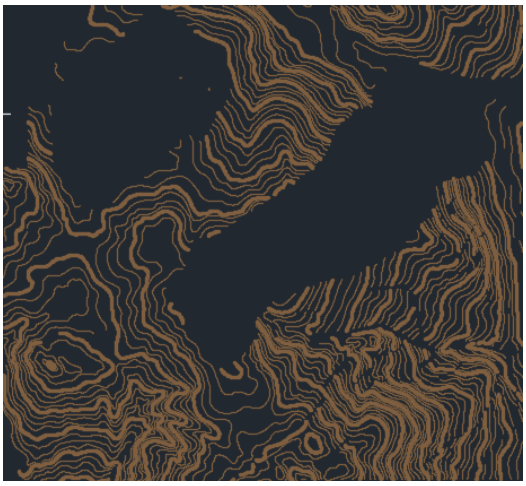


Figure 3. DTM of test area based on topography map

From our test area, our institute had a LiDAR survey from 2009. Its resolution is 1 meters*1 meters (Figure 4.) for ease of use, the test area was reduced to 500*500meters. The area has mixed vegetation, it is covered by grass, and forests. From this point of view, it is an ideal test area, since a hard-to-detachable plant cover will cause a regular error in producing DTM. Using classification algorithms, we have the possibility to separate the different plant cover, this is necessary if the multiple reflection is not counted.

Some of our investigations were carried out with commercial software (Civil 3D). For this purpose, we prepared the space indexed point cloud from LiDAR survey (Figure 5) and we built a DTM from it in CAD program.

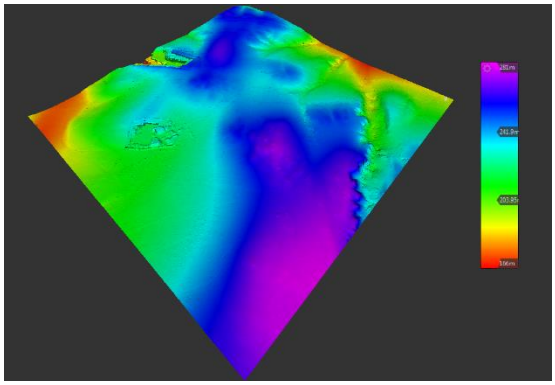


Figure 4. Indexed point cloud

Figure 5. Text area in commercial CAD program (LiDAR-based point cloud)

Generally, three types of interpolation can be taken into consideration when building a TIN model from a point cloud in general CAD environment:

- No interpolation,
- Plane average
- Kriging interpolation

All three methods are used to construct TIN model. In order to compare the three methods we calculated the volume differences between the three TIN models.

Base surfaces	Volume differences (m3) on 250.000m2 test area		
	No interpolation	Plane average	Kriging interpolation
No interpolation	-	-27900	-18300
Plane average	+27900	-	+9700
Kriging interpolation	+18300	-9700	-

Table 1. Volume differences between interpolation methods

We also calculated other statistical measures. The biggest difference between the two surfaces (e.g. No interpolation and Plane average interpolation) is -6 meters and 0.7 meters. What seems to be excessive, however, there is an active mine and significant natural valleys in the area with high altitudes, and these differences occurred in these areas. The average deviation in the area was -11cm. Our calculations are summarized in the following two tables (Table 2, Table 3.).

Base surfaces	Min/max distances (m)		
	No interpolation	Plane average	Kriging interpolation
No interpolation	-	-5.8/+0.7	-4.9/+0.9
Plane average	-0.7/5.8	-	-1.2/3.3
Kriging interpolation	-0.9/+4.9	-3.3/1.2	-

Table 2. Min/max distances between interpolation methods

Base surfaces	Average distances (m)		
	No interpolation	Plane average	Kriging interpolation
No interpolation	-	-0.11	-0.07
Plane average	+0.11	-	0.04
Kriging interpolation	+0.07	-0.04	-

Table 3. Average distances between interpolation methods

The Table 3. shows what kind of regular error the TIN model generates in the interpolation solution. For a geological calculation, these are negligible, but are not acceptable for many civil engineering applications.

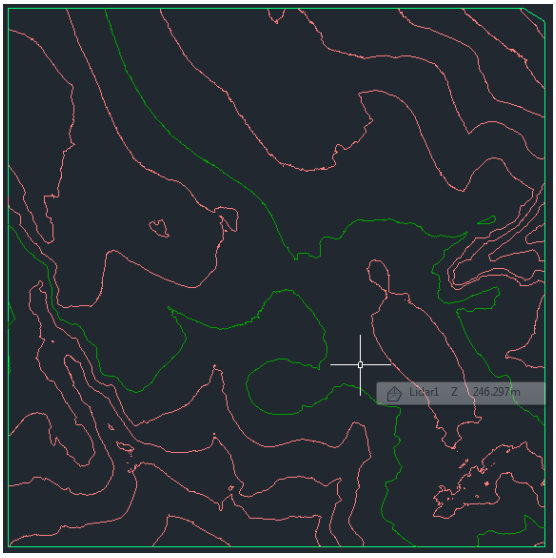


Figure 6. TIN model in commercial CAD program (LiDAR-based point cloud)

In the next step we prepared the DDM of the area based on the topographic map. We cut out those parts that have changed since the topographic map and the aerial scan (Figure 3.). These are mainly areas affected by mining. Because of the large surface and the 100 meter deep side, they would cause significant errors in the calculation.

Base surfaces	Volume differences (m3) on 150000m2 test area			
	Topographic map	No interpolation	Plane average	Kriging interpolation
Topographic map	-	+136000	+108000	+120000
No interpolation	-136000	-	-26000	-16000
Plane average	-108000	26000	-	+9000
Kriging interpolation	-12000	16000	-9000	-

Table 4. Volume differences between Topographic map and different interpolation methods

The result tables are, of course, symmetrical to the diagonal.

Base surfaces	Min/max distances (m)			
	Topographic map	No interpolation	Plane average	Kriging interpolation
Topographic map	-	-5.3/5.0	-6.8/+6.4	-6.9/+5.0

Table 5. Min/max distances between Topographic map and different interpolation methods

Base surfaces	Average distances (m)			
	Topographic map	No interpolation	Plane average	Kriging interpolation
Topographic map	-	+0.9	+0.74	+0.80

Table 6. Average distances between Topographic map and different interpolation methods

The values of the min/max distances are in dotted areas and in wooded areas. The "average distances" includes the editing errors of DDM created from the topographic map of course. The impact of vegetation on the survey is clear. The positive values of the averages distance table mean the surfaces of the Lidar based DDMs are systematic higher than the topographic based DDM. The results can be used for mining research.

IV. CONCLUSIONS

Overall, it can be stated that LiDAR surveys are perfectly suited to mining purposes. For reasons of economics, their domination is not expected, but the rapid development of UAV based LiDAR systems (cheaper, lighter sensors) is likely to contribute to the spread of technology. In addition to current economic conditions, it is possible to use it as a by-product of other LiDAR surveys in mining. This is possible in the survey of abandoned mines and research areas. In these cases, there is a great advantage in capturing multiple reflectors against image-based procedures.

REFERENCES

[1.] Ládai András, Tóth Zoltán (2005): Egy bányatérkép újjászületéséről, *BÁNYÁSZATI ÉS KOHÁSZATI LAPOK-BÁNYÁSZAT* 138:(5) pp. 13-15.

[2.] Csörgits Péter (2018) Mobil lézerszkenner a bányamérési gyakorlatban. LVII. Bányamérő Továbbképző és Tapasztalatcsere, Tatabánya.

[3.] Hoványi Lehel (1979): Külfejtéses Bányák felmérése. Műszaki könyvkiadó. ISBN: 963-10-2924-7

[4.] Jason R. Kack (2017): UAS LiDAR in support of mining https://www.linkedin.com/pulse/uas-lidar-support-mining-jason-r-kack?articleId=6305869371092934656

[5.] Pikli Károly, Tóth Zoltán (2012): A térfogat meghatározások pontossági kérdései LI. Bányamérő Továbbképző és Tapasztalatcsere: Bányamérés Selmectől napjainkig - 50 éves a fehérvári földmérő képzés. Konferencia helye, ideje: Székesfehérvár, Magyarország, 2012.06.13-2012.06.15. Székesfehérvár

From real – time manufacturing to IoT digital technologies

Elisabeta Mihaela Ciortea
"1 Decembrie 1918" University of Alba Iulia, Faculty of Engineering Sciences
Str. Unirii nr. 15-17, Alba Iulia, Romania
ciortea31mihaela@yahoo.com

***Abstract*— The paper deals with several aspects of Petri networks, their implications in the modeling, analysis, synthesis and implementation of systems in the field of manufacturing. Since these aspects are being analyzed based on discrete events systems and because everything is done in real time, it is recommended that digital technology is accessed using IoT devices. For the chosen manufacturing system, we will only model and analyze the sectors in which the importance of the component systems can be highlighted in real time. Even though Petri networks are no longer secret in use, they are easy to use in simulation and modeling, and the resulting diagrams are easy to interpret. The synthesis and implementation proposed in this paper are considered cloud-based because the security of information is much more complex. The tracking and control system is proposed to be managed using IoT systems and all levels to provide the most accurate security of the manufacturing process. This system can be used in enterprises that adopt the implementation of IoT systems, implement digital technology. The system can only be applied to manufacturing sectors, and should not cover the whole system from supply to disposal.**

I. INTRODUCTION

The specialized literature deals very carefully with details of Petri networks and their implications in the modeling, analysis and implementation of manufacturing systems. [4]

As digital technology follows the path of technological development of the future and more IoT is an integral part of it, I will try to consider examples for which conceptual complexity is extracted from specific systems.

In the paper are used concept and techniques dedicated to classical systems, but they are private from the perspective of digital technology. [9]

The starting point is a manufacturing system for which the flow of transport is analyzed. Only one sector in the whole system is modeled and the difficulties encountered in the monitoring and control system imposed by IoT are highlighted. This analysis is necessary because very crowded systems use systems with discrete events that most often lead to hybrid and continuous approximations. [9]

Because production systems are large and conceptual complexity is due to particular situations, general analyzes are not only considered complete examples but also partial sectors.

Cloud manufacturing is a modern concept that uses computing and IT, representing a distributed set of computing, applications, access to information and data storage without the need for the user to know the location and physical configuration of the systems that provide these services. [3]

II. PRESENTATION AND ANALYSIS OF THE MODEL

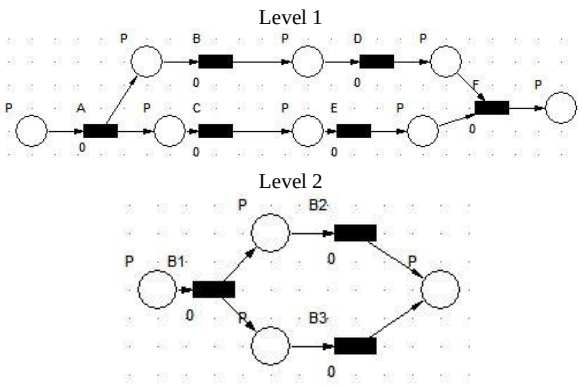
The analyzed model contains two processing centers and an assembly center. In this system are added some additional elements to ensure good functioning. These additional elements have the role of ensuring a continuous flow and avoiding any errors. These errors can be so caused by various interruptions (maintenance) but also accidental interruptions.

In order to model these types of systems, an important role is played by level analysis, [7]

To be able to express security requirements relating to physical devices, the model must be designed to contain a description of the working environment to adapt their behaviour to situations that may occur (not provided).

Planning complex processes face the problem of describing the workload characteristics and their interactions in a concise and easy to understand. One of the basic strategies for achieving this function is a hierarchical approach.

Within the complex processes by analysing the characteristics of the system according to the blank transported considering the general case. A particular case where it can be studied in future work in the system is to be used for the transport of hot blanks, that is, in the steel rolling system. [4]



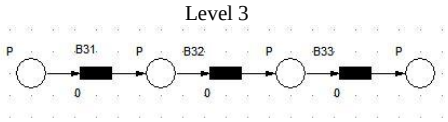


Figure 1. The hierarchical levels with Petri Nets [8]

In level 1 Petri Net is presented to the highest level network, a hierarchical transition and is used for level 2 Petri network. At level 2 transition Petri Net is a lower level hierarchical Petri Net. This mechanism can be used to achieve hierarchical decomposition of a complex system.

Under this system could be divided into different levels of detail from top to bottom. The concept of hierarchical structure information can be used effectively. The upper level offers a higher level of abstraction and detail a complete plan without extensive and complex hierarchical system. At the lowest level it offers a high level of detail on Petri Nets modelling form and is working at the complex tasks of system. [8]

Figure 2 presents the classical structure modelled with Petri nets, from which depart in order to analyse hierarchical system. The classic is the representation of Petri Nets model with factors indicated in the first chapter.[7]

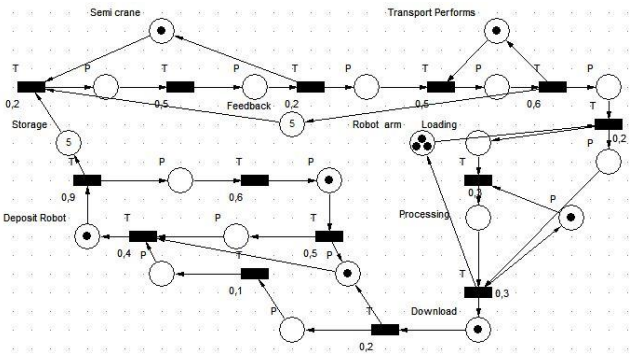


Figure 2. Classical structure modelled with Petri nets[7]

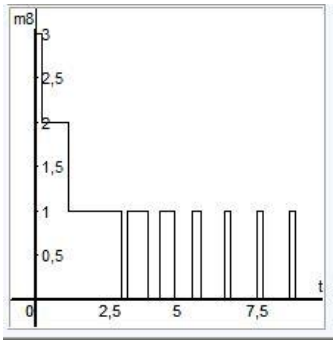


Figure 3. Increased flow after initial processing [7]

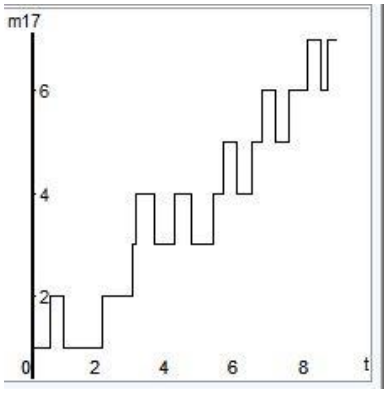


Figure 4. Transport blank in the classical system [7]

In the initial state, the processing centers and the robot are inactive and the warehouses are empty.

The following types of handling systems are used in manufacturing systems:

-Transfer linear for automatic lines,

-Conveyors,

-Transfer palletizing on processing centers,

- Industrial robots for handling materials.

The factors that infuse material handling at all levels, especially at the level of abstraction, are:

-The amount of material that may be small or large, continuous or discontinuous,

-The rhythm imposed by the material flow is determined by time, an important factor in the establishment of architecture,

-The Schedule of Material Flows routes, aims at the distribution, management and dispatching of the transported materials,

-Transport itineraries, defining each transport route, manipulation including distance and time.

For smart systems, common problems that may occur are failure, machine repair, variation in processing times.

The advantages of the Petri networks used in the modeling and analysis of the manufacturing systems are:

- Explicit relations between events.

-The modeling language can serve to describe the system on different levels of abstraction.

- System properties analysis to validate the solution.

For the chosen manufacturing system, I will design and evaluate on the basis of factory data collected but reduced so that a complete simulation can be made. Because it is a theoretical research, it is attempted to model stochastic petrochemical networks, which with the help of the chosen technologies can be achieved with IoT devices.

The following types of handling systems are used in manufacturing systems:

- Linear transfer for automatic lines,

- Conveyors,

- Pelleted transfer to processing centers,

- Industrial robots for handling materials.

The factors that infuse material handling at all levels, especially at the level of abstraction, are:

- The amount of material that may be small or large, continuous or discontinuous,

- The rhythm imposed by the material flow is determined by time, an important factor in the establishment of architecture,

- Schedule of material flows routes, aims at the distribution, management and dispatching of the transported materials,

- Transport itineraries, defining each transport route, manipulation including distance and time.

The advantages of the Petri networks used in the modeling and analysis of the manufacturing systems are:

- explicit relationships between events.

- The same modeling language can serve to describe the system on different levels of abstraction.

- system properties analysis, validating the solution.

Cloud monitoring is essential for automated and autonomous data control systems, but also for service logic feedback. Scalability and security are critical to cloud monitoring. In general, monitoring is done locally, and shared information is monitored and not available outside of resources. [2]

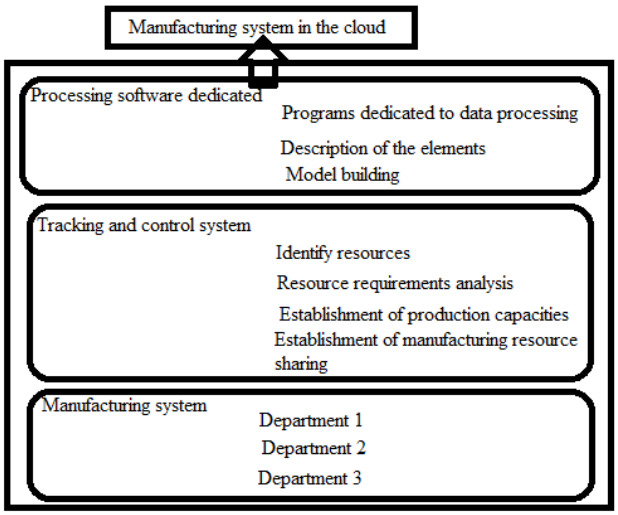


Figure 5. Cloud manufacturing systems [3]

This paper analyzes the path on the three components, but bases from Component 1 on Building the Model as close as possible to reality so that there are as few errors as possible in the simulation, Component 2 analyzes the Determination and Analysis of Resources, but also Establishment of Manufacturing Capacities, where required by the introduction of temporary warehouses or storage facilities. For Component 3 only one department is selected that deals with actual production.

Cloud manufacturing service providers are motivated to manage their activity because the ability to obtain fast and accurate information is grouped together in an advantageous way.

Cloud manufacturing service users are requesting information at all stages of the system's lifecycle.

In this paper, a cloud manufacturing system is defined as a global activity of an enterprise that has activities for both manufacturing resources, which can be resources one or more manufacturing resources. [3]

The cloud manufacturing service is dedicated to implementing semantic interoperability between manufacturing enterprises.

The major advantage of this type of manufacturing is that depending on the flexibility of the system, the description of the proposed model of each manufacturing resource can be customized.[3]

III. MODEL FOR IOT

A model for preparation for the adoption of production is presented below:

- Size and investment capacity of the manufacturing and supply chain. Micro-enterprise ecosystems are the fastest and most destructive innovators, due to insufficient investment. Large industries are considered champions of ICT-based innovation, but their migration processes are often slow and bureaucratic. Feasibility and economic sustainability is the major maturity criterion addressed in this dimension;

- Industry dimension and ICT awareness, in which high technology industries already have knowledge of certain technologies and well-trained young people in digital competences. Low-tech industries are heavily dependent on the knowledge and experience of older workers and engineers, and migration in many cases implies the meaning of knowledge transfer. Social sustainability is the main criterion approached in this dimension;

- Another dimension considers the political and societal environment in which the supply chain operates.

Some changes to IoT include:

- communication and interaction solutions,

- capitalizing data on system evolution,

- the use of specialized services for the extraction, synthesis and use of information.

The advantages of IoT are that users can focus on connecting the equipment, selecting the necessary options, using the results provided by the platform.

A useful platform that can be used in manufacturing systems is Oracle Integrated Cloud IoT, which provides real-time IoT data analysis, device visualization, data collection management, user notification.

Among the evaluation criteria are the IoT Platform type, which is often provided through cloud, either as a Platform as a Service (PaaS) or as a Software as a Service (SaaS). In PaaS, platforms provide cloud computing for IoT devices and data, while SaaS focuses on interconnecting data sources using cloud computing capabilities.

IV. ANALYSIS OF THE PETRI NETWORK

A Petri net type Location / Transition is form [4][5][6],

$\sum = (P, T, F, W)$ where:

- P, T are two non-empty sets, which represents the set of places and transitions respectively crowd,

$P \cap T = \phi$,

- $F \subseteq (P \times T) \cup (T \times P)$ is a binary relation, called the relation of the network flow,

- $W : F \to N$ is the function of the network share $\sum(W(f))$ is called the weight of the element (f).

If $\sum = (P, T, F, W)$ a network location type / transition is called the marking network $\sum$ any function $M : P \to N$ with the property $M(p) \le K(p)$, for any $p \in P$, where $K : P \to N \cup \{\infty\}$ is the function of network capacity $\sum$.

If $\sum = (P, T, F, K, W)$ Network P / T, where notes N^P set of all the indications of the network, obtain

$N^P = \{M \mid M : P \to N \wedge (\forall p \in P)(M(p) \le K(p))\}$

If your network has infinite capacity only when N^P coincides with the set of applications from P to N.

A network P / T is marked by a pair $\gamma = (\sum, M_0)$, where $\sum$ is a network of support network γ, and M_0 is marking the initial network γ.

Network transitions $\sum$ is considered functions $t^-, t^+ : P \to N$ i and $\Delta t : P \to Z$ defined by:

$t^-(p) = W(p, t)$,

$t^+(p) = W(t, p)$,

$\Delta t(p) = t^+(p) - t^-(p)$, for any $p \in P$.

From the entire processing flow we chose only the assembly sector for this work.

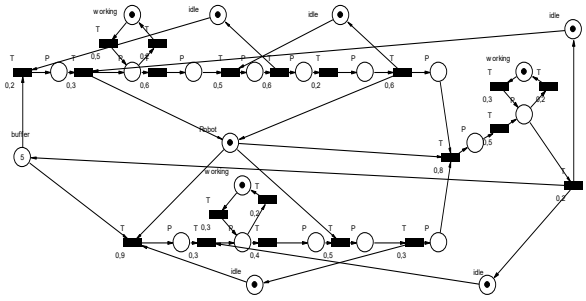


Figure 6. Model under analysis

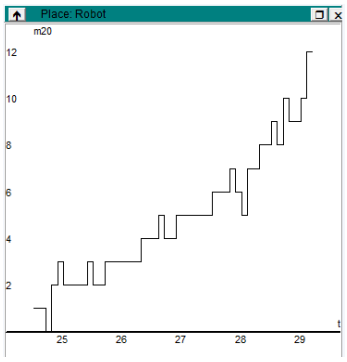


Figure 7. Robot activity variation for a complete cycle

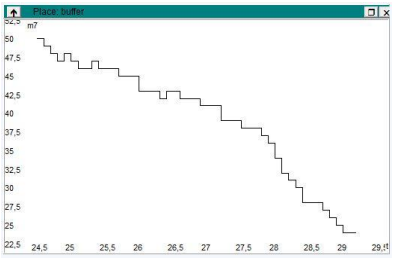


Figure 8. Activity flow from the central repository - output to processing

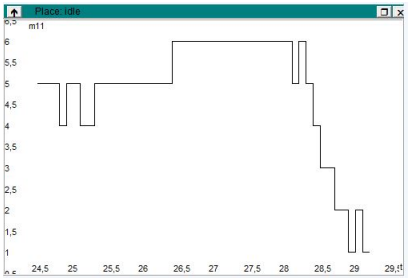


Figure 9. Temporary storage site activity

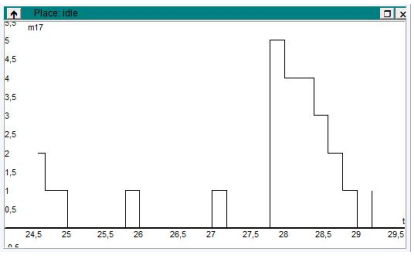


Figure 10. Assignment site storage center assembly temporarily

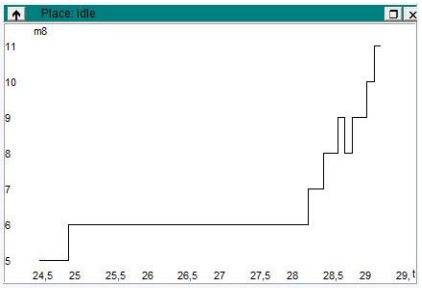


Figure 11. The activity of the permanent storage site

This system is linear. Because during normal operation, that is to say, in compliance with pre-established operating conditions, no errors are encountered, real-time monitoring of the entire process on levels will be attempted. For this I will appeal to the application of cloud systems on levels. This system was presented and analyzed in the paper [3]. Here we come to the conclusion that the proposed mechanism is flexible because the cloud manufacturing services are decoupled from manufacturing resources. The resources used can be located and obtained according to the production capacities and imposed constraints. This is how to define and outline the dynamic and automated services on which the cloud system is based.

V. HARDWARE, SOFTWARE, AND NETWORK

What would hardware need to create a digitization and turn an automated factory into a smart factory?

The general infrastructure may be something like [1]:

a) sensors, IoT devices (devices that can transfer / transmit data over the Internet) to read / capture some data from the main, if not all, automated modules involved in this manufacturing process

b) physical network infrastructure to traffic for specified data (inside an intelligent factory also specified externally between smart factories)

c) servers / modules that can process the data received from the production modules

d) Modules for storing processed data or data to be processed

e) other modules designed to organize this complex (internally) system designed for a smart factory

Maintenance for this structure of this system can be created / administered by humans (partly for the whole process) or it can also be controlled by the artificial intelligence module.

The main disadvantage is created by the high costs generated by the purchase of necessary hardware and software. Additional costs are generated by the maintenance system that must be implemented for the computing technique. In addition to the costs of purchasing and maintaining equipment that allows for virtual engineering implementation, the costs of employee training and the costs of time lost during technology implementation and verification are added.

The deficiency resulting from the application of this complex manufacturing concept, based on virtual reality, is that the digital information of a product becomes very complex and a single computer can not satisfy the necessary system resources. It takes the development of a whole computer system.

Empowering a manufacturing plant / entire system of this "brain and autonomous muscles" also contains this section of software applications that need to run on dedicated hardware / machines to have a "live system".

By analyzing the cloud, he provides all the features of a high performing software with the ability to track online data, accessing information from any device with an Internet connection.

This section may contain different types of software, such as [1]:

a) Embedded software / embedded software

b) Operating systems

c) dedicated software for tracking the entire manufacturing process or for a specific process of manufacturing the product

d) Process automation software (brain, can learn and make decisions)

e) Security Software

f) Software for data traffic

g) Comprehensive data storage software (may be in the cloud)

h) Software designed for self-control (to check the entire system for valid data and processes)

i) Software that can generate reports with / for the human interface

j) Dedicated software for system-wide support (for specified system updates and bug fixes)

All in all, this system must be able to process complex information and also be autonomous in decision-making on management processes for this intelligent factory based on the manufacturing steps of a product. It also has to have the right results (data, time and processes).

A major disadvantage is the need to use specific software tools to access digital information. In most cases, these tools are very expensive and require the same software version that created the original digital information. Given the large amount of software vendors on the market, this can be a big problem in a collaborative work environment. The solution to this problem is to simplify digital models according to the needs of each user so that the amount of digital information is reduced. In this way, the necessary hardware and software resources can be kept to a minimum.

The network is a very important part of this automated management system for a smart factory.

This network must be organized into subnetworks (kernel-based interconnected subnetworks as division-based software modules because there is no need to upload a lot of resources / overload resources) [1]. These subnets need to be connected to a larger network / main network for an entire intelligent factory and also this network may be a subnet for another larger network consisting of several internal networks forming a group of intelligent factories .

Some dedicated servers can be used with a multitude of processing power and a minimal Gigabit data system to get faster communication between the different parts of the factory and management modules.

The specified subnetworks can also be divided into automated processing point locations / points, such as robots or other hardware devices.

Cloud computing can be used to solve a lot of traffic-related issues and the status of stored data.

Also, some software applications can be directed directly into the cloud for certain tasks / process rules.

Another disadvantage is that due to the flexible manufacturing lines required, employees need to be trained faster and more efficiently so that they can cope with rapid changes.

VI. CONCLUSIONS

The main benefit of IoT is the ability to collect data and turn them into activities.

IoT is a launching platform for quantitative and qualitative transformation, monitoring and control of complex manufacturing systems.

The IoT platform is appreciated for the facilities offered, related to:

- collecting and preparing data,
- communication protocols,
- monitoring, follow-up and control,

- security,
- data analysis, processing, easy interface from system functions.

The technological advantages are: security, data analysis, monitoring, low power networks, real-time event flow processing.

The implementation of digitized technologies by large companies may hinder their suppliers, small companies. They are forced to adopt a computing technique that is compatible with that used by customers to meet the requirements, for which reason they have to increase the price of the products provided to repay their investments.

REFERENCES

[1] Nicu-Avram Pocris, Mihaela Ciortea, *Digitized production based on workflow and IoT*, MDIS 2017, Universitatea „Lucian Blaga" Sibiu

[2] Ovidiu Vermeşan, Peter Friess, *Internet of Things – From Research and Innovation to Market Deployment*, River Publishers, pp315-333, ISBN: 978-93102-94-1

[3] Mihaela Ciortea, *Prototyping manufacturing in cloud*, IOP Conf. Series: Materials Science and Engineering 227 (2017) 012028, doi:10.1088/1757-899X/227/1/012028

[4] Camerzan Inga, „*Proprietăţi structurale ale reţelelor Petri temporizate*", Teza de doctor în informatică, Universitate de Stat din Tiraspol, Chişinău, 2007-12-15

[5] Păstrăvanu Octavian, „*Sisteme cu evenimente discrete. Tehnici calitative bazate pe formalismul reţelelor Petri*", Editura MatrixRom, Bucureşti, 1997, ISBN 973-9254-61-6

[6] Păstrăvanu Octavian, Matcovschi Mihaela, Mahulea Cristian, „*Aplicaţii ale reţelelor Petri în studierea Sistemelor cu evenimente discrete*", Editura Gh. Asachi, 2002, ISBN 973-8292-86-7

[7] Mihaela Ciortea, *Analysis Hierarchical Model for Discrete Event Systems*, IOP Conference Series: Materials Science and Engineering, Volume 95, conference 1, http://apps.webofknowledge.com/Search.do?product=UA&SID=Y2bvqs5SXzZtn9g9CVD&search_mode=GeneralSearch&prID=62d1e42b-949e-4fe7-878a-0aa3a252b05a, doi:10.1088/1757-899X/95/1/012093, iunie 2015

[8] A. Sawhney, *Petri net based simulation of construction schedules*, Proceedings of the 1997 Winter Simulation Conference

[9] Ciortea Elisabeta Mihaela, *Aspecte ale producţiei digitizate utilizînd IoT*, Conferinta Nationala Multidisciplinara PROFESORUL ION D. LAZARESCU, FONDATORUL SCOLII ROMANESTI DE TEORIA ASCHIERII, ISBN 978-973-126-712-8, Cugir, Septembrie 2017

[10] Mihaela Ciortea, *IoT aspecte pentru fabricaţia viitorului*, Conferinta Nationala Multidisciplinara PROFESORUL ION D. LAZARESCU, FONDATORUL SCOLII ROMANESTI DE TEORIA ASCHIERII, ISBN 978-973-126-712-8, pp. 119-126, Cugir, Septembrie 2018

[11] Mihaela Ciortea, *Producţia digitizată bazată pe fluxul de lucru*, Conferinta Nationala Multidisciplinara PROFESORUL ION D. LAZARESCU, FONDATORUL SCOLII ROMANESTI DE TEORIA ASCHIERII, ISBN 978-973-126-712-8, pp/ 179-184, Cugir, Septembrie 2018

Supervision of the operation of digital circuits by Embedded Microcontroller

György Györök, András Dávid, Nikoletta Tolner, Bertalan Beszédes, Dániel Cseh
Alba Regia Technical Faculty, Óbuda University
H-8000 Hungary Székesfehérvár, Budai str. 45, Hungary
{gyorok.gyorgy, david.andras, tolner.nikoletta, beszedes.bertalan}@amk.uni-obuda.hu; cseh.daniel@gmail.com

Abstract: There are many electronic circuitry applications that can be used to determine circuit behaviour and operation only with the probability of time. For certain applications (medical electronics, safety engineering, operational-critical industrial applications, transport ... etc.) it is important that the quality of the circuit can be determined.
In our present article, we describe a method in which we compared the digitally produced model of a circuit, and the circuit function can be determined in relation to that-time invariant model. For this process, an embedded microcontroller environment is formed which has one of the embodied test states for combinational and sequential digital circuits.

I. PRELIMINARIES

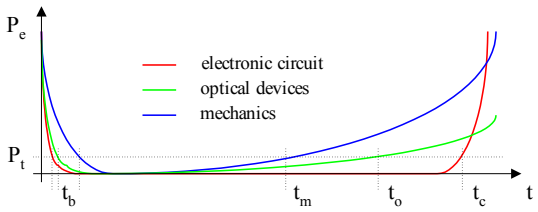


Fig. 1. Failure probability-time functions of three different technology devices.

For technical applications is often used, the bath-curve functions which is characterized failure-probability and lifetime correlation (Fig. 1.), where P_e is the probability of failure, while the horizontal axis is time (t) dimensional. We can observe three probability at three different technologies; t_m, t_o, t_c are the life expectancy typical of mechanic, optical and electrical devices, and t_b is the threshold time of burning-in failures. Above these, the probability of failure reaches the critical (P_t) value [13]. For high-reliability circuits and equipment, it is not necessarily enough to appeal to statistical probabilities.

II. STATECONTROLL OF ELECTRONIC CIRCUIT BY EMBEDDED MICROCONTROLLER

For this, we recommend the layout of Fig. 2. Here, an embedded microcontroller interrogates the voltage of the relevant points in the supervised circuit [11].

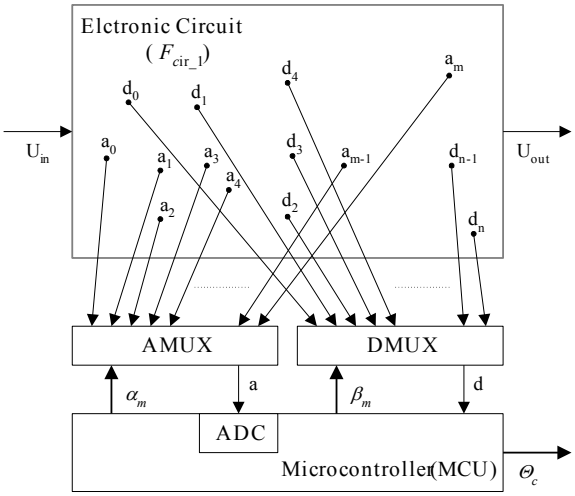


Fig. 2. Embedded microcontroller at the nodal voltage-pooling methods.

Traditional hybrid electronic circuits (Fcir-1) consist of between input (U_{in}) and output (U_{out}) a lot of electronic parts in well designed connection, a circuit network. This electronic circuit has got any significant analog (a_0, a_1, $a_{2...}a_m$) and digital (d_0, d_1, $a_{2...}d_n$) nodes [9] [10]. Microcontroller on α_m surface addressing a multiplexer (*AMUX*), and a selected voltage of a node appears on output (*a*) connected to input of analog-digital converter (*ADC*). The last one is an inner periphery of the microcontroller. Case of digital signals is used a digital multiplexer (*DMUX*), and its select surface is β_m, and output of DMUX is *d*. Important to note that multiplexing of digital signals can solve by pooling of an digital PORT of the microcontroller [14] [15].
If the voltage of the polled nodes deviates significantly (eralier gived) from the desired value or the digital signal levels are inadequate, the microcontroller generates an error signal (Θ_c). The Θ_c can either contain the serial number of the defective node and the nature of the error [1].
The current values of the nodes can be determined in three ways; or by electronic circuit simulation or manual-, or automatic measurement of a reference circuit. The values obtained, together with the tolerance value, can be compressed in a table [9].

III. STATECONTROLL BY INPUT MANIPULATION

The procedure of Fig. 2 can also be extended to input manipulation [12].
For this, it is very important to determine the moment when the test can be performed without disturbing the operation of the circuit (the whole system), now with excited control. Figure 3 shows the proposed arrangement. Here, the input signal of the circuit (U'_{in}) is taken from the normal environment (U_{in}) or is produced by a microcontroller (U_c) for testing purposes. To do this, we use an analog multiplexer (*AMUX*$_i$) that is addressed through the σ surface [3] [4].
Analog signals can be generated by a PWM method to obtain a microcontroller while digital signals can be generated with a bit-type output [2].

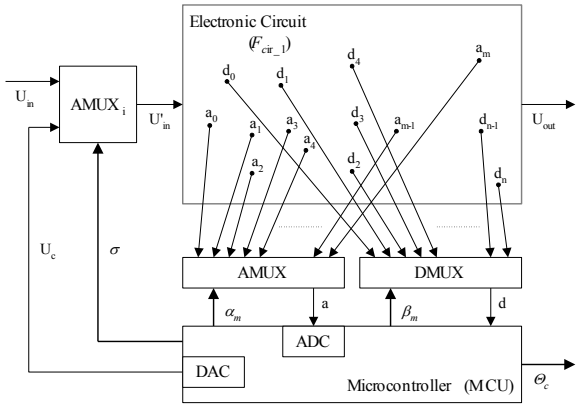


Fig. 3. The input manipulation by a hybrid circuit.

IV. VERIFICATION OF THE OPERATION OF THE PROCESS USING THE SYSTEM IDENTIFICATION METHOD

Fig. 4 shows the test procedure for most errors in the use of possible electronic circuits. Thus, we can test the system's functioning with the stimulated errors.
To test the process itself, we make a reference circuit whose operation is trivial. Thus, we use a digital circuitry where commonly used circuit solutions occur, including combinatorial and sequential networks, and tristate circuits [6][7][8].

For combination networks, it is appropriate to produce the correct total input excitation. For sequential networks, after a well-defined reset, the clock is generated by the testing.ndi networks, after a well-defined reset, the clock signals are generated and the testing can be done [5].

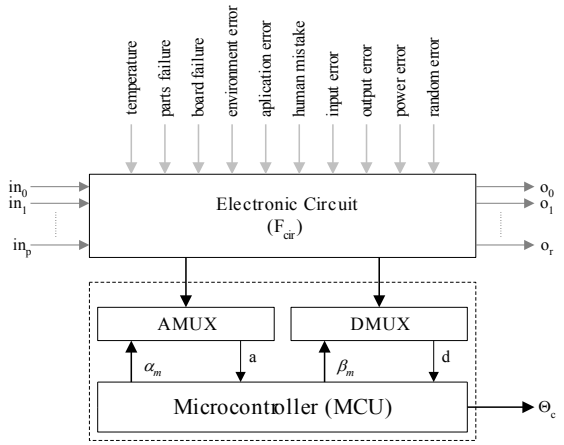


Fig. 4. The proposed circuit supervisor arrangenment with system identification approach.

V. THE TESTED CIRCUITS

The digital network contains the following combinational and sequential circuits:

Combinational networks:

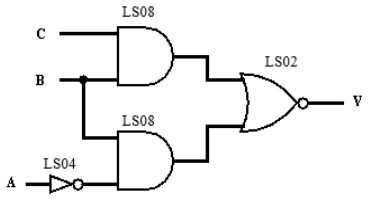


Fig. 5. Simple combinational network.

In the case of the simple combinational logic network the output is 0 if both the inputs B and C are 1's or B is 1 and A is 0.

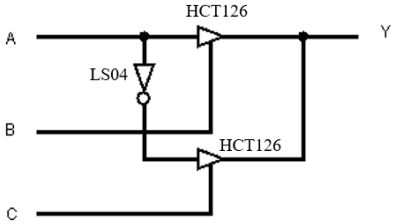


Fig. 6. Tri-state output network

In the case of the tri-state network the output is logic level 1 if the inputs A and B are 1's and the C is 0 or if the input C is 1 while A and B are 0. Applying 1's at the same time to B and C inputs must be avoided, because it can ruin the circuit.

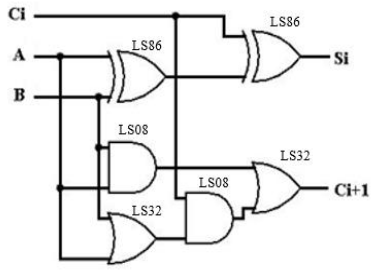


Fig. 7. Full adder

In the case of the full adder the output Si is logic level 1 if one of the inputs A, B, Ci is 1 or all of them are 1's. The output Ci+1 is 1 if at least two of the inputs are 1's.

Sequential circuits:

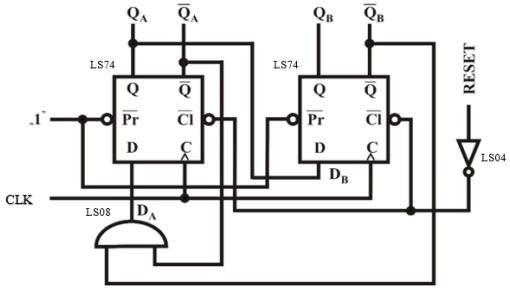


Fig. 8. Sequential circuit using flip-flops

With Reset flip-flop circuit the counter will step into 0. The count sequence is 1, 2 and then 0 again.

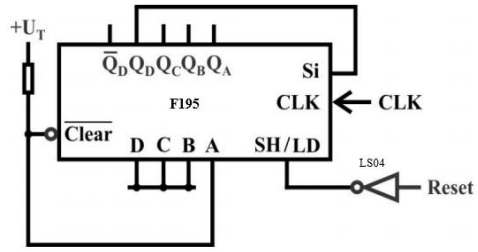


Fig. 9. Ring counter using shift register

With Reset the ring counter will step into 1. The count sequence is 1,2,4,8 and then 1 again.

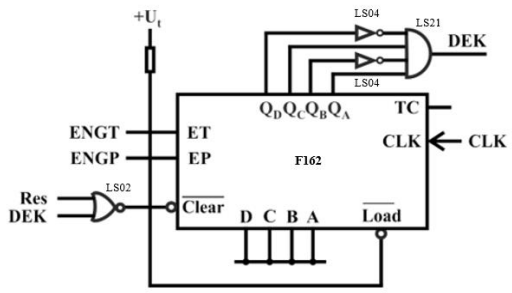


Fig. 10. Network with synchronous counter

With Reset the network will step into 0. The count sequence is 0, 1, 2, 3, 4, 5 and then 0 again. By applying logic level low to the input ET the network won't step into the next state, it will hold its present state.

VI. TESTING THE CIRCUITS

We checked the operation of the combinational and sequential networks discussed in part II by simulation and measurement. We will describe our experiences We checked the operation of the combinational and sequential networks discussed in part V by simulation about the simulation and measurement in another article [15].

VII. THE TESTED CIRCUITS ON BREADBOARD

To test the measurement control program we built the above described circuits on breadboard. We used the same type of IC's which can be seen on the schematics (Fig. 5, 6, 7, 8, 9, 10).

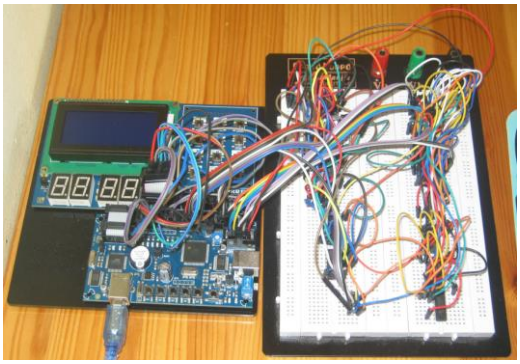


Fig. 11. The test circuit on breadboard and the development environment

VIII. DEVELOPING THE MICRO-CONTROLLER BASED TEST ENVIRONMENT

To prepare the test environment with microcontroller we used the AVR development board "T-Bird 2" with T-Bird expansion board made by BioDigit Kft. This development tool contains ATMEL AVR – Atmega128 microcontroller. On the development board we can find 5 push-buttons, which can be used to navigate in the menu. The results are shown on the LCD screen which is placed on the expansion board. The control program was written in C language.

The microcontroller board operates at 5V DC, which is connected from PC (via USB) or from external power supply. All of the IC's on the board operate at 5V DC,

which comes from the microcontroller board. The microcontroller has 128KB flash program memory, 4KB EEPROM and 4 KB SRAM data memory. On the expansion board we can find a HD44780 compatible LCD screen with 4x16 display format. In the measurement control program the clock and the reset signals which are needed to test sequential networks are generated by software. The clock has a time period of 1ms; the reset is logic high for 1 clock cycle [16][17].

Fig. 12. AVR development board „T-Bird 2”

IX. THE MEASUREMENT CONTROL PROGRAM

The already written measurement control program implements the following functions:

- quick test
- show the results
- step by step test
- manual measurement

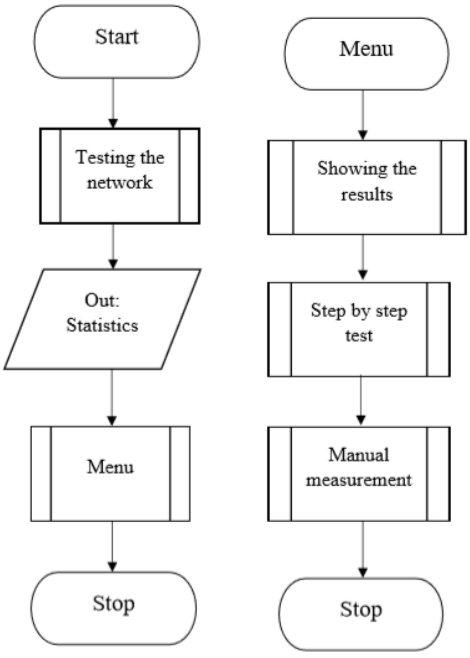


Fig. 13. The global flowchart of the measurement control program

a) Quick test

In the quick test the program measures the logic level of the output nodes. After that it compares the results with the values in the data (reference) table and prints out the quantity of errors and the number of the faulty nodes.

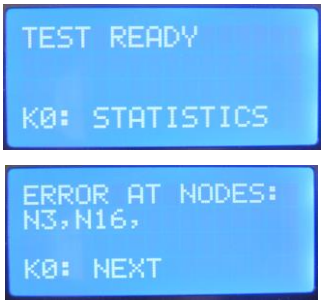


Fig. 14. An example for a running quick test

b) Showing the results

In this mode we can print out particular results of the measurement in details. In the case of a combinational network the input combinations (Fig. 15 column CBA), in the case of a sequential circuit the number of clock cycle and the reference value (Fig. 15 column R) and the actually measured value (Fig. 15 column M) appear on the LCD screen. In the picture below we can see the results of the test of a combinational network.

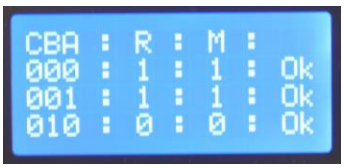


Fig. 15. An example for printing out the results

c) Step by step test

In this mode we can test the network step by step. The results are shown on the LCD screen, the signals appear on the nodes and pins, so we can measure them by multimeter or oscilloscope. By pressing the K0 button we can step further to the next combination and clock cycle. With this mode we can detect the cause of failure more accurately. On Fig. 16 we can see a detail of the test results of the synchronous counter (Fig. 10). “O” means the number of clock cycles after RESET, R is the reference value, M is the measured value. The output order is: $Q_DQ_CQ_BQ_ADEK$.

Fig. 16. Example for step by step test (tested with the network on Fig. 10)

d) Manual measurement

In this mode the combinations and the clock cycles appear on the appropriate inputs of the digital network with the timing set by the program. We can measure output signals on the appropriate nodes and pins with oscilloscope. In our previous article [15] we performed the measurement in this mode.

X. THE MEASUREMENT CONTROL PROGRAM IN DETAILS

a) Assigning the ports to the signals in case of combinational network

In case of the combinational network the inputs A and B of the schematics 5, 6, 7 are connected, so they are controlled together. The inputs C and Ci of the schematics 5 and 7 are also connected, in the case of schematic 6 the input C is an individual input, it is not connected to the input C of the schematics 5 and 7. The reason for this is that there are two input combinations of the tri-state output network of Fig. 6 in which case both tri-state gates would be enabled. Of course this combination must be avoided.

b) Assigning the ports to the signals in case of sequential network

In case of sequential network the inputs CLK and RES are common by all three schematics (Fig. 8, 9, 10), input ET is individual input.

c) Quick test

The essence of the quick test is that the program measures the logic level of the given nodes without human intervention and compares the measured value to the values of a reference data table. The program handles the combinational and the sequential networks separately. The data tables are 2x8 int type matrixes. The order of the bits of the matrixes is not the same as the pinout of the connectors. Row 0 of the matrixes contains the reference values (in case of correct operation these equal with the logic levels to be measured), row 1 contains the actually measured logic levels. In case of combinational network the indexes of the columns are the combination itself (C is the most significant bit, A is the least significant bit), in case of sequential network it is the order of the clock cycles.
We tested the tri-state output network (Fig. 6.) when connected to digital input ports, and we also tested it when connected to analogue input ports. We used different reference data table when we connected the tri-state output network to analogue input. This data table is a 2x6 char type matrix.

d) The algorithm of the quick test:

We the algorithm of the quick test into two main parts. In the first part we test the combinational network, in the second part the sequential network.

Testing the combinational network:
The testing process of the combinational network was taken further apart into two sub pieces, because to test the tri-state output network an analogue measurement is necessary as well.
The essence of the test is that we put out the 3-bit input combination to the appropriate pins cyclically. After the right timing we measure and read the logic level of the output pins, and then we save it into the first row of the reference data table.

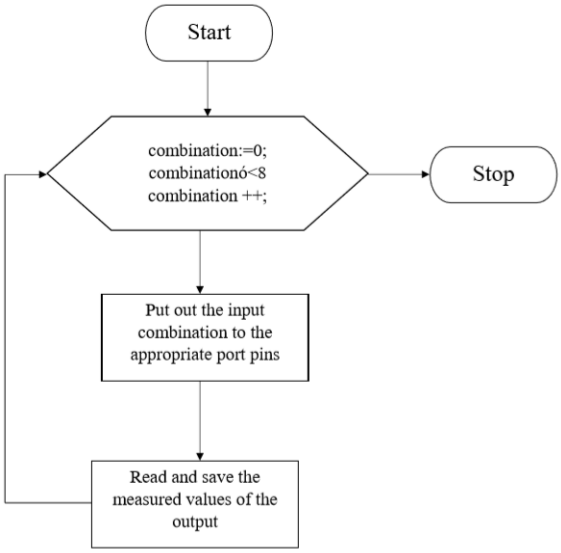


Fig. 17. The algorithm for the test of the combinational network

We tested the tri-state output network also in the way that we simply connected the output to a port that was defined as a digital input. When the output should have been in high impedance state according to the truth table, we measured 4.5V on the output, which is logic level high. The reason for this is that if a port is set as input, then we can attach an inner pull up resistor to the given port. In our case the inner pull up resistor was enabled, which pulled up the output to logic high. When we disabled the pull up resistor on the given port, we measured 1.2V on the output in high impedance state. When we tested the tri-state output network (Fig. 6), we connected the output (Y) through a voltage divider to an analogue input, and we measured the voltage of the output Y. Based on the specified voltage intervals we were able to decide if the output is logic 0, 1, or Z (high impedance state).

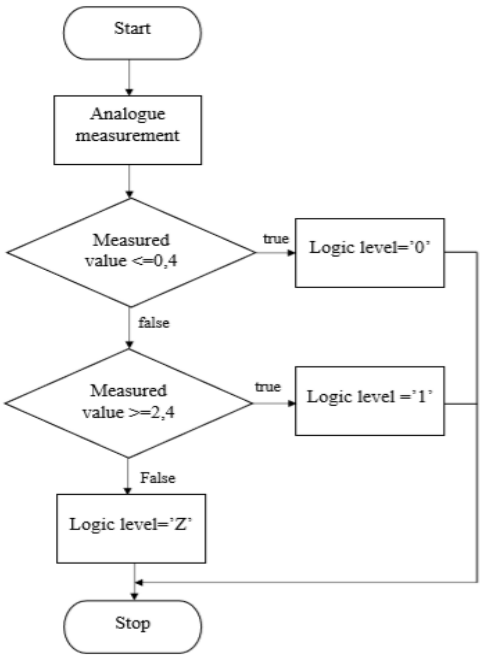


Fig. 18. The principle of the testing of the tri-state output network

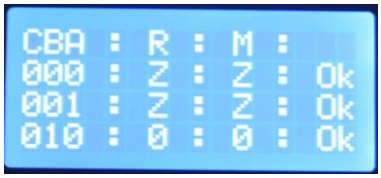


Fig. 19. Result of the testing of the tri-state output network (with the network on Fig. 6)

Testing the sequential network:

The essence of the test is that after the RESET we cyclically query and read the logic level of the outputs after the right timing during 8 clock cycles and we save them into the first row of the data table.

e) Making statistics:

When making statistics we compare the bits in the rows 0 and 1 of the reference data table one by one (row 0: reference value, row 1: actually measured value), so we can decide if the operation is correct or faulty. Every output has got a number (beginning from 0). If the measured and reference values of the given output are different from each other, we set the given bit (=1) of the statistics vector, otherwise we reset it (=0). 1 means error, 0 is the sign of the correct operation. After the program is finished the statistics is printed out. The program prints out the quantity of errors and the number of the faulty nodes (Fig. 14).

f) Printing out the results and the algorithm of the step by step test

In these two menus we can choose combinational or sequential network and after that we can choose which particular network we want to test.

When testing combinational and sequential networks this two menu offer separate algorithms for each type. In both cases we solved the printing out of the results and the step by step test with merged, parameterized functions. One parameter shows if only printing out or step by step test is needed (the difference between the two is that in case of printing out we simply read out the data from the first row of the data table, in case of step by step test we need to perform a new measurement to get the measured value). The other parameter shows which network we talk about (for example the network on Fig. 10).

```
void sorrendi_halozat_teszt(void)
{
    szinkron_reset();
    PORTC |= 0b00000100; //ET = 1

    for(int i=0; i<8; i++)
    {
        //scanning output signals
        //flip-flop
        int port_be_flip_flop = 0x00;
        port_be_flip_flop = PINC;
        port_be_flip_flop >>= 4;
        //register
        int port_be_regiszter = 0x00;
        port_be_regiszter = PINB;
        port_be_regiszter &= 0x0F;
        //counter
        int port_be_szamlalo = 0x00;
        port_be_szamlalo = PIND;
        port_be_szamlalo &= 0x0F;
        int dek_be = 0x00;
        dek_be = PINC;   //DEK
        dek_be &= 0b00001000;
        port_be_szamlalo <<= 1; dek_be >>= 3;
        port_be_szamlalo |= dek_be;

        //generating data
        int mert_adat = 0x0000;
        mert_adat |= port_be_flip_flop;
        port_be_regiszter <<= 4;
        mert_adat |= port_be_regiszter;
        port_be_szamlalo <<= 8;
        mert_adat |= port_be_szamlalo;
        adattabla_sorrendi_halozat[1][i] = mert_adat;

        orajel();
    }
}
```

Fig. 20. Detail of the program – testing of the sequential network

```
void orajel(void)
{
    PORTC |= 0b00000010;  //CLK = 1
    _delay_ms(1);
    PORTC &= 0b11111101;  //CLK = 0
    _delay_ms(1);
}

void szinkron_reset(void)
{
    PORTC |= 0b00000001;   //RES = 1
    orajel();
    PORTC &= 0b11111110;   //RES = 0
}
```

Fig. 21. Detail of the program – the code of the software generated RESET and CLK signals

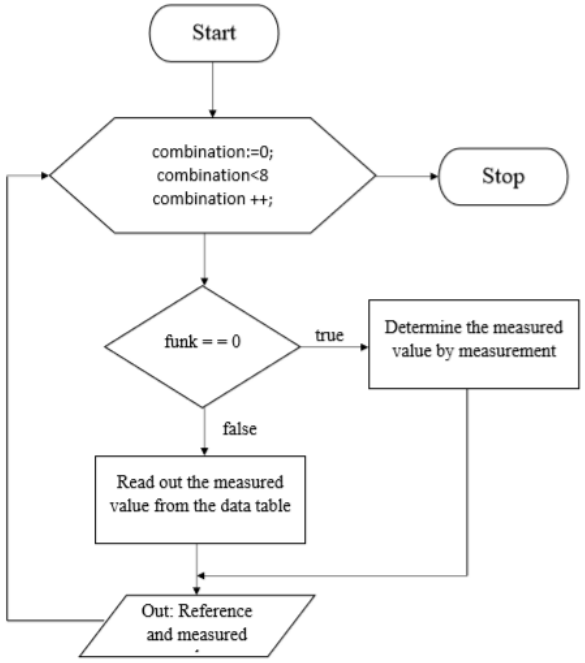


Fig. 22. The algorithm for printing out the data and for the step by step test

g) Manual measurement

In this menu we can choose if we want to measure combinational or sequential network. The program sends out the appropriate timed combinations with the RESET and CLK signals until we push the STOP button (K0), so the signals can be measured on the appropriate measurement pins. The RESET and the clock (CLK) signals are generated by software.

XI. SUMMARY

The description in this article has demonstrated the applicability of the method of digital circuits, for some testing functions.
Further important research, implementation of time-dependent measurement, introduction of the input select function and presentation of the monitoring of analog circuits.
Another interesting application is to create real-time simulation.

XII. REFERENCES

[1] N. Ádám. Single input operators of the df kpi system. Acta Polytechnica Hungarica, 7(1):73–86, 2010.

[2] J. Hurtuk B. Mados, N. Adam and M. Copjak. Brain-computer interface and Arduino microcontroller family software interconnection solution. 2016 IEEE 14th International Symposium on Applied Machine Intelligence and Informatics (SAMI), ISBN 978-1-4673-8739-2, pp. 217-221, Herlany, 2016.

[3] Gy. Györök A. E. Baklanov and B. Beszdes. Extension of Nodal Voltage Method with the Thermosensing. AIS 2017 - 12th International Symposium on Applied Informatics and Related Areas organized in the frame of Hungarian Science Festival 2017 by buda University, ISBN:978-963-449-032-6 p.p. 201-204, Hungary, 2017.

[4] Leon O. Chua and Pen-Min Lin. Computer-Aided Analysis of Electronic Circuit. 1975 Prentice-Hall, INC Englewood Cliffs, New Jersey, USA, 1975.

[5] Gy. Györök, M. Mak´o. Configuration of EEG input-unit by electric circuit evolution. Proc. 9th International Conference on Intelligent Engineering Systems (INES2005), pages 1–7, September 2005.

[6] Gy. Györök. Self configuration analog circuit by FPAA. Proc. 4th Slovakien – Hungarien Joint Symposium on Applied Machine Intelligence (SAMI2006), pages 34–37, January 2006.

[7] Gy. Györök. Self organizing analogue circuit by monte carlo method. Proc. IEEE International Symposium on Logistics and Industrial Informatics (LINDI2007), pages 34–37, September 2007.

[8] Gy. Györök. A-class amplifier with FPAA as a predictive supply voltage control. Proc. 9th International Symposium of Hungarian Researchers on Computational Intelligence and Informatics (CINTI2008), pages 361– 368, November 2008.

[9] Gy. Györök. The FPAA realization of analog robust electronic circuit. Proc. IEEE Internacional Conference on Computational Cybernetics (ICCC2009), pages 1–5, November 2009.

[10] Gy. Györök. Crossbar network for automatic analog circuit synthesis. Proceedings (Liberios Vokorokos, Ladislav Hluch, Jnos Fodor szerk.) of the IEEE 12th International Symposium on Applied Machine Intelligence and Informatics (SAMI 2014). IEEE Computational Intelligence Society,

Budapest: IEEE Hungary Section, ISBN:978-1-4799-3441-6, pages 263– 267, January 2014.

[11] Gy. Györök and B. Beszédes. Duplicated Control Unit Based Embedded Fault-masking Systems. IEEE 15th International Symposium on Intelligent Systems and Informatics : SISY 2017, Szabadka, Szerbia, 2017.09.14-2017.09.16. New York: IEEE, 2017. pp. 283-288. (ISBN:978-1-5386-3855-2, DOI, IEEE Xplore), 2017.

[12] Gy. Györök and B. Beszédes. Highly reliable data logging in embedded systems. SAMI 2018: IEEE 16th World Symposium on Applied Machine

Intelligence and Informatics, proceedings. 237 p., Koice; Herlny, Szlovkia, 2018.02.07-2018.02.10. Seattle (WA): IEEE, 2018. pp. 49-54. (ISBN:978-1-5386-4771-4, REAL), 2018.

[13] Gy. Györök, T. Trifonov, A. E. Baklanov, B. Beszédes, S. V. Grigoryeva, and A. Grigoryeva. A Special Robust Solution for Battery Based Power Supply. 11th International Symposium on Applied Informatics and Related Areas (AIS 2016), Székesfehérvár, Magyarország, 2016.11.17 Budapest: Óbudai Egyetem, 2016. pp. 32-35, Hungary, 2016.

[14] R. Kovacs L. Fozo, R. Andoga. Thermo-dynamic cycle computation of a micro turbojet engine. IEEE, Óbudai Egyetem, ISBN 978-1-5090-3909-8, Budapest, 2016.

[15] Nikoletta Tolner and András Dávid: Examination of digital circuits with simulation and measurement.

[16] Tamás Sándor, Sándor Bedő, Szabolcs Román: Use of 32-bit ARM-based microcontrollers in project tasks at the University of Óbuda. 29th Kandó Conference. Budapest, Magyarország, 2013.11.21 Budapest: Óbudai Egyetem, 2013.

[17] Tamás Sándor, Zolltán Topor, Sándor Bedő: Experiences of 8- and 32-bit microcontroller education at Atmel's Atmel laboratory at the University of Óbuda 28th Kandó Conference. Budapest, Magyarország, 2012.11.08 Budapest: Óbudai Egyetem, 2012.

Examination of digital circuits with simulation and measurement

Nikoletta Tolner and András Dávid and Bertalan Beszédes
Campus of Alba Regia Technical Faculty of Obuda University
H-8000 Hungary Székesfehérvár, Budai str. 45, Hungary
tolner.nikoletta@amk.uni-obuda.hu
david.andras@amk.uni-obuda.hu
beszedes.bertalan@amk.uni-obuda.hu

Abstract: We teach Digital Technology to the Students of the Óbuda University AMK for many years. In the curriculum there are lectures and laboratory exercises. Initially, all the measurements were performed with an oscilloscope, but later part of the measurements were simulated. From here came the idea to examine the operation of the logic circuits by simulation, then oscilloscope, and compare the results. We also made comparisons for combination and sequential networks. The results of this work are described in the article.

I. TESTED CIRCUITS

A. Combination circuits

The equation for the output of the first tested combination network: $Y = \overline{C * B + B * \overline{A}}$

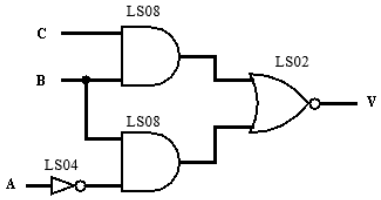


Figure 1 Simple combination circuit

We connected tri state outputs, in the second tested combination network. Such outputs should only be connected if it is necessary to ensure that only one gate is enabled at a time. This is achieved by the B and C inputs of the corresponding control.
If the upper gate is enabled, the input A is displayed on the output, if the lower gate is enabled, the input A negated is displayed on the output.

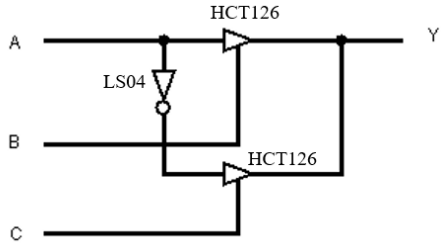


Figure 2 Tri state output

The third examined combination network is a full adder.
The equation of the output:
$Si = A * B + (A + B) * Ci$
$Ci = A \oplus B \oplus Ci$

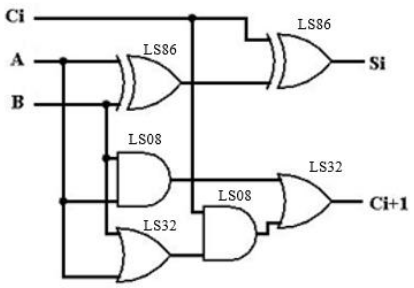
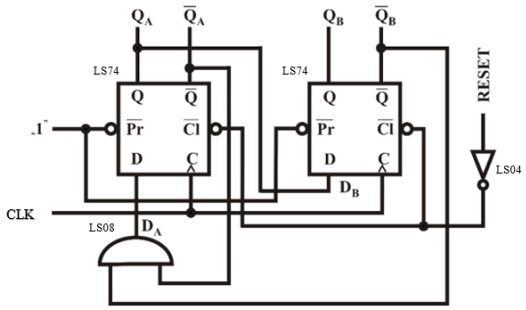


Figure 3 Full adder

B. Sequential circuits

The first sequential circuit is a network of two D flip flops. Reset the network to 0, then from there to the rising edge of clock 1, 2 and again 0.



Figurer 4 Sequential circuit with flip flop

The second sequential circuit contains a shift register, its name is ring counter. Reset is loaded on the 1, then on the rising edge of the clock, the 2, 4, 8, and again 1 states the network.

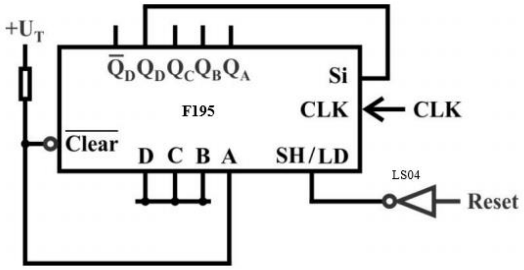


Figure 5 Ring counter

The third sequential circuit contains a decimal synchronous counter. Reset is deleted, that is synchronous, that is, the rising edge of the clock. Reset 0, 1, 2, 3, 4, 5, and 0 again.

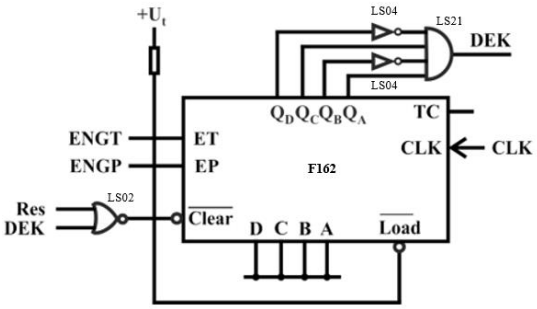


Figure 6 Sequential circuit with synchronous counter

II. THE PROGRAMS USED FOR THE SIMULATION

A. Logisim

Logisim is an open source program. It is free download and you do not have to install it immediately to run. We were looking for a program which can be used easily, by which we can demonstrate the function of the digital electronic circuits. So we found the Logisim program. It is easy to operate and has a transparent design. It was for educational purposes. The simple toolbar interface and the simulation of circuits make it easy to get familiar with the most important concepts related to logical circuits. It runs on any machine that supports Java 5 or later. Color coded leads help to simulate and troubleshoot a circuit. The finished circuits can be saved in a file, exported to a GIF file, or printed. We tested the operation of our circuits with this program.

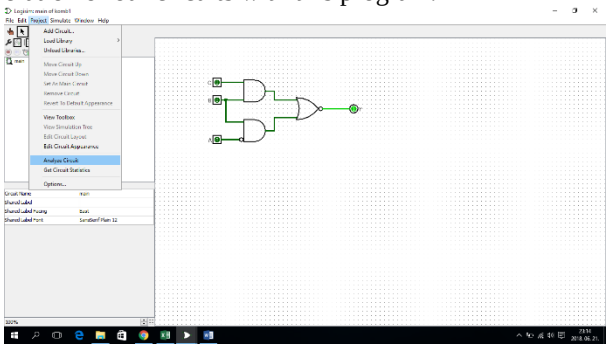


Figure 7 Logisim program

B. ISIM simulator

The Xilinx ISE WebPack (Integrated Software Environment) is a free software developed by Xilinx for FPGAs and CPLDs, which is freely downloadable from the internet at the company's website. The development system includes all the elements needed for circuit diagram basis and hardware description language basis to development. The designer his ideas, his plans:

- can input it in Schematic, using the wiring diagram creation and input program.
- can enter hardware description language. This is supported by HDL editor. Supported languages: Verilog and VHDL. After entering, we can check if our circuit is working properly. The control is done by simulation. The simulator of the WebPACK system is the Xilinx ISE Simulator. For simulation, the circuit must be "energized", variable signals must be given in the inputs of the circuit . This is done by test vectors. Test vectors can be specified by the designer with VHDL description (test bench). Circuits tested at the measurements were implemented by a colleague in CPLD. We had the program system in place so we chose this to test our logical networks. With the help of the simulator we examined the timing of the circuits

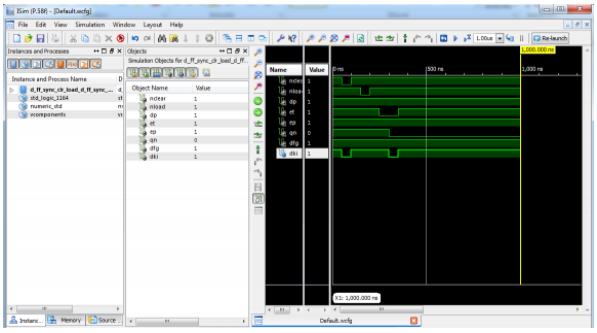


Figure 8 ISIM simulator

III. THE RESULTS OF SIMULATIONS

A. With Logisim

This program was used to test the logic of digital circuits. The circuits were built from the gates in the program library and from the basic sequential circuits (flip-flop, shift register, counters). In the case of combination networks, we have recorded the truth table with the program. It is not possible to add a status table of sequential networks because the clock cannot be considered as input by the program. For sequential circuits, we tested the outputs with the proper excitation of the inputs. We outline below the results of the simulation. When simulating a simple combination network, we have the following truth table:

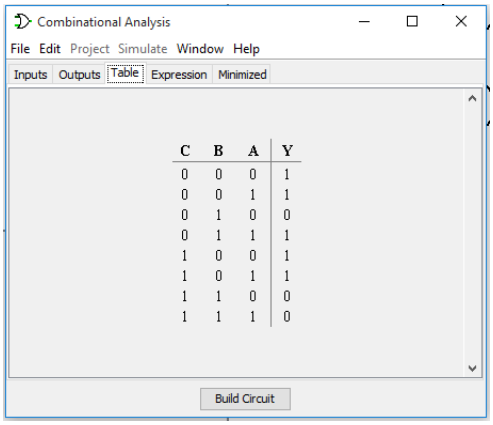


Figure 9 The truth table of simple combination circuits

In the circuit with tri-state gates, we get the following truth table

A	C	B	Y
0	0	0	x
0	0	1	1
0	1	0	0
0	1	1	!!
1	0	0	x
1	0	1	0
1	1	0	1
1	1	1	!!

Figure 10 The truth table of tri-state circuit

The output x of the output (Y) actually indicates a high impedance state when B and C are low. The „!!" means the control is prohibited when both HCT126 are enabled (B = C = 1).

At the full adder the result of the simulation is as follows:

Ci	A	B	Si	Cil
0	0	0	0	0
0	0	1	1	0
0	1	0	1	0
0	1	1	0	1
1	0	0	1	0
1	0	1	0	1
1	1	0	0	1
1	1	1	1	1

Figure 11 The truth table of the full adder

For sequential circuits, we could not add a state table to the program, but we tested the outputs by the excitation of the inputs. In the case of a Flip Flop Circuit, we get a sequence of states: Res-> 0,1,2 and again 0.

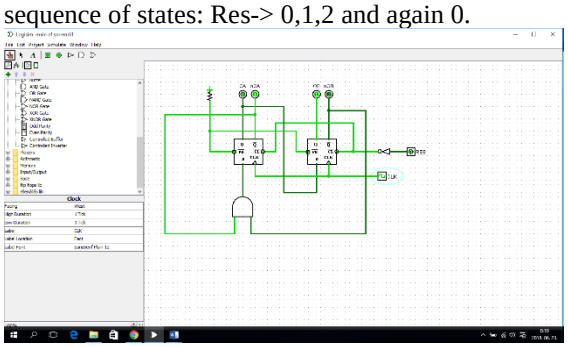


Figure 12 Simulation of the flip-flop circuit

In the case of ring counter is a sequence of states: Res-> 1,2,4,8 and again 1.

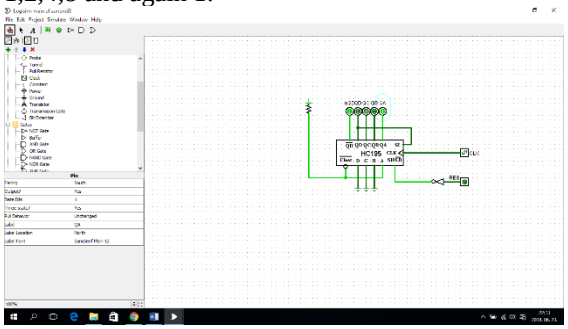


Figure 13 Simulation of the ring counter

In the case of synchronous counter is a sequence of states: Res->0,1,2,3,4,5 and again 0If the EP or ET input is 0, the circuit has not gone further, the outputs have not changed.

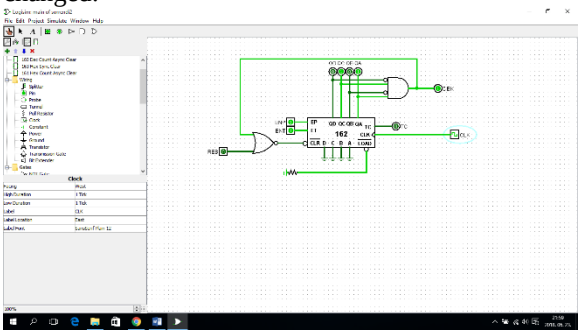


Figure 14 Simulation of the synchronous counter

We received the expected results for all the simulations.

B. With ISIM simulator [6]

In order to come closer to the properties of the circuit elements we used real components in the simulation, we have described them in VHDL as their function. Using the components in the description language of hardware (making them a symbol), we built up the combination and sequence networks outlined at the beginning of the article. Using the built-in symbols, we constructed the tested circuits with the XILINX company's free program package (ISE WebPack).

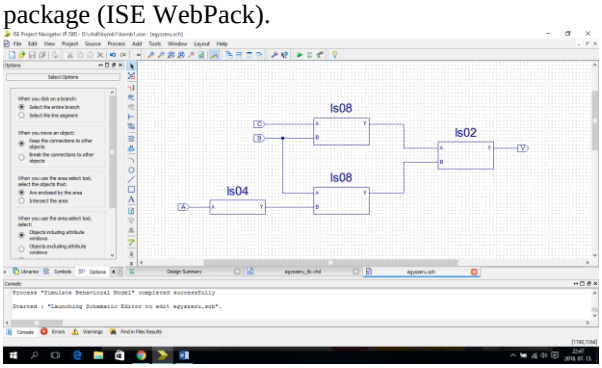


Figure 15 The implementation of a simple combination network with Ise Webpack

For the simulation of the circuits, the state of the inputs had to be determined. This was described in VHDL. After filling the test bench, we started the ISIM simulator, which outlined the circuit time diagram of the input excitation signals. The 8 possible logical values differentiated by the program are represented by different colors in the time diagram. Some colors refer to what we will need, The green color indicates a definite logic values 0 and 1. The blue color is characteristic of the high impedance state. The orange color indicates that the test signal is not determined. The red color indicates a fault simulation. The tri-state gates outputs connection, means that each tri-state gate was enabled. This control should be avoided. The clock signal period time is 100 ns for the flip-flop circuit, the clock signal period time of the shift-register and synchronous counter circuit is 200 ns. The Reset signal is high for a clock period. These were described in the test bench for each sequential circuit.

The detail of the test bench set in the ring counter is the following:

```
CLK_process :process
  begin
        CLK <= '0'; wait for CLK_period/2;
        CLK <= '1'; wait for CLK_period/2;
  end process;
-- Stimulus process
  stim_proc: process
  begin
    RES <= '0';
    wait for clk_period;
    RES <= '1';
    wait for clk_period;
    RES <= '0'; wait;
  end process;
```

In a time diagram, the delay times are also visible.

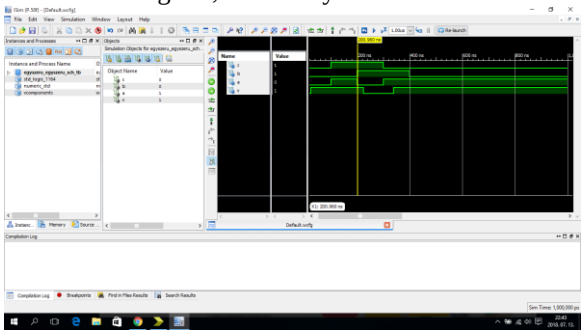


Figure 16 The simulation of simple combination network

In the simple combination network, the output is logically 0 when input B and C are 1 or B is 1 and A is 0.

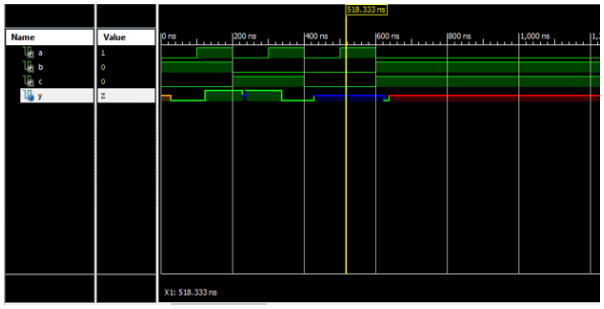


Figure 17 The time diagram of tri-state circuit

The tri-state network output is a logic value 1 when the input B is 1, the C input is 0, and A input is 1, or the C input value 1, B 0 and A is 0. The blue color indicates the high impedance state (B=C=0), the red color shows the forbidden control (B=C=1).

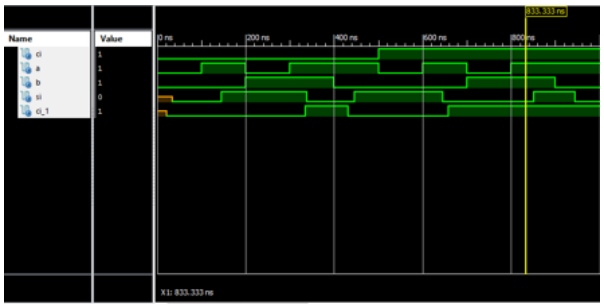


Figure 18 The time diagram of the full adder

Si output of the full adder, this is 1 if one of the A, B, Ci inputs is 1 or all 1. The Ci + 1 output is 1, if at least two inputs are 1.

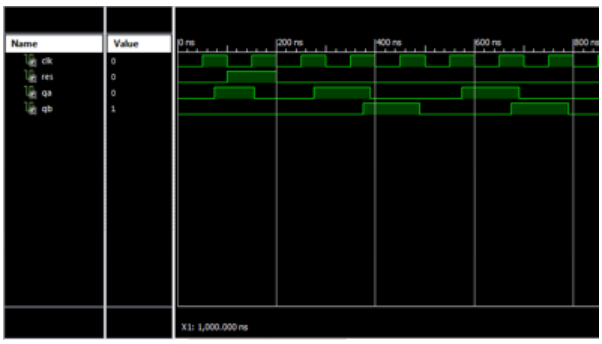


Figure 19 The time diagram of flip flop circuit

The flip-flop circuit reset to 0, then 1, 2, and again 0 will appear on the outputs.

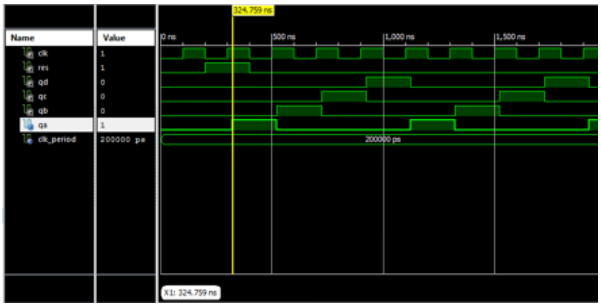


Figure 20 The time diagram of the ring counter

The ring counter to Reset 1, then 2, 4, 8 and again 1 to the outputs.

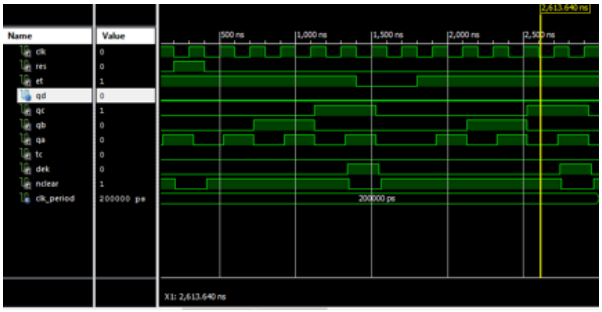


Figure 21 The time diagram of the synchronous counter

The network resets to 0, 1, 2, 3, 4, 5 and again 0. When the ET is low, the network keeps the current status.

Each simulation result was in line with expectations. In the tri-state network, due to the timing of the excitation signals, we saw spikes in the time diagram. In the case of the blue spike, tri-state gates was not allowed for very short periods of time. (Figure 22 top time diagram). Blue color indicates the high impedance state. In the case of a green spike, due to the delay times, one gate was enabled for a short time. (Figure 22 below).

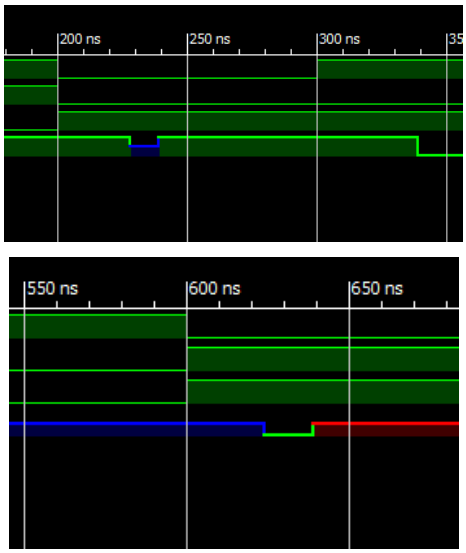


Figure 22 Tri-state network is an enlarged detail of the signals

IV. OSCILLOSCOPE MEASUREMENTS

A. The built circuit

We built the digital circuits on a test panel. We worked with on the circuit diagrams (Figures 1, 2, 3, 4, 5, 6) visible IC number. We generated input excitation signals using a microcontroller development environment. The microcontroller test environment is described in another article. [4] The built circuit operates on a 5V supply voltage. We used the T-Bird 2 AVR development platform produced by BioDigit Kft. It includes ATMEL AVR - Atmega128 microcontroller. The microcontroller panel is operated by a 5V power supply received from a USB connector of the computer or an external adapter.

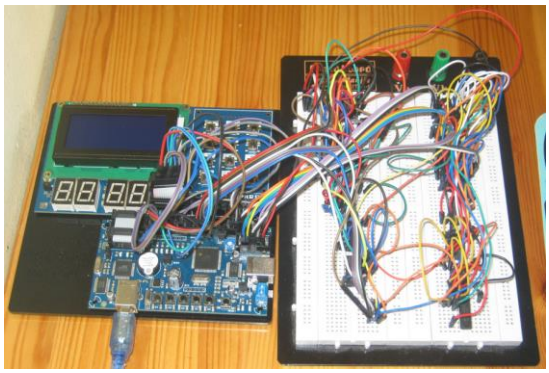


Figure 23 The circuit built on the test panel and the development environment

B. Used oscilloscope

Tektronix TDS 1012B oscilloscope was used for the measurements. is a digital storage oscilloscope with 2 channels, a separate trigger source, 100 MHz bandwidth, 1 GS / s sampling rate, and monochrome LCD display.

Figure 24 TDS 1012B oscilloscope

C. Measurement results

We generated the clock and reset signal for sequential networks by a software. The period of the clock signal is 2 ms, the reset signal is high for 1 clock cycle. We have examined the operation of all combination and sequential networks. For each measurement, the input excitation signals were generated using the microcontroller software. For oscilloscope measurements we can only examine two signals at a time, so it is a bit lengthy to examine the operation of the circuits. The triggering signal is always selected for the desired switching. The trigger came from an external source, set to edge and falling edge. For all measurements, the expected result was obtained. The following figure shows the waveforms displayed on the oscilloscope display.

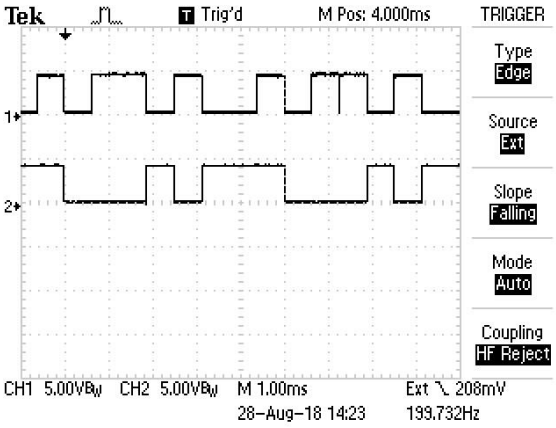


Figure 25 The waveform measured at the Si and Ci + 1 outputs of the full adder

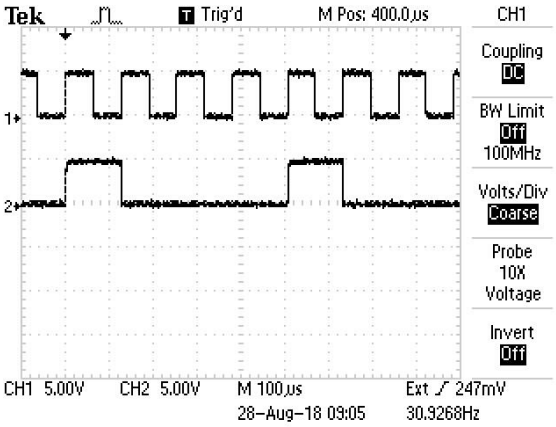


Figure 26 At the ring counter, waveforms of the clock signal and the QA output

V. COMPARISON OF VARIOUS MEASURING RESULTS

We tested three different methods for the operation of digital circuits. Out of the three methods, testing with the Logisim program is the simplest. With the help of the program we introduced the circuits and then checked their operations. We could only check the logic function with Logisim. Delay times could not be taken into account. The program gave a truth table of combination networks. For sequential networks, we manually changed the clock to check the operation of the circuits step by step. We recommend this program for beginners who are familiar with digital technology. In the second case, we tested the operation of the digital networks by the Xilinx software package with ISE WebPack. We described the operation of the circuit elements in VHDL. From the base elements we built the examined circuit, which were then examined with the ISIM simulation program. Input excitation signals are also described in VHDL. Digital networks were implemented using the schematic editor of the program package. This type of solution requires more serious knowledge. The operation of the combination and sequential networks was also examined in the form of a time table. We could measure the time data using the markers, the program is not suitable for voltage measurement. The possible logical values are indicated by different colors in the simulation program. We could not describe all the attributes of the real circuit elements in the program. There were differences between the built circuit and the simulation circuit. We could set delay times but we could not take into account setup and hold times. The period time of the clock was 100 ns and 200 ns, for the actually built circuit, this data was 2 ms. The third solution is the oscilloscopic measurement. In this case we built the circuit to be tested on the test panel. At a voltage of 5 V, we controlled our digital circuits with a microcontroller. In oscilloscope measurements, we could only examine two signals simultaneously, and the trigger signal was also important. In this case, we could accurately measure time data and measure voltage. This is lengthy, but the most accurate measurement is achieved by this method. In the built circuits, the tri-state gates (HCT126) we had to be careful not to allow both gates to be simultaneously enabled. This was not the case with the other two methods, but the built-in circuit would have caused the ICs to fail in this forbidden control.

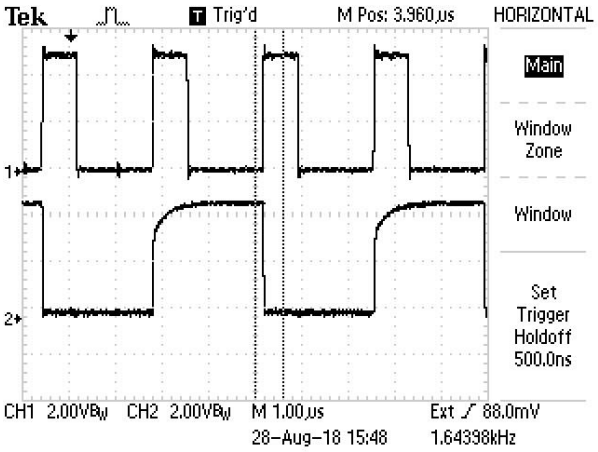


Figure 27 For a synchronous counter circuit, the clock signal and the QA output signal on the oscilloscope

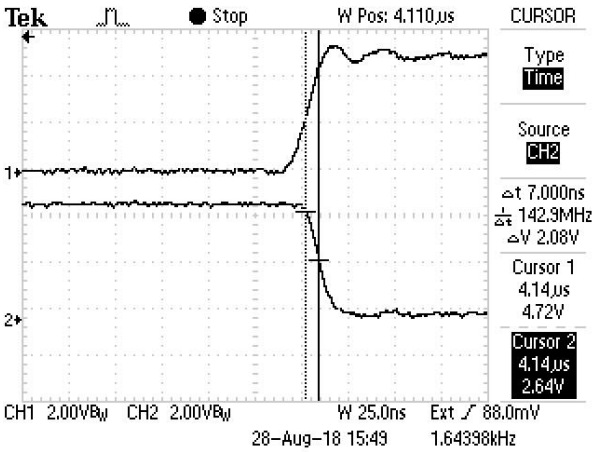


Figure 28 For synchronous counter circuit, the clock signal and QA output signal are enlarging

VI. REFERENCES

[1] http://www.cburch.com/logisim/
[2] https://www.xilinx.com/products/design-tools/ise-design-suite/ise-webpack.html
[3] https://www.xilinx.com/products/design-tools/isim.html
[4] Supervision of the Operation of Digital Circuit by Embedded Microcontroller (György Györök, András Dávid, Nikoletta Tolner, Bertalan Beszédes, Dániel Cseh)
[5] https://www.tek.com/oscilloscope/tds1012b
[6] The usage of programmable circuits in the education of Digital Technology (Nikoletta Tolner, András Dávid) AIS 2017

Modern Experience of Dementia Classification

Maryam Naderan*, Jozsef Halasz**, Seebauer Marta***

* Igor Sikorsky Kyiv Polytechnic Institute/MMSA/Kiev/Ukraine

**Obuda University/Alba Regia Technical Faculty/Szekesfehervar/Hungary

***Obuda University/Alba Regia Technical Faculty/Szekesfehervar/Hungary

Email1: ma.naderan@gmail.com

Email2: seebauer.marta@amk.uni-obuda.hu

Email3: halasz.jozsef@amk.uni-obuda.hu

Abstract. Experience of recognition Alzheimer`s dementia at the earliest stage needs attention of a wide range of information of the practicing staff, involved in diagnostics and treatment. Using Convolutional Neural Network provides additional possibilities for clear and detailed recognition of the actual stage of the disease, comparison with another objective data present in the database. The medical procedure needs complex implementation of the described experience into clinical practice. The main aim of this article is to develop an alternative approach of using modern technical data, new methods in recognition and treatment of Alzheimer`s dementia explained in scientific literature and practical activities using Convolutional Neural Network to provide further development of medical science. This paper focused on comparing Converted and Nondemented patients and used the ensemble learning methods to the development of the diagnosis of Alzheimer's Disease.

Keywords: Alzheimer's Disease, Classification, Convolutional Network, Diagnosis, Model

I. INTRODUCTION

Development of a new network provides patients with Alzheimer`s dementia with a possibility to obtain a clear and detailed diagnostic of the effective status, that has special impact on further progress of the disease, methods of treatment, quality of patient`s life etc.

Convolutional neural network is a modern effective tool for at-site recognition and deep learning of actual status of patients with diagnosis of Alzheimer`s dementia. As we noted in our previous publications, this disease has common features with another psychological and neurological diseases, thus resulting into incomplete diagnostics and inadequate treatment. Modern tools of Convolutional neural network provide additional technical and statistical data providing effective recognition of Alzheimer`s dementia at the earliest stage, that has special importance for the patient. Convolutional neural network constitutes a type of computerized deep learning, that provides analysis of effective data obtained and classified according to the main groups of patients, using objective data: MRI images of different stages of Alzheimer`s dementia, images of patients without such status for objective comparison, video, sound and text information describing the mentioned data. Application of Convolutional neural network algorithms let us as professionals not only to participate in deep learning of actual patient`s status, but also to contribute to effective recognition and further treatment of a patient.

II. RELATED WORK

Practical side needs special attention, as the network unites objective information not only of general character: these are important images data, fixing date-to-date progress of Alzheimer`s dementia patients, as well as information about qualitative and quantities changes in objective data. It may include images, sound fixed information and documents, faces and scenes, that gathered together constitute a unique database for further research and analysis. Convolutional neural network provides information about the most important aspects of recognition in case of Alzheimer`s dementia, these are as follows: object recognition; computer vision; self-driving vehicles; face-recognition application etc. These data provide the researcher with objective data and sources for further treatment and progress of disease. Convolutional neural network uses special models, such as scratches, pre-trained models. These models may be applied to different databases, built on examples, uniting patients from different countries, age etc.[1]

Convolutional neural network is a modern tool for sharp recognition not only of Alzheimer dementia, it may be effectively used for many other cases and incidents. As Convolutional neural network eliminates the need for writing, sending, manual exam, that is substituted by computer analysis and technical data recognition. The obtained results have a very high level of trust, as are built on objective data, fix up-to-date progress and shows exact changes in the patient`s status. Moreover, modern facilities provide the possibility to change recognition tasks, databases.

"Convolutional neural network provides an optimal architecture for image recognition and pattern detection. Combined with advances in GPUs and parallel computing, Convolutional neural networks are a key technology underlying new developments in automated driving and facial recognition" [1, p.2].

General feature of Alzheimer`s disease was described by scientists S. Sarraf, G. Tofighi and J. Anderson in publication devoted to classification of the considered features using convolutional neural networks, that contain results necessary for distinguishing classification criteria and significant features taken into consideration for the patients with the mentioned disease. Authors stated, that "early detection and classification of Alzheimer`s disease are critical for proper treatment and preventing brain tissue damage. Alzheimer's disease has a certain progressive pattern of brain tissue damage. It shrinks the hippocampus and cerebral cortex of the brain and enlarges the ventricles" [2].

Convolutional neural networks are used in different approaches, for example there exist special recourses R-CNN – Regions with Convolutional neural networks, that unite knowledge, data and methodology for object detection, recognition and deep learning. For example, nowadays methodology for Alzheimer`s dementia recognition includes such tools as: Parallel computing toolbox, neural network toolbox, statistics and machine learning toolbox etc.

"Instead of classifying every region using a sliding window, the R-CNN detector only processes those regions that are likely to contain an object. This greatly reduces the computational cost incurred when running a CNN." [1, p.2]

In Ukraine effective recognition of Alzheimer`s dementia is usually performed using MRI, as cortical atrophy leads to such a state of patient`s brain, that it`s volume becomes smaller. "On MRI, this is manifested by an increase in the ventricles and fissures of the hemispheres. And although MRI in Alzheimer's disease reveals changes, mainly in the moderate and severe stages of the disease, already in the early stages of Alzheimer's

on MRI, a decrease in the size of the hippocampus is revealed. And if in half a year there is an increase in the atrophy of the cortex and the hippocampus - this is a sign of the progression of the disease" [3].

III. METHODS

For this research, Kaggle's data about patients Nondemented and Demented was used. We want to focus on what could be important causes which are lead to AD and find the methods for early diagnosis with high accuracy. By using the GLM analysis, there was a statistically significant difference according to MMSE values between the three mentioned groups. In the GLM analysis (table 1), there was a statistically significant difference according to MMSE values between the three mentioned groups. [5]

TABLE 1

GLM ANALYSIS OF MINI-MENTAL-STATE-EXAMINATION

GLM MMSE	SS	Degr. Of Freedom	MS	F	P
Intercept	69701.16	1	69701.16	13542.23	0.00
Group	556.36	2	278.18	54.05	0.00
Error	756.60	147	5.15		

TABLE 2

NUMBER OF PATIENTS FOR EACH GROUP

Group	Count	Cumulative Count	Percent	Cumulative Percent
Nondemented	72	72	48.00000	48.0000
Demented	64	136	42.66667	90.6667
Converted	14	150	9.33333	100.0000
Missing	0	150	0.00000	100.0000

In contrast, Newman-Keuls post hoc comparison revealed significantly lows MMSE values in demented groups vs. Nondemented groups, while the converted group data were not significantly different from nondemented group data (table 3):

TABLE 3

COMPARING THREE DIFFERENT GROUPS IN POST-HOC NEWMAN-KEULS

Post-hoc Newman-Keuls	Group	{1} 29,194	{2} 25,328	{3} 29,357
1	Nondemented		0.000009	0.782223
2	Demented	0.000009		0.000022
3	Converted	0.782223	0.000022	

IV. RESULTS

Based on the results and Random Forest Algorithms, we determined three important features for diagnosis AD:

- Gender;
- Age;
- Education.

Practical aspects of application of Deep Learning algorithm. This algorithm provides grounds for classification into main groups of data gathered from the examined patients. It has particular role for improvement of modern diagnostic of Alzheimer`s dementia in clinics and hospitals as unites clear and detailed data about all recorded patients and stages of the disease.

Deep Learning can be considered as an important approach in computer learning developed for human brain analysis. The applied technologies combine algorithms and neural network architecture for further neuroscience based research.

Deep Learning or Deep machine learning "focuses on computational models for information representation which exhibits characteristics like those of the neocortex (Jia et al., 2014) (Ngiam et al., 2011)" [6]. To perform Deep Learning there are applied special methods, divided into separate stages: pre-processing; data conversion stage and classification stage.

In practice, recognition of patients with Alzheimer's dementia features can be made using different methods, that requires analysis of objective status at the above-mentioned stages, and involves machine learning of the defined features, that makes it possible to reveal functional and structural dissimilarities in human brains and disease progress.

As described in publication of Saman Sarraf, Ghassem Tofighi "DeepAD: Alzheimer's Disease Classification via Deep Convolutional Neural Networks using MRI and fMRI", In current paper, there were demonstrate and provide two strong pipelines and reproducible results. Inside the early block, intensive knowledge preprocessing was performed against fMRI and magnetic resonance imaging knowledge, that removed potential noise and artifacts from the information. Next, a convolutional layer of CNN design consisting of a collection of learnable filters, and that conjointly is a shift and scale in-variant operator, extracted low- to mid-level options (as well as high-level options in GoogleNet). Within the fMRI pipeline, each adopted LeNet and GoogleNet design were trained and tested by a huge variety of pictures created from 4D fMRI statistic. moreover, removal of non-functional brain pictures from knowledge improved the accuracy of recognition [6, p.9].

V. DISCUSSION

To conclude the performed analysis, we would like to make the following recommendations as to development of using Convolutional neural network tools for early recognition of Alzheimer`s dementia:

1. The combination of study of anamnesis, psychological testing, laboratory and instrumental research methods allows to recognize Alzheimer's disease with great accuracy. Using Convolutional neural network provides new research tools and practical possibilities for detection of concrete stage of disease, to evaluate its progress and further treatment proposals.
2. Based on research, there are three main factors, which could be used for diagnosis AD. Although, genetic, healthy lifestyle and medical history are not less important factors.
3. Unfortunately, in Ukraine studies on Convolutional neural networks are a rare example, so we would recommend to the Government to implement special program oriented on complex integration of new methods for computerized recognition of Alzheimer`s dementia in clinical practice.
4. The explained statistical basis and Convolutional neural network database sources might constitute grounds for future theoretical and practical research of new methods, tools, positive and negative experience of patient's treatment at different stages of Alzheimer's dementia progress.
5. Another important conclusion refers to establishing modern registries and databases in Ukraine, containing actual information and MRI results regarding different groups of patients, classified according to the defined criteria into: non-demented, early demented (very mild), mild, moderate Alzheimer`s disease. Such registries play the leading role in clinical practice, as provide exact information as to the quantity of patients, qualitative changes fixed in their disease progress, it also provides a necessary basis for comparison with close examples and possible consequences.

REFERENCES

1. Convolutional Neural Network. 3 things you need to know. P.1-4. https://www.mathworks.com/solutions/deep-learning/convolutional-neural-network.html

2. S. Sarraf, J. Anderson, and G. Tofighi, “Deepad: Alzheimer’s disease classification via deep convolutional neural networks using mri and fmri,” bioRxiv, p. 070441, 2016.

3. Альцгеймера болезнь. Диагностика. Анализы и инструментальные исследования. URL: http://demenciya.com

4. J. Islam, Y.Zhang. An Ensemble of Deep Convolutional Neural Networks for Alzheimer's Disease Detection and Classification, Computer Vision and Pattern Recognition. https://arxiv.org/pdf/1712.01675.pdf

5. Jacob Boysen. Magnetic Resonance Imaging Comparisons of Demented and Nondemented Adults. URL: https://www.kaggle.com/jboysen/mri-and-alzheimers

6. Saman Sarraf, Ghassem Tofighi. DeepAD: Alzheimer's Disease Classification via Deep Convolutional Neural Networks using MRI and fMRI, Alzheimer's Disease Neuroimaging Initiative2016, p.2

7.T. Brosch, R. Tam. Manifold learning of brain mris by deep learning, International Conference on Medical Image Computing and Computer-Assisted Intervention.Springer,2013, pp. 633–640.

8.A. Gupta, M. Ayhan, and A. Maida, Natural image bases to represent neuroimaging data, International Conference on Machine Learning, 2013, pp. 987–994.

Performance Analysis of Character case-sensitive and case-insensitive Classification in Handwritten Character Recognition

Gaye Ediboğlu Bartos*, Éva Hajnal **, Yasar Hoşcan *
* Anadolu University/Department of Computer Engineering, Eskişehir, Turkey
**Óbuda University Alba Regia Technical Faculty, Székesfehérvár, Hungary
gayeediboglu@anadolu.edu.tr, hajnal.eva@amk.uni-obuda.hu, hoscan@anadolu.edu.tr

***Abstract*— In character recognition classes are letters, numbers and punctuations. Therefore, the number of classes depends on the number of characters in the language. However many classes (such as upper case "C" and lower case "c") have very similar characteristics therefore merging such classes is also an option for a more successful recognition. In this study, we aim at evaluating the effects of abovementioned phenomenon for handwritten character recognition by performing the recognition in three different set of classes namely case-sensitive (52 classes), case-insensitive (26 classes) and similarity based (38 classes) using Deep Feedforward Networks. Looking at the results, as expected case-insensitive classification outperformed case-sensitive classification. Surprisingly, similarity based classification having a greater number of classes resulted in better accuracy rate compare to case- insensitive classification.**

I. INTRODUCTION

Handwriting recognition is heavily researched both by researchers and technology makers since recognition of handwritings has become a feature of mobile devices such as smart phones and tablets. However, the success is still limited to discretely written characters. When it comes to cursive handwritings or characters with special accents, the success of the recognition starts to decrease [1]. The challenge with such recognition mostly derives from individuals having various types of handwritings and certain characters having similar shapes such as "o" and "O". These bring about the complications and ambiguities for the recognition process. Additionally, having a large set of class label also has a negative effect on the success of the classification. In order to tackle such errors, it is possible to put the similar letters into the same class and perform classification with a smaller number of classes. Having more samples in each class and fewer class labels is believed to increase the accuracy rate in most cases. In this study, we aim at evaluating the performance of three different classification strategies using a cursive handwritten character database which are namely case-sensitive (52 classes), case-insensitive (26 classes) and similarity based (38 classes) classes two layer feed-forward networks

II. RELATED WORKS

In 2003, Koerich published a study in which he evaluated the performance of different classification strategies on NIST handwritten database [2], [3] using Multilayer Perception Classifier (MLP). The results indicated that 26 letter meta class classifier outperformed the other two strategies which were combination of two 26-class classifier and 52-class classifier. In 2012 Sulistiyo et. al. adopted another classification strategy for an alphanumerical dataset including letters a-z, A-Z and numbers 0-9[4]. However, they used a two level classification method using Nested MLP. For the first level classification, instead of having 62 class labels; they divided the entire set into 15 similar groups (classes) using a Fuzzy C-Means (FCM) clustering method. After completion of the group level classification, each group was classified for individual characters. The results indicated that, accuracy for classifying into correct groups was 88.5 % and into correct character classes was 64.6%. Additionally, recognition of the upper case characters yield the highest accuracy by 84.38% followed by digits with 78.92% and lower case characters with 76.43%. Another handwritten character dataset EMNIST was published in 2017 by Cohen et. al. as an extension of its previous versions NIST and MNIST [3], [5]–[7]. The EMNIST dataset contains 814255 samples of letters and digits. In addition to NIST and MNIST, EMNIST not only provides two class hierarchies namely By Class (every character into a different class with a different label) and By Merge (similar characters into the same class with the same label) but also provide four more options namely: balanced dataset which is easy to apply due to its balanced subset of all the By Merge classes; letters dataset generated to increase the number of errors occurring from case confusion by merging all the uppercase and lowercase classes to form a balanced 26-class classification task; digits dataset being a balanced subset of the digits dataset containing 28,000 samples of each digit and a copy of MNIST dataset. In their study they used a three layer Extreme Learning Machine (ELM) network and Online Pseudo-Inverse Update Method (OPIUM). For recognition purposes, increasing the size of hidden layers in the network gave higher accuracy naturally requiring higher memory. The results presented that the Digits dataset gave higher accuracy rate compare to other sets including letters. Amongst the ones including letters, Balanced dataset gave the highest accuracy followed by By Merge and By Class classes.

III. THE DATA SET

In this work, we adopted C-Cube (Cursive Character Challenge) cursive character database [8]. The C-Cube dataset includes 57293 characters including 26 upper and 26 lower case versions of each Latin letter. In order to collect the characters, several postal Plants in the United States were used. This way the images used in the data set two different type of resolution 212 DPI and 300DPI depending on the source plant. Additionally for the same reason, the dataset is not a balanced dataset which means the number of each character are not the same. Some letters can be found in a larger quantity than other characters depending on how frequently they were used in the original papers. In addition to characters, several features for each character are also provided by the dataset. However, applying a deep learning algorithm (Convolutional Neural Networks) in the study we will not be using the features provided. Finally, the dataset is represented into training and test set containing respectively 38160 and 19133 characters.

A. Modification of the Data Set

The dataset is available in ".chr" file format. In the file, five integer numbers corresponding to the five features namely width, height, distance baseline-upperline, position upper extreme and position lower extreme can be found for each character [8]. After the five above integer numbers, the bitmap representation of the characters can be found ("0" referring to white pixels and "1" referring to black). And finally after each bitmap representation, class label is written in the next row. A sample character with its features and class label can be seen in Fig. 1.

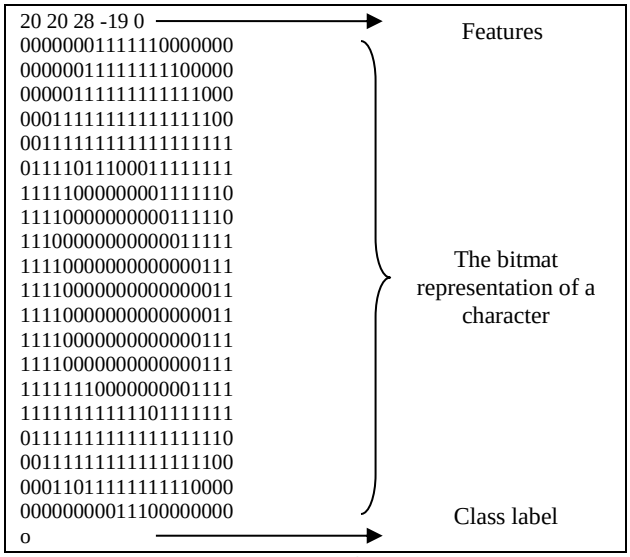


Figure 1. A sample letter from the dataset

The characters are written one after another without blank spaces between one character and the following one therefore it is not easy to use the dataset in the given format. The characters, binary numbers and integers are saved into the same file. Another challenge is that the characters are represented in different sizes, in other words they are not normalized into a standard size such as 28x28 pixels each. In order to be able to use the characters as an input to CNNs, we manipulated the characters given in the dataset. The steps followed are described below. Firstly, we copy the ".chr" file format into the MATLAB environment. It imports the data into a cell array and saves every row into one cell. Therefore we start the process with splitting the separate characters into separate cells. Since the file not only contains binary characters but also the class label and 5 features represented with numbers, the splitting process is more complex. Splitting the features is straightforward since they are separated with empty space. Secondly, we split the cells containing binary values representing each character in binary. For such cells, we split bit by bit, having one single bit in every cell. Lastly, we split the class label into a cell as a character. Having all the information in one cell array is not practical for classification therefore later we split the cell array into two separate matrices. The first matrix (1xN, N= number of characters) consists of the class label and the second one (Nx784) consists of the character representation for every character normalized into a 28x28 pixels and saved as 1x784 row for each character.

IV. EXPERIMENTS

Having all the letters saved into a matrix, we created three versions of the set with different class labels which are case-sensitive (52 classes), case-insensitive (26 classes) and similarity based (38 classes) classes as can be seen from Table 1. Merging upper and lower case of the characters based on their similarity was manually performed by looking at the shapes of the characters. However in the future, we plan to adopt a clustering method for the detection of similar classes. As for classification a two-later feed-forward network adopted. The experiments were carried out using MATLAB 9.3 Environment and Deep Learning Toolbox provided by MATLAB [9]. MATLAB is a programming platform which can be used in many fields such as control systems, image and video processing, medicine and finance[10]–[12].

Table 1 Similarity Based Merged Classes

	Merged Classes		**Merged Classes**
1	c- C	8	s-S
2	i-I	9	u-U
3	j-J	10	v-V
4	k-K	11	w-W
5	m-M	12	x-X
6	o-O	13	y-Y
7	p-P	14	z-Z

The unbalanced nature of C-Cube dataset was a disadvantage for the experiment since some letters have several samples and others only have a few.

V. RESULTS

The dataset was divided into sets namely train (70%), test (15%) and validation (15%) sets. As for the hidden layers 20 layers gave the highest accuracy for the dataset. The percentages for accuracies for 10 and 20 hidden layers are shown in the Table 2 below.

Table 2 Results

	No of Classes	**Number of Hidden Layers**	
		10	20
Case-sensitive	52	79,2 %	80,1%
Case-insensitive	26	82,5%	82,9%
Similarity Based	38	82,5%	83,1%

Looking at the results, it can be said that case insensitive classification gives better results than character sensitive classification for every single character. The main reason for such performance can be related with a smaller number of class labels and greater number of samples in each class. Having fewer number of classes from the case– sensitive and more number of classes from case–insensitive classification; similarity based classification outperforms other classifications as expected.

More detailed recognition results for case-sensitive and case-insensitive classification for 20 hidden neurons can be found in Table 3 below.

Table 3 Detailed Results for case-sensitive and case-insensitive classification

Case-sensitive				**Case– insensitive**	
Class	**Recognition Rate (%)**	**Class**	**Recognition Rate (%)**	**Class**	**Recognition Rate (%)**
a	79,1	A	79,5	A-a	80,1
b	78,8	B	80,9	B-b	80,2
c	69,3	C	72,0	C-c	75,0
d	78,1	D	80,2	D-d	81,5
e	85,0	E	84,5	E-e	85,5
f	71,5	F	72,9	F-f	80,0
g	77,5	G	80,0	G-g	82,0
h	77,3	H	77,5	H-h	80,0
i	71,5	I	70,0	I-I	72,5
j	75,0	J	74,3	J-j	77,5
k	77,5	K	75,0	K-k	80,0
l	70,1	L	76,6	L-l	78,4
m	84,5	M	85,5	M-m	87,5
n	77,8	N	80,1	N-n	82,5
o	71,5	O	74,6	O-o	75,2
p	75,0	P	80,0	P-p	87,5
q	75,0	Q	75,0	Q-q	80,0
r	75,0	R	77,5	R-r	80,0
s	82,3	S	85,5	S-s	87,5
t	80,3	T	81,3	T-t	83,0
u	78,1	U	72,3	U-u	81,3
v	70,0	V	77,8	V-v	81,5
w	83,2	W	85,0	W-w	87,5
x	85,0	X	86,1	X-x	88,0
y	75,5	Y	71,1	Y-y	81,1
z	82,5	Z	80,0	Z-z	85,3

VI. CONCLUSION

In this study, different classification strategies were applied on C-Cube handwritten character dataset in order to investigate the effects of grouping the classes. Putting upper case and lower case version of each character into one group provided higher accuracy rate compare to keeping case sensitive class labels. However, having a higher number of class labels, similarity based strategy outperformed the others by a small difference. We believe the results would have been more distinctive if the number of samples in each class was more balanced. The number of instances in each class in the dataset was highly unbalanced therefore; the results may differ in more balanced datasets.

VII. FUTURE WORK

A more sophisticated method such as clustering for finding metaclass labels rather than putting similar characters into same classes manually could improve the performance of the merging. The next step of the study is going to be application of such automated grouping for similar classes. Additionally, application of the same strategies on EMNIST dataset is planned to be performed.

REFERENCES

[1] G. E. Bartos, É. Hajnal, and Y. Hoşcan, "Comparison of Feature Extraction Techniques for Handwriting Recognition," in *IEEE 12th International Symposium on Applied Computational Intelligence and Informatics (SACI 2018)*, Temesvár: IEEE Hungary Section; IEEE Romania Section, 2018, pp. 405-410 PG–6.

[2] A. L. Koerich, "Unconstrained Handwritten Character Recognition Using Different Classification Strategies," 2003.

[3] P. J. Grohter, "NIST Special Database 19 Handprinted Forms and Characters Database," 1995.

[4] M. D. Sulistiyo, D. Saepudin, and Adiwijaya, "Optical character recognition using modified direction feature and nested multi layer perceptrons network," *Proceeding - 2012 IEEE Int. Conf. Comput. Intell. Cybern. Cybern. 2012*, no. October 2014, pp. 30–34, 2012.

[5] G. Cohen, S. Afshar, J. Tapson, and van A. Schalik, "EMNIST : an extension of MNIST to handwritten letters," 2017.

[6] P. J. Grother and K. K. Hanaoka, "NIST Special Database 19," pp. 1–30, 2016.

[7] M. Thoma, "The HASYv2 dataset," pp. 1–8, 2017.

[8] F. Camastra, M. Spinetti, and A. Vinciarelli, "Offline Cursive Character Challenge : a New Benchmark for Machine Learning and Pattern Recognition Algorithms .," *18th Int. Conf. Pattern Recognit.*, vol. 2, pp. 913–916, 2006.

[9] "MATLAB and Statistics Toolbox Release 2017b." The MathWorks, Inc., Natick, Massachusetts, United States.
[10] M. Kalechman, *Practical MATLAB Applications for Engineers*. Boca Raton: CRC Press, 2009.
[11] O. Demirkaya, M. Asyali, and P. Sahoo, *Image Processing with MATLAB*. Boca Raton: CRC Press, 2009.
[12] K. Széll and G. Manhertz, "Development of a web-based training module in robotics," in *2016 International Symposium on Small-scale Intelligent Manufacturing Systems (SIMS)*, 2016.

Consistency test of surface deformation determination using Sentinel-1A and Sentinel-1B images for the 2018 Oaxaca earthquake

G. László*, L. Földváry*,**

* Institute of Geoinformatics, Alba Regia Technical Faculty, Óbuda University, Pirosalma u 1-3, 8000 Székesfehérvár, Hungary

** Geodetic and Geophysical Institute, Research Centre of Astronomical and Earth Sciences, Hungarian Academy of Science, Csatkai Endre u. 6-8, 9400 Sopron, Hungary

laszlo.gergely@amk.uni-obuda.hu, foldvary.lorant@amk.uni-obuda.hu

***Abstract*—By the launch of Sentinel-1B, a new satellite has become available for InSAR determination of surface deformation apart from Sentinel-1A. The two missions are essentially identical, which were planned to reduce the repeat time of imaging. Even so, it is important to analyze the consistency of the data each satellite can provide. In this study, the 2018 Oaxaca earthquake has been processed using images of these satellites independently, and based on that a comparison of their solution has been provided.**

I. INTRODUCTION

The basic concept of the dedicated joint Copernicus program of European Space Agency (ESA) and European Union (EU) is to provide continuous monitoring programmes over timescales of decades [1]. The Sentinel missions of the Copernicus program are primarily designed for routine observations of operational GMES (Global Monitoring and Environmental Security) services [2]. Among the 6 Sentinel mission, Sentinel-1 satellites are providing radar imaging for land and ocean services. Accordingly, Sentinel-1 images enable determination of surface deformation over large region even with some cm amplitude over short and long time scales as well.

Volcanic and seismic events are often accompanied by measurable surface deformations over large area within the time frame of some days. As it was already shown by [3], pre-, co- and post-seismic deformations can efficiently be determined depending only on the timing of the images with respect to the event. The timing of the images is,

Figure 1. Shake map (intensity) of the Oaxaca 2018 event [6]

however, haphazard, cannot be planned or controlled. Thus, the densification of similar images by the extension of the Sentinel-1 satellite mission with additional satellites is highly justified.

Meanwhile, the extension of mission with new satellites is valuable only if the different satellites provides equivalently useful result with similar accuracy at the same spatial resolution. As in 2018, Sentinel-1 mission contains two satellites: 1A and 1B [2], while two more satellites, 1C and 1D are already contracted for an unknown date of launch [4]. The repeat cycle of a Sentinel-1A satellite is 12 days, which has already been reduced to 6 days by the launch of Sentinel-1B satellite, which is an identical twin of Sentinel-1A with 6 days delay in its orbit. As demonstrated by [5], successful combination of Sentinel-1A and Sentinel-1B images can be achieved.

In this paper now we investigate whether the results achieved by of Sentinel-1A and Sentinel-1B images are consistent. It is tested by independent processing of the same event. Therefore, a suitable event was chosen for the analysis - an event generating notable magnitude of deformation over a very short time (i.e. some days). The short duration of the deformation is demanded to be able to find useful data by both satellites.

II. The Test Event

The test event is the deformation caused by the earthquake at 23:39:39 UTC of 16 February, 2018 in Oaxaca, Mexico (location: 16.386N, 97.979W) with a magnitude of 7.2 Mww, which has resulted in an intense shake over a huge region of South Mexico, c.f. Fig. 1 [6].

According to [6] the event is a shallow earthquake (depth of the hypocenter is 22.0 km), thus we may expect visible deformation over a finite region from Sentinel images. The focal mechanism shows a normal slip with a dip angle of the nodal planes of 12° (NP1) and 78° (NP2), c.f. the beach ball model of the event on Fig 2. The direction of the slip is completely downward (rake of NP1 is 91°, rake of NP2 is 90°), along the Middle America Trench, where the strike angle of NP1 is 297°, and of NP2 116° [6].

This event can be easily separated from others: there have been no relevant pre-shocks (the largest one has occurred 2 days before the main event with an M4.4 magnitude). The post shocks contain two notable events: there has been a quake approximately an hour after the main shock (at 00:36:52 UTC of 17 February with a magnitude of 5.8 Mww and depth of 3.5 km), and another one two days later (at 06:56:58 UTC of 19 February with a magnitude of 5.9 Mww). Even though the effect of these post-seismic events on surface displacement cannot be separated from the main event, they are negligible, as the seismic moment of the main event has is $7.996*10^{19}$ Nm, while for the post shocks is only $5.817*10^{17}$ Nm and $8.329*10^{17}$, respectively for 17 and 19 February events. It means that due to the post shocks the seismic moment (practically related to ground motions) became less than 2% of that of the main shock, so surface deformations are dominated by the main shock. The same can be understood when the released energy of the main shock and of the two largest post shocks are compared: the energy released by these events are $3.98*10^{15}$ J, $3.16*10^{13}$ J and $4.47*10^{13}$ J, respectively, so the energy released by the post-shocks is less than 2% of that of the main shock.

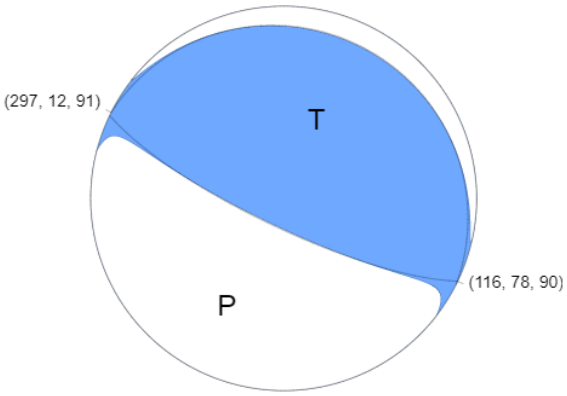


Figure 2. The beach ball model of the Oaxaca 2018 event [6]

Note finally, that this event has already been processed by [7] using only the Sentinel-1B satellite data. The deformation rings were observed to be essentially clear, confirming the choice of the event to be appropriate for overall analysis.

III. Data Processing

In order to analyze consistency of Sentinel-1A and Sentinel-1B, the same processing sequence has been performed for both satellite's images independently. First, images from the Copernicus Open Access Hub has been downloaded [8]. From the Hub images on 27 January and 20 February for Sentinel-1A, and on 5 February and 17 February for Sentinel-1B has been found accessible. Even though the time span is notably different, presumably differences in the deformation may be regarded mainly to the pre-seismic deformations, in which case the difference of starting at 27 January or at 5 February may be relevant, and partially may be regarded to the post-seismic differences, as the effect of the 19 February post-shock event is excluded from the Sentinel-1B, but included in the Sentinel-1A image.

A relevant difference of the acquired images is that in the case of Sentinal-1A images, these were taken along a descending orbit, while in the case of the Sentinel-1B images, it was an ascending orbit.

Using these images, processing was performed by using the Sentinel-1 Toolbox of the Sentinel Application Platform (SNAP), an open-source software provided by the ESA, jointly developed by Brockmann Consult, Array Systems Computing and C-S for processing images of orbiting Sentinel satellites [9]. The processing (including co-registration of the images, formation of the interferograms, deburst of the images, removal of the topographic phase, the Goldstein phase filtering, phase unwrapping, determination of the phase displacement, georeferencing and removal of the ellipsoidal correction) is done routinely, the sequence is detailed in [10].

The resulted phase maps are shown on Fig. 3 and Fig. 4. The structure of the deformation is visually similar, a difference in coverage of the test area can be observed due to the different areal coverage of taking the different images.

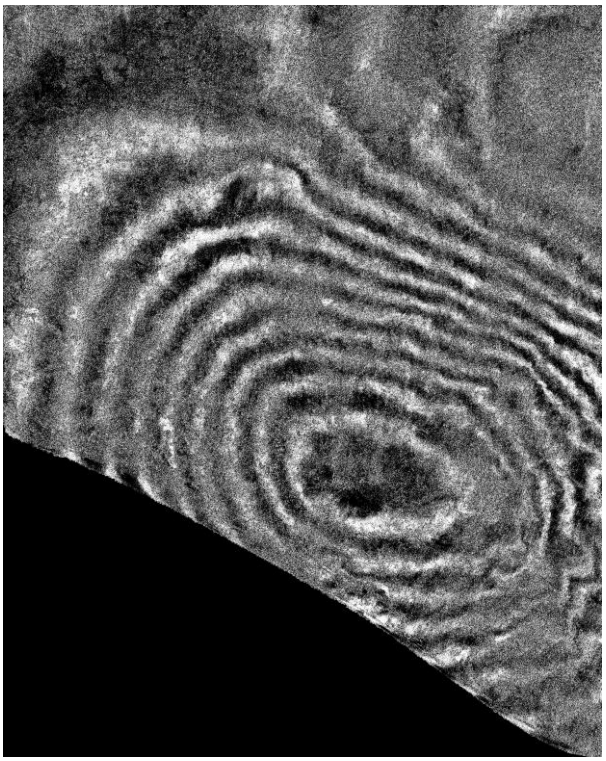


Figure 3. The phase map of the Oaxaca 2018 event based on observations of the Sentinel-1A satellite

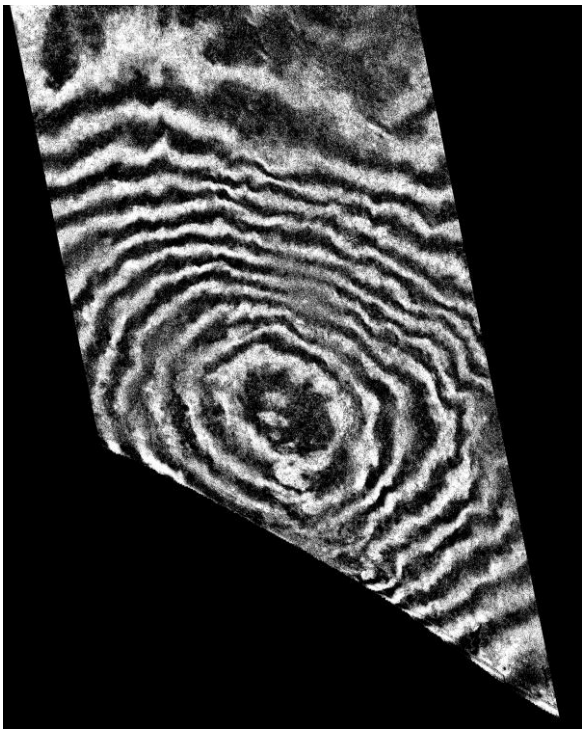


Figure 4. The phase map of the Oaxaca 2018 event based on observations of the Sentinel-1B satellite

IV. COMPARISON OF THE RESULTS

For numerical comparisons the phase maps shown on Fig. 3 and Fig. 4 should be georeferences. Fig 5., Fig 6 and Fig. 7 shows the georeferenced Sentinel-1A, Sentinel-1B and their difference phase maps, respectively.

According to this images it is obvious that there is a displacement of the center of the location of the maximal deformation. It may arise from the different circumstances of taking the images, i.e. taken along an ascending or a descending orbit.

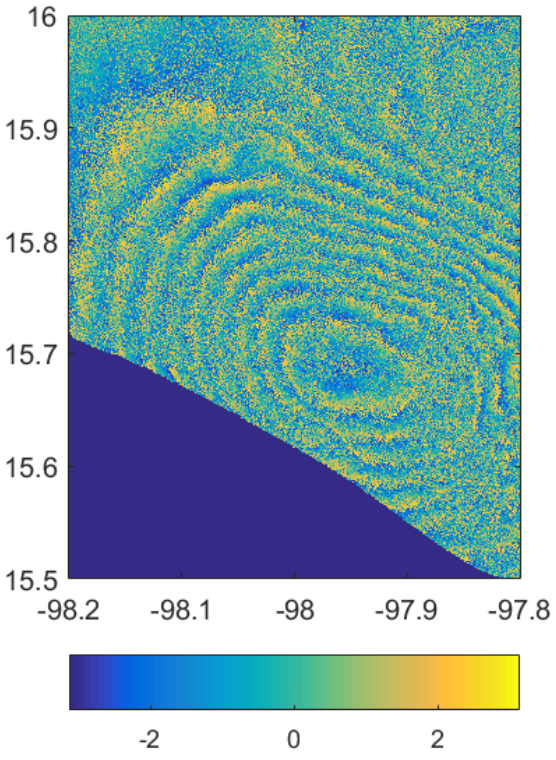


Figure 5. Georeferenced Sentinel-1A phase map

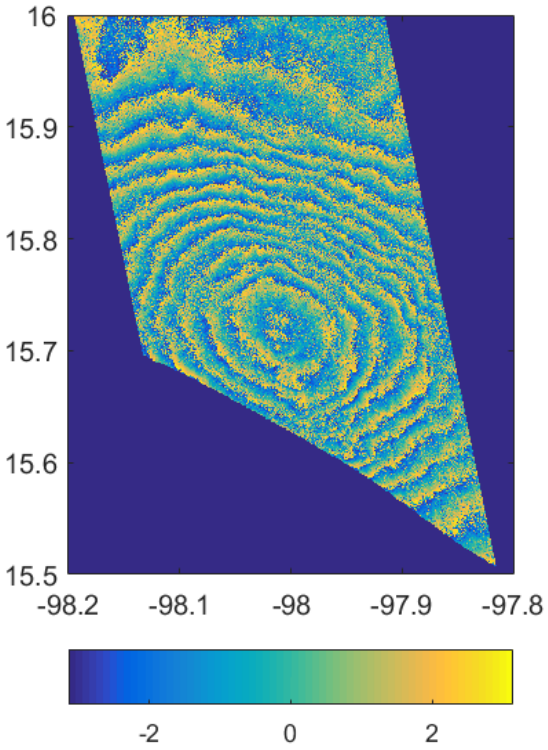


Figure 6. Georeferenced Sentinel-1B phase map

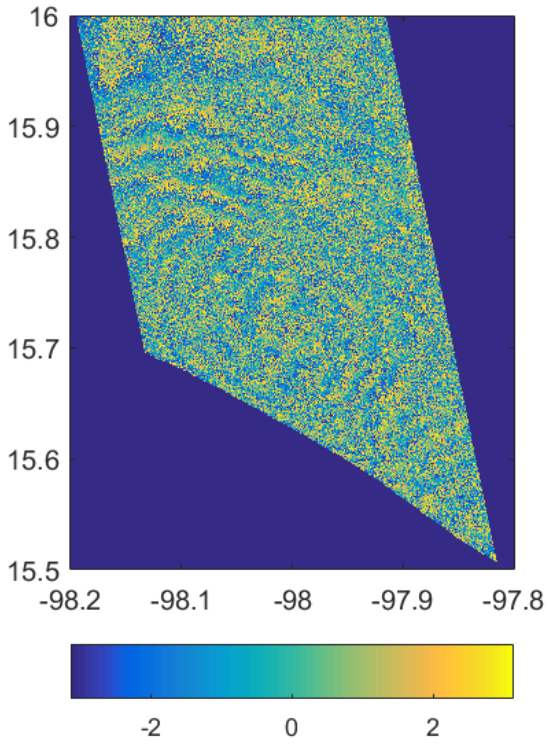


Figure 7. The difference phase map

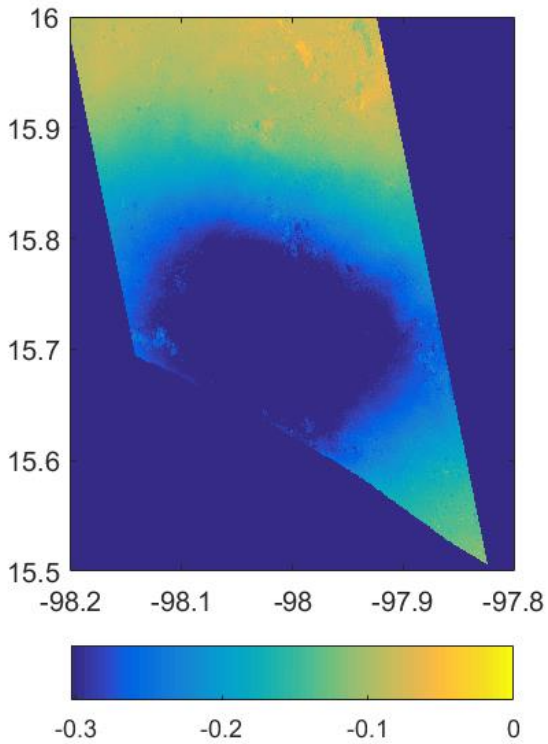


Figure 8. Deformation derived from Sentinel-1A images. Unit: [m]

The corresponding deformations have been determined by the Snaphu software [9]; these are displayed by Fig. 8 and Fig. 9, and their difference is on Fig 10.

According to the figures, no values below a threshold, in the present case below -30.34 cm, was observed, but all values stored to be the threshold value. Definitely, the deformation cannot be considered to be flat and fully horizontal over a huge region, therefore this is obviously a bug of the software.

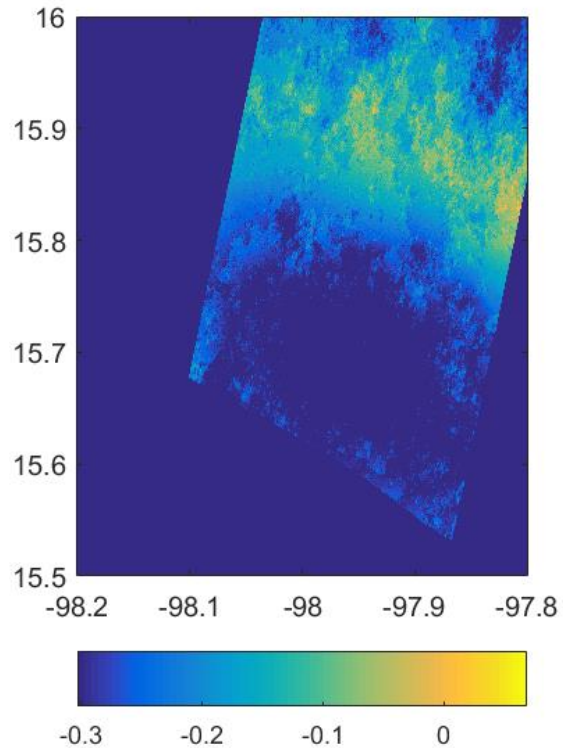


Figure 9. Deformation derived from Sentinel-1B images. Unit: [m]

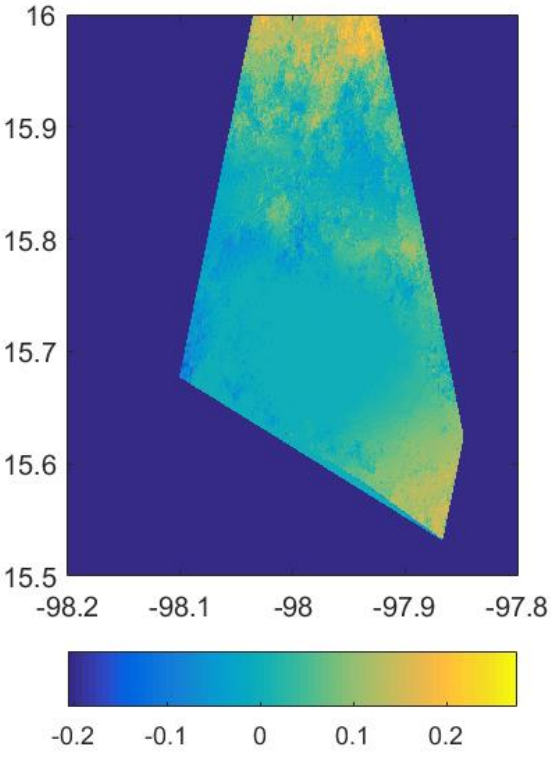


Figure 10. Deformation differences. Unit: [m]

Still the results are interesting. Apart from the threshold truncation problem, there are relevant features can be detected. First of all, the shape of the deformation seems to depend on direction of the beamline of acquiring the image as the tilting of the displaced area is in accordance with that of the imaging beamline, c.f. Fig. 8 and Fig. 9.

Also, it is obvious that close to the edge of an image that acquired data is less reliable, it results in a relevant difference at the Northern edge of the ascending and descending images. It may be a consequence of the different direction of the beamline, as it may reflect that in one case the beamline could reach the ground from one direction, while it is shadowed from the other.

V. Conclusions

A statistical comparison of each cases are also derived. The deformation was observed to occur in the [-30.34 cm, +0.01 cm] and in the [-30.34 cm, +0.07 cm] intervals for the Sentinel-1A and Sentinel-1B cases, respectively. The maximal sinking is a bug; in reality the deformation for sure extend this value. The range, however, is similar for both cases, as the bug influenced region has a similar area.

Fig. 10 shows the difference deformation map. It is dominated by the Northern edge difference. Including this area as well, the residual RMS was found to be 5.05 cm, showing that the two solutions are similar to each other in the cm level. It may not be considered to be an adequate validation as a notable area of the test is infected by software bug, still it is a notation that the satellites can contribute identical results to surface deformation analysis in the cm level.

Fig. 11 and Fig 12. Present two different cross-sections of the deformation map. These cross-sections indicate the nonsense characteristics of the maximal deformation as both data suggest a flat and horizontal bottom of the deformed area, while no indication of such flatness in the phase maps, c.f. Fig. 3 to Fig. 6 can be observed. The cross-sections, on the other hand shows that the slope of the deformed area is similar from the two independent sources, which is a notable indication that the satellites observes some real change with a similar accuracy. This similarity is distorted by approaching to the Northern edge of the test area, as it has already been mentioned before.

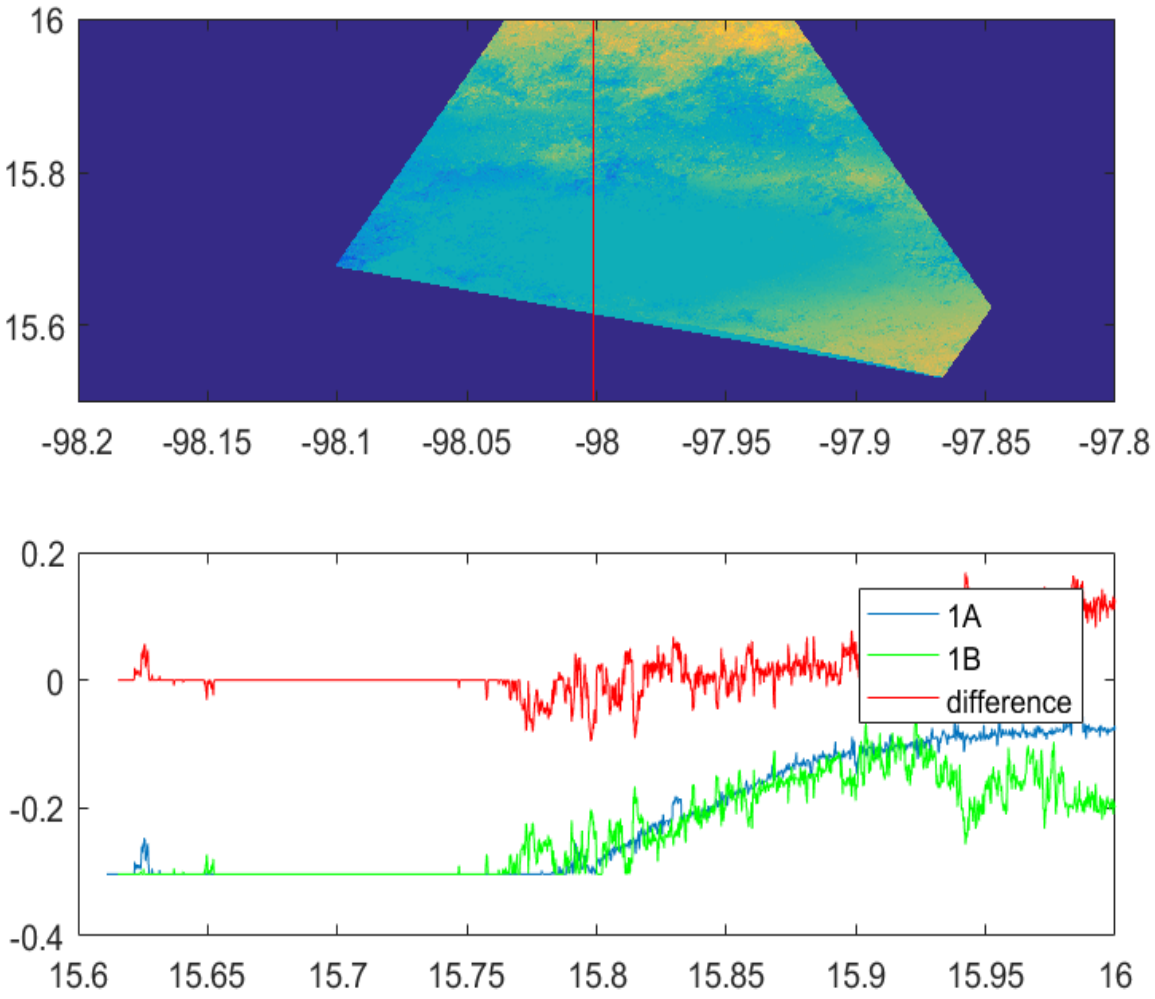


Figure 11. Cross section at E98-00-04 longitude.

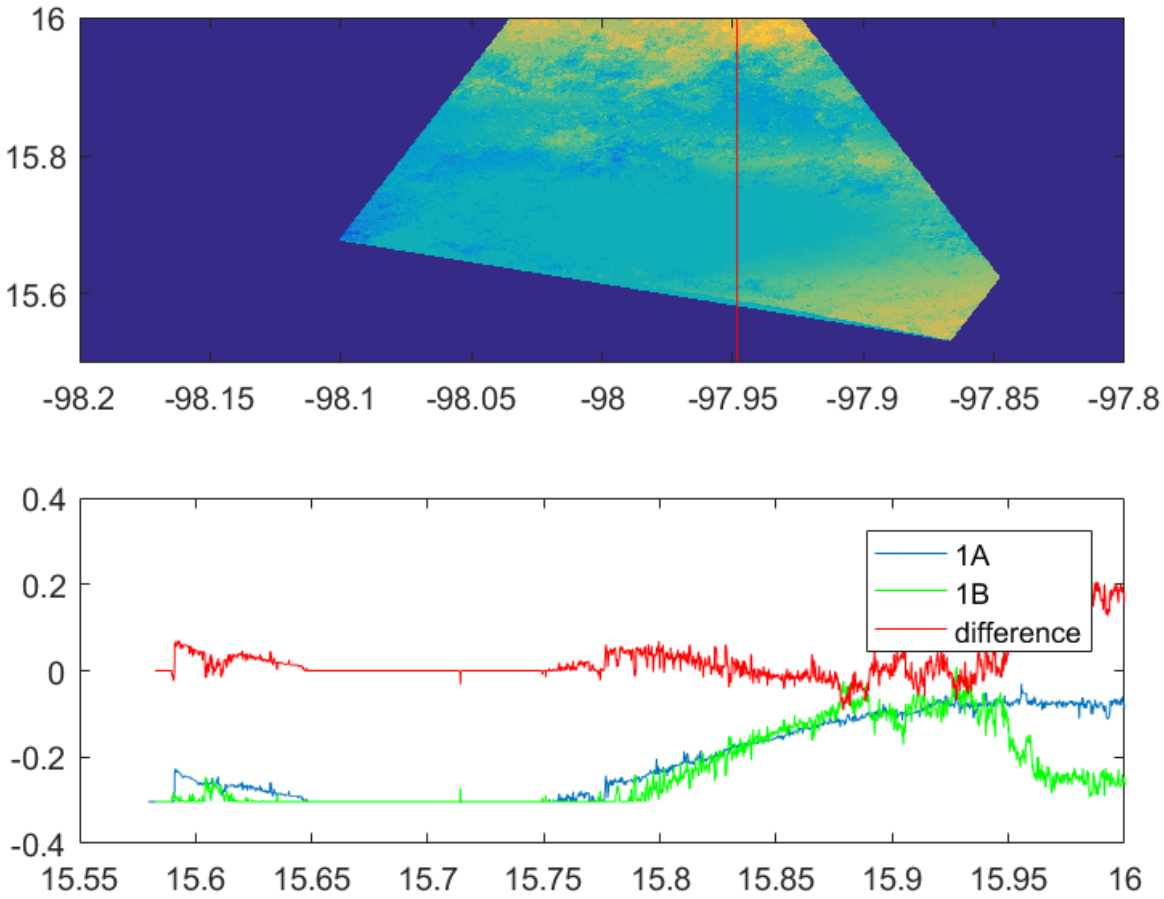


Figure 12. Cross section at E97-56-54 longitude.

The strength of the present analysis that it has dealt with images taken by two different satellites, along two differently oriented orbits, still the similarity at cm level could be detected. Definitely, further analyses may take place making also use of combinations of identical satellite-borne images along two different orientations, and also different satellite-borne images taken from identical orientations.

.

REFERENCES

[1] M. Berger, J. Moreno, J. A. Johannessen, P. F. Levelt, R. F. Hanssen, "ESA's sentinel missions in support of Earth system science", *Remote Sensing of Environment*, 2012.

[2] Sentinel-1, *ESA's radar observatory mission for GMES operational services*, ESA Special Publication (SP-1322/1), 2012.

[3] G. László,, L., Földváry: "Application of Sentinel-1 data for determination of surface deformation due to volcanic and seismic events", 17th edition International Technical-Scientific Conference Modern Technologies for the 3rd Millennium, 55-60, ISBN 978-88-87729-49-8.

[4] Ch. Rossi, "Decolla La Space Economy Italiana", *AirPress*, 15.12.2015, retrieved on 06.09.2018, link: http://www.airpressonline.it/8562/decolla-la-space-economy-italiana/

[5] G. László, L. Földváry, "Combined use of Sentinel-1A and Sentinel-1B data for determination of post-seismic vertical surface deformation - a case study: Perugia, 2016", *Proceedings of the 12th International Symposium on Applied Informatics and Related Areas (AIS 2017)*, Székesfehérvár November 9, 2017, pp. 61-65, ISBN: 978-963-449-032-6, 2017.

[6] USGS, "M 7.2 - 3km S of San Pedro Jicayan, Mexico earthquake information", *USGS homepage*, retrieved on 06.09.2018, link: https://earthquake.usgs.gov/earthquakes/eventpage/us2000d3km#executive

[7] S. Hábel: *InSAR adatfeldolgozások paramétereinek vizsgálata és optimalizálása*, BSc thesis, Óbuda University, pp. 52, 2018.

[8] Sentinel Data Hub, *Copernicus Open Access Hub*, retrieved 06.09.2018, link: https://scihub.copernicus.eu/dhus/#/home

[9] Sentinel-1 Toolbox of the Sentinel Application Platform (SNAP), *Science Toolbox Exploitation Platform (STEP)*; retrieved on 06.09.2018, link: http://step.esa.int/main/toolboxes/snap/

[10] G. László, L. Földváry, "Determination of the surface deformation during the eruption of the Sakurajima volcano in February, 2016", *Proceedings of the 11th International Symposium on Applied Informatics and Related Areas (AIS 2016)*, Székesfehérvár November 11, 2016, pp. 74-78, ISBN: 978-615-5460-92-0, 2016

The applicability of alternative point-targets for laserscanning of underground and surface formations

B. Petrétei

Institute of Geoinformatics, Alba Regia Technical Faculty, Óbuda University, Pirosalma u 1-3, 8000 Székesfehérvár, Hungary

Abstract - **In this paper a research of finding the best point type, replaces the professional Leica HDS, without losing accuracy, and enables automating possibilities of measurement and processing process, during terran and and subterranean laser scanning of natural formations like rock outcrops and caves will be presented. The research started 3 years ago with many possible target-candidate and by the time the list narrowed to the two best in practise options and proven to be almost as accurate as the professional targets but for fractions of costs.**

I. INTRODUCTION

In my research I studied that during the laser-scanning measure of caves, in addition to the four Leica HDS (2 pcs 3 "and 2 pcs 6") of the Geoinformatics Institute, what alternative, point clouds connection capable targets can be used, and how accurate they are compared to the solution of the Leica factory-solutions, and how it is possible to automate the processing of measurements at alternative point marks. [4.]

The test is useful because, in the measurement of caves, especially if they have multiple sidebranches, the four targets are often prove to be insufficient, but without easily identifiable discrete dots, aligning the point clouds manually can be a very cumbersome task.

During the measurement, the Leica HDS target used as a reference for the study and were placed with other discreet objects such as:

- printed HDS Black & White Targets,
- Table-tennis balls,
- Golfballs
- Tennis balls
- Sponge balls and
- Styrofoam spheres.

Figure 1. Target group

I applied anchorages that take into account the particularities of the site, to ensure that the targets were stationary.

The site I have choose for the first measurement was the Hospital-cave, part of the Tapolca-cave system in the town of Tapolca. The reason of this choice because it can be easily approached by walk, has a very high ceilings, so it was easy to move with the instrument and equipment, but I had also the opportunity to investigate how my methods work in narrow passages. The measurement took place on 2016.10.14.

The second phase of my research was in the Csákvár cave, which can be located in the western part of the Csákvár. Here, during the measurement I have already tried to focus on those point targets which based on last year's measuring achieved good results, and in addition of those I used a new target type, the geodetic reflector stickers, available at the

University, which I examined also, in laboratory measurements.
The third stage of my research site was a rocking stone, called "Woolenoch" near Sukoró. In the nature reserve of Velence Hills, one of the most unique sights. The location was easy to reach, and it was possible to go by car to the rocks to be measured. Here I only used the target types gave - based on my previous measurements the best results (ping-pong balls), and as a new point type the Styrofoam globes, which I choose its diameters to have almost identical of the Leica HDS Sphere, but for obvious reasons it has the advantage for being much less expensive.
The orientation of this research phase were also broken down into two measurements.

II. PHASE ONE

First, there was a laboratory measurement in which I examined the usability and fitting accuracy of the Styrofoam sphere, and a field measurement where the new point type was tested, and the results were examined and compared against the ping pong balls and again in reference the HDS Targets .
To process the measured data, in all cases Leica Cyclone software were used where I have completed the various types of points for manual point-cloud joins and in the light of the results I made a proposal to use the most optimal point type for the laser-scanner measurement of caves.

The first results During the research Developed as follows:

The table shows that after the HDS targets, the ping pong balls and the sponge balls gave the best values. Of the two, however, the ping pong balls could be given minimal value in three categories, as compared to the two of the sponge ball, as there was a rough defect in the case of the ping-pong balls, so that I could not make it clear that it would be a better choice.

Aggregation of results						
	Ping pong	Bounce	Tennis Ball	Sponge Ball	Sheet of paper	HDS Target
Expected value	0.0018	0.0021	0.00200	0.0018	0.0031	0.0014
Mode	0.001	0.002	0.002	0.002	0.003	0.001
Minimum	0	0.001	0	0.001	0.001	0
Maximum	0.003	0.004	0.004	0.003	0.008	0.004
Number used	52	44	37	43	36	25

Table 1. Aggregated results of the first part of the research

Based on my results I couldn't give a definite answer which alternative point marks are the best choice in each field, so the research has continued. [1.]

III. PHASE TWO

In addition to the spring measurement in the second part of the research, apart of the ping pong balls and A5-size printed targets I used earlier I tried three new point signal type out

- White and yellow geodetic reflector signal
- A4-Sized printed mark
- Styrofoam Sphere

I have decided to test the advantages of using Reflectors as well-discernible high reflecciating signal in a point cloud, and study what accuracy can be achieved by this target type.

The A4 page was tested answer to which extent the signal is affected by the size of the targets the identification in automatic, semi-automatic and manual mode.
Finally, in addition I tested the Styrofoam sphere, for its size cheapness.

In the calculation of the fair resolution in the ground level, I created a program in Excel, in which by knowing the resolution (f), the resolution distance (t) and the instrument height (h) I determined with the plane of the position at a height at a given distance's (d) mid-point resolution (Fd).

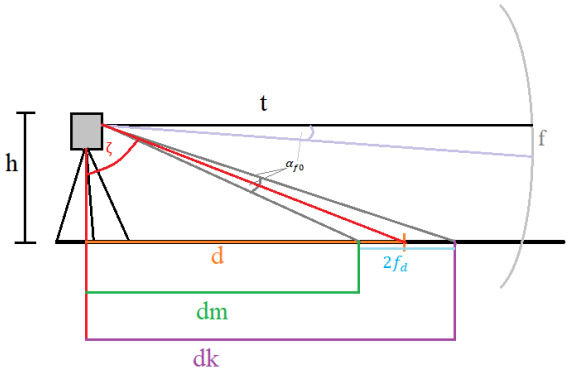


Figure 2.The geometrical representation of the calculation

The calculation developed as follows:

1. The laser head stepping angle (α_{F0}) is calculated at the resolution distance (t_{My}) and Resolution (f_{cm}) using

$$\alpha_{f0} = \frac{360\frac{f_{(cm)}}{100}}{2t_{(m)}\pi}$$

2. Then I count on the basis of the instrument height (h) and signal distance (d) (ζ):

$$\zeta = \tan^{-1}\frac{d}{\frac{h}{100}}$$

3. The third step is to calculate how many steps of the instrument should start from nadir direction, as defined in point α_{F0} to the respective ζ. Angle.

$$f_i \approx \frac{\zeta}{\alpha_{f0}}$$

4. Thereafter, it shall be calculated that previous $f_i - 1$ and the next $f_i + 1$ The Angles are:

$$\alpha_{fi-1} = \alpha_{f0} \cdot (f_i - 1)\ and\ \alpha_{fi+1} = \alpha_{f0} \cdot (f_i + 1)$$

5. Finally, the resolution (f_{ncm}) calculated on the basis of previous and subsequent stepping angles,(d_m) and (d_k).

$$f_{d(cm)} = \frac{d_k - d_m}{2}$$

Where:

$$d_m = \tan\alpha_{fi-1} \cdot \frac{h}{100}\ illetve\ d_k = \tan\alpha_{fi+1} \cdot \frac{h}{100}$$

The program's simple controlThe following illustration shows the stake:

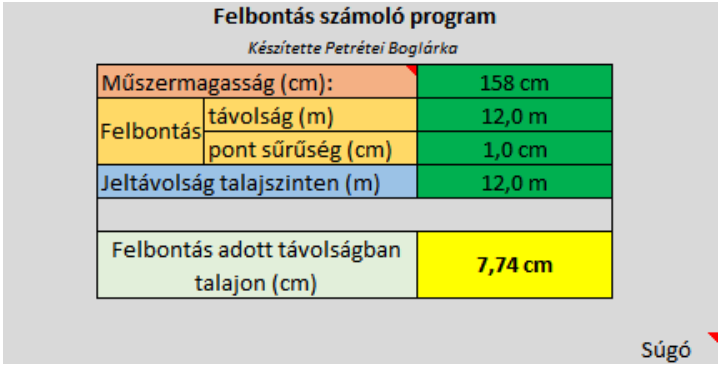


Figure 3. User interface of the resolution calculator

Overall, measurements and processing were successful. I received important results. It is important to create documentation of the placement of the ping pong ball for easy processing, because without it there have been some identification problems that could easily be repaired, but exploring is time consuming if there are multiple signs near each other.

Using paper targets may be justified, because they are partially auto-recognized (at a certain resolution, they are already identified during an import), there is a way to search semi-automatic (when the program searches and selects them automatically in the surrounding of a specific point) and manual selection, if seen on the point cloud identifying its center point is simpler than in the case of using the spheres. Problem alone was to place the marks in a location that can be seen on multiple positions. For this reason, it is usually only used for a maximum of 2 adjacent positions, as the program I developed demonstrates that the effective resolution at long distances for plain targets can be very low.

The most difficult to consider in terms of research is the perception of the reflectors. Due to its good reflective properties, they can be easily identified in the point cloud visually, but there are some situations where they cannot be seen, in which we haven't found explanations for the by now. Conversion to point target is also cumbersome because the instrument directional noise is significant for reflector stickers, the overall use of them is more work-intensive in the point of processing than the paper sheet, because of its strange reflective behavior, is unreliable. The use of this form only redundantly recommended, and their price is also not makes them ideal.

	Ping pong	*Reflector*	*Sheet of paper*	*HDS Target*
Expected value	0.006	0.017	0.002	0.001333333
Mode	0.005	0.017	0.002	0.002
Minimum	0.002	0.003	0.001	0
Maximum	0.016	0.036	0.003	0.003
Number used	11	8	4	9

Minimum Reference

Table 2. Aggregated results of the second part of the research

From the table it can be read that the previously not-so-good sheet of paper jumped to the first position in all four tested categories, almost reaching the HDS Targets. The formal favorite ping pong ball with a larger error has produced a slightly weaker accuracy than before. The reflector can only be reached by a minimum of 3 mm in the exact same category compared to others.

Thus, in the minds of the results, further tests were needed, in order to find the most appropriate score. [2.]

IV. PHASE THREE

In the third part of the research, the objective pursued was double. First, consider whether is with similar than Leica HDS Sphere spherical diameters it is possible to perform this automatic measurement with styrofoam spheres and, if so, what accuracy.

The other objective was to try to automate the processing process, accelerate by own algorithm. This time, the area tested was on the surface and not in cave, it was a rocking stone, thus proving the universal applicability of the procedure.

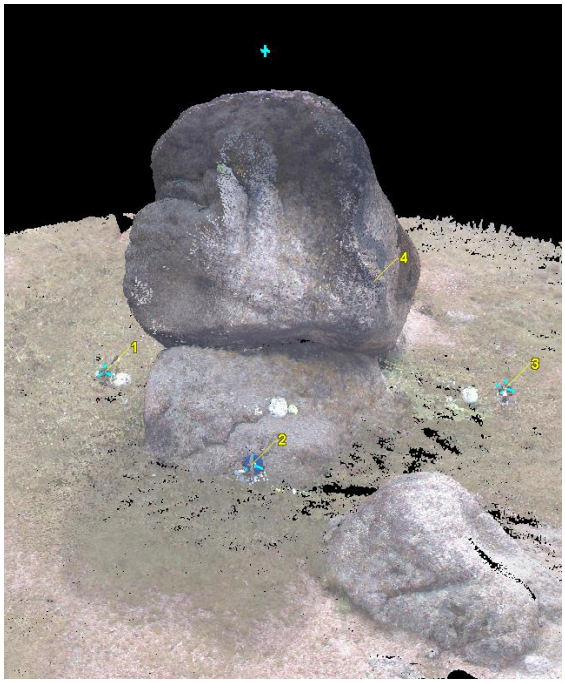


Figure 4. Pointcloud image of the third measurement

The first result for the question was that, at the measurement of target, even a few millimeters difference in relation to the standard diameter compared to HDS spheres can lead to significant errors.

	S (152mm)						G (145mm)					
#	x	y	z	Horz	Vert	Error	x	y	z	Horz	Vert	Error
12	0,009	0,000	-0,001	0,009	0,000	0,009	0,000	0,000	-0,003	0,000	0,000	0,003
9	0,010	0,000	-0,001	0,010	0,000	0,010	0,000	0,001	-0,001	0,001	0,001	0,001
6	0,009	0,000	0,000	0,009	0,000	0,009	0,001	0,001	-0,002	0,001	0,001	0,002
3	0,009	0,001	0,001	0,009	0,001	0,009	-0,001	0,001	0,003	0,001	0,001	0,003

Table 3. Results of processing spheres with different diameter

Apart from this, the spheres can be used well with a post-processing method, although the best results are achieved not the a sphere with the nominal radius, but with 145 mm which is less by 4mm. The reason for this could be the pole-orientation flattening of the sphere, which we were unable to justify, but we later plan to investigate with another instrument.

The second goal was realized with MATLAB, wherein to a Ply format exported point cloud were used to select the approximate location of the spheres, to be used to get a spatial midpoint for a join. The program is prepared for rough defective cases (when sphere is greater than a given value) or to write the results directly to a result file. [3.]

V. CONCLUSIONS

So comparing the above results I took a big step closer to be able to perform high precision laser-scanner measurement and processing not only substituting the targets but to provide an alternative for the processing software. [3.]

REFERENCES

[1.] B. Petrétei: Alternatív pontjelek pontossági vizsgálata barlangok lézerszkenneres méréséhez, TDK dolgozat, ÓE AMK, Székesfehérvár, 2016

[2.] B. Petrétei: Különböző jellegű illesztőpontok vizsgálata barlangok felméréséhez, TDK dolgozat, ÓE AMK, Székesfehérvár, 2017

[3.]Petrétei B.: Gömb jellegű illesztőpontok lézerszkenneres feldolgozásának automatizálási lehetőségei, TDK dolgozat, ÓE AMK, Székesfehérvár, 2018

[4.]Berényi A., Lovas T., Barsi Á., Tóth Z., Rehány N., Tarsoly P. (2011): Földi lézerszkennerek minősítő vizsgálatának lehetőségei, Geomatikai Közlemények XIII/2, HU ISSN 1419-6492, MTA GGKI, Sopron, 2010, 87-95. oldal

Special Classes of Shifted k-dimensional Vectors Modulo a Prime

Borbély József, Csala-Takács Éva

Óbudai Egyetem, Székesfehérvár, Hungary

borbely.jozsef@amk.uni-obuda.hu

csala.takacs@amk.uni-obuda.hu

***Abstract*— In this paper we will discuss a problem concerning vectors with integer coordinates modulo a prime number. We will proceed from a problem that was set in the Kürschák competition in 2018, and we will give a strong generalization of it. We will also formulate some open questions.**

I. Introduction

There are a lot of standard linear algebraic questions, how many vectors one can choose in the k-dimensional plane, if one expects some prescribed properties to be fulfilled But one can also ask similar questions with some number theoretical restrictions. In this paper we will discuss the generalization of a Kürschák problem. Throughout the whole paper k will denote a positive integer, and k will denote a prime number.

II. A Kürschák problem from 2018 and its generalization

We formulate the second problem of the Kürschák competition as a theorem:

Theorem 1: *Let p be an odd prime number, and let* $\boldsymbol{v_1, v_2, \dots, v_n}$ *be distinct three-dimensional vectors with integer coordinates, whose lengths equal p. If for every* $1 \le i < j \le n$ *there exists an integer t such that* $0 < t < p$ *and all the coordinates of the vector* $\boldsymbol{v_i - t v_j}$ *are divisible by p, then n (which is the number of the vectors) is at most 6.*

We will prove a generalization of this problem in every finite dimensional case.

We formulate the generalization as a theorem:

Theorem 2: *Let p be an odd prime number, and let k be a positive integer, and let* $\boldsymbol{v_1, v_2, \dots, v_n}$ *be distinct k-dimensional vectors with integer coordinates, whose lengths equal p. If for every* $1 \le i < j \le n$ *there exists an integer t such that* $0 < t < p$ *and all the coordinates of the vector* $\boldsymbol{v_i - t v_j}$ *are divisible by p, then n (which is the number of the vectors) is at most 2k.*

Obviously, Theorem 2 is a generalization of Theorem 1. The result of Theorem 1 follows from the one of Theorem 2, if we consider the case k=3.

It is easy to see that in Theorem 2 the two parameters p and k can be chosen independently from each other.

The condition $\mathbf{v_i - t v_j}$ being divisible by p means, that the vectors $\boldsymbol{v_j}$ and $\boldsymbol{v_i}$ are in some sense multiples of each other (in the sense of calculating modulo p).

Before proving Theorem 2, we need some lemmas and some definitions.

III. The Cauchy-Schwarz inequality

Lemma 1: *Let k be a positive integer. For any k-dimensional vectors* ***u*** *and* ***v*** *holds the inequality*

$$|\boldsymbol{uv}| \le |\boldsymbol{u}||\boldsymbol{v}|$$

Lemma 1 is called the Cauchy-Schwarz inequality. In other words we can say, that for any vectors $\mathbf{u}=(u_1, u_2, \dots, u_k)$ and $\mathbf{v}=(v_1, v_2, \dots, v_k)$ with real coordinates holds the inequality

$$|u_1 v_1 + u_2 v_2 + \dots + u_k v_k| \le$$

$$\le \sqrt{u_1^2 + u_2^2 + \dots + u_k^2} \cdot \sqrt{v_1^2 + v_2^2 + \dots + v_k^2}$$

Equality occurs in the Cauchy-Schwarz inequality if and only if the vectors **u** and **v** are multiples of each other, i.e. there exists a real number r such that one of the equalities $r\mathbf{u}=\mathbf{v}$ or $r\mathbf{v}=\mathbf{u}$ holds.

Now we will define linear dependence and independence.

IV. Linear dependence and independence

One of the most important terms in linear algebra is the one of linear independence. We will define linear independence and dependence to be able to prove Theorem 2.

Definition : *Let V be a vector space over the field F. The vectors* $\boldsymbol{v_1, v_2, ..., v_k}$ *form a linearly independent system of vectors, if for the elements* $t_1, t_2, ..., t_k$ *of F the equality* $t_1\boldsymbol{v_1}+t_2\boldsymbol{v_2}+...+ t_k\boldsymbol{v_k}= \boldsymbol{0}$ *implies that* $t_1=t_2=...= t_k=0$. *If a system of vectors is not linearly independent, we call it linearly dependent.*

We will use also the following well-known theorem:

Theorem 3: If the vectors $\boldsymbol{v_1, v_2, \dots, v_n}$ with *k coordinates from a vector space over the field F are linearly independent, then* $n \leq k$.

In order to have a more simple formulation we will use the (standard) terminology that the vectors $\mathbf{u}=(u_1, u_2, ..., u_k)$ and $\mathbf{v}=(v_1, v_2, ..., v_k)$ are perpendicular if and only if

$u_1v_1+u_2v_2+...+u_kv_k = \mathbf{uv} = \mathbf{0}$.

We need another lemma:

Lemma 2: *If the nonzero vectors* $\boldsymbol{v_1, v_2, \dots, v_k}$ *of a vector space over the field F are pairwise perpendicular, then they form a linearly independent system of vectors in this space.*

Proof of Lemma 2:

Let $t_1, t_2, ..., t_k$ elements of F such that

$t_1\mathbf{v_1}+t_2\mathbf{v_2}+...+ t_k\mathbf{v_k}= \mathbf{0}$

holds. Let us multiply this equality by $\boldsymbol{v_1}$, then we get that

$t_1\mathbf{v_1}\mathbf{v_1} = 0$.

$\mathbf{v_1} \neq \mathbf{0}$, therefore $t_1 = 0$. A similar argumentation shows that

$t_1=t_2=...= t_k=0$, thus the vectors are linearly independent **QED**.

Now we are able to prove Theorem 2.

V. Proof of our main theorem

Let $\mathbf{u}=(u_1, u_2, ..., u_k)$ and $\mathbf{v}=(v_1, v_2, ..., v_k)$ be distinct k-dimensional vectors with integer coordinates and with length p. Suppose that there is an integer t such that $0<t<p$ and all the coordinates of the vector $\mathbf{u}$-t$\mathbf{v}$ are divisible by p. That means that all the numbers u_i-tv_i are divisible by p. Thus the sum

$$(u_1 - tv_1)^2+(u_2 - tv_2)^2+...+(u_k - tv_k)^2$$

is divisible by p^2.

Now, using the condition, that $|u|=p$ and $|v|=p$ we have

$$(u_1 - tv_1)^2+(u_2 - tv_2)^2+...+(u_k - tv_k)^2=$$

$$=u_1{}^2+u_2{}^2+ \dots u_k{}^2+ t^2(v_1{}^2 + v_2{}^2+ \dots v_k{}^2)-$$

$$-2t(u_1v_1 + u_2v_2+\dots+u_kv_k) =$$

$$= p^2 + t^2p^2 + -2t(u_1v_1 + u_2v_2+\dots+u_kv_k).$$

The number $p^2 + t^2p^2$ is divisible by p^2, therefore

$2t(u_1v_1 + u_2v_2+\dots+u_kv_k)$ must be divisible by p^2 as well.

The numbers 2t and p are coprimes because p is odd and $0<t<p$. That means that the number

$$u_1v_1 + u_2v_2+\dots+u_kv_k$$

is divisible by p^2.

Now we will use the Cauchy-Schwarz inequality.

We know that

$$|u_1v_1+u_2v_2+...+u_kv_k| \leq$$

$$\leq \sqrt{u_1{}^2 + u_2{}^2 + \dots + u_k{}^2} \cdot \sqrt{v_1{}^2 + v_2{}^2 + \dots + v_k{}^2}.$$

The right hand side of the inequality equals p^2, because the length of the vectors u and v is p. Thus the expression

$$u_1v_1+u_2v_2+...+u_kv_k$$

is divisible by p^2 and its absolute value cannot be greater than p^2. So we have two possibilities:

Case 1:

$|u_1v_1+u_2v_2+...+u_kv_k|=0$
Case 2:

$|u_1v_1+u_2v_2+...+u_kv_k|= p^2$.

In Case 1 the vectors u and v are perpendicular, in Case 2 there must be a real number r such that one of the equalities r**u**=**v** or r**v**=**u** holds. **u** and **v** are different vectors, their length is the same, and for some real number r one of the equalities r**u**=**v** or r**v**=**u** holds, therefore in Case 2 we get **u** = - **v**.

Thus if we have the vectors distinct $\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_n}$ k-dimensional vectors, which corespond to the conditions of Theorem 2, then for every 1≤i<j≤ one of the following conditions holds:

Condition 1: $\boldsymbol{v_i}$ and $\boldsymbol{v_j}$ are perpendicular

Condition 2: $\boldsymbol{v_i}+\boldsymbol{v_j} = \mathbf{0}$

We will show that in this case n ≤ 2k.

Let us choose a linearly independent subset of the set $\{\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_n}\}$ with the maximal number of elements, let this subset be $\{\mathbf{w_1}, \mathbf{w_2}, \dots, \mathbf{w_j}\}$**.** Due to Theorem 3 the inequality j ≤ k holds.

Let **v** be an element of the set
$\mathbf{H = \{v_1, v_2, \dots, v_n\} \setminus \{w_1, w_2, \dots, w_j\}.}$
If the vector **v** was perpendicular to all the elements in the set $\{\mathbf{w_1}, \mathbf{w_2}, \dots, \mathbf{w_j}\}$**,** then the set $\{\mathbf{w_1}, \mathbf{w_2}, \dots, \mathbf{w_j}, \mathbf{v}\}$ would be a linearly independent set of vectors, but that would contradict to the maximality of the subset
$\{\mathbf{w_1}, \mathbf{w_2}, \dots, \mathbf{w_j}\}$**.**

Therefore for every element **v** of H there is a **v*** vector in the set $\{\mathbf{w_1}, \mathbf{w_2}, \dots, \mathbf{w_j}\}$ such that **v+v*=0**. Because for every vector **v** in the whole vector space there is exactly one vector **v*** such that **v+v*=0**, this implies that

$|H| \leq |\{\mathbf{w_1}, \mathbf{w_2}, \dots, \mathbf{w_j}\}| = j \leq k.$

Thus we have

$n = |\{\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_n}\}| = |H \cup \{\mathbf{w_1}, \mathbf{w_2}, \dots, \mathbf{w_j}\}| \leq$

$\leq k + k = 2k$, as desired.

Thus we proved of the statement of Theorem 2 **QED**.

VI. SOME QUESTIONS CONCERNING THEOREM 2

Naturally arises the question whether one can choose exactly 2k vectors with the prescribed conditions. The answer is yes. Let us take the following vectors:

$\boldsymbol{g_1}$=(p, 0, 0, ..., 0, 0)

$\boldsymbol{g_2}$=(0, p, 0, ..., 0, 0)
...
...
...
$\boldsymbol{g_k}$=(0, 0, 0, ..., 0, - p)

$\boldsymbol{f_1}$=(- p, 0, 0, ..., 0, 0)

$\boldsymbol{f_2}$=(0, - p, 0, ..., 0, 0)
...
...
...
$\boldsymbol{f_k}$=(0, 0, 0, ..., 0, - p)

It is easy to see that for every 1 ≤ j ≤k the equality

$\boldsymbol{g_j}$= - $\boldsymbol{f_j}$ and $|\boldsymbol{g_j}| = |\boldsymbol{f_j}|$ = p hold,

and for every 1 ≤ i , j ≤k we have

$\boldsymbol{g_j f_j}$=0.

That means that the estimate is tight.
But as one could notice in our construction we have only vectors whose (k-1) coordinates are zeros.
If k=2z is an even number and p is a prime number that can be written as a sum of two squares, i.e.
p = $a^2 + b^2$ for some integers a and b, then let us define our vectors in the following way:

$\boldsymbol{q_1}$=(a, b, 0, 0, 0, 0, ... 0, 0, 0, 0)

$\boldsymbol{q_2}$=(0, 0, a, b, 0, 0,...., 0, 0, 0, 0)
...
...
...
$\boldsymbol{q_z}$=(0, 0, 0, ..., 0, 0, 0, 0, a, b),

$\boldsymbol{q_{z+1}}$=(b, -a, 0, 0, 0, 0, ... 0, 0, 0, 0)

$\boldsymbol{q_{z+2}}$=(0, 0, b, -a, 0, 0,...., 0, 0, 0, 0)
...
...
...
$\boldsymbol{q_{2z}}$=(0, 0, 0, ..., 0, 0, 0, 0, b, -a),

and let us take $\boldsymbol{q_{2z+j}} = - \boldsymbol{q_j}$ for every $1 \leq j \leq 2z$.

It is easy to see that this construction is also appropriate.

But generally one can ask, whether one can find constructions with vectors which does not contain zero as a coordinate? If yes, then for which prime numbers p is it possible to do it?

Another way to generalize the problem would be omitting the condition that p is a prime number. In this case we would have to consider a lot more diviibility relations (see our proof).

A third way to investigate this problem would be to extend (or modify) the condition concerning the lengths of our vectors.

The generalizations above would be very difficult and interesting, and probably one should use deep number theoretical tools to handle them.

REFERENCES

[1] Freud, Számelmélet, Nemzeti Tankönyvkiadó, 2006.

[2] Kuros, Felsőbb Algebra, Tankönyvkiadó vállalat, 1968.

Practical possibilities for photogrammetric point detecting in open source environment

K. Epresi*, Z. Tóth**
* Óbuda University, Székesfehérvár, Hungary
** Óbuda University, Székesfehérvár, Hungary
epresi.konrad@gmail.com
toth.zoltan@amk.uni-obuda.hu

Abstract- By the spread of UAVs in regular use has increased the importance of automatic recognition of land marks in photogrammetry. The proposed method for automatic recognition is developed under open source software, which results in a time and cost efficient solution.

I. INTRODUCTION

Our aim was to investigate the possibilities for automatic recognition of land marks in photogrammetry by using open source software. Development of automatic point detection in many areas would accelerate and make the process more economical. The UAV (Unmanned Aerial Vehicle) records in several areas are part of everyday technology. For example: the important role in the crime scene, where recording the circumstances is quick and accurate.

For our algorithm, we used the Python programming language and its OpenCV library for its image processing.

First of all, based on our preliminary research [10], we created the “perfect” landmark.

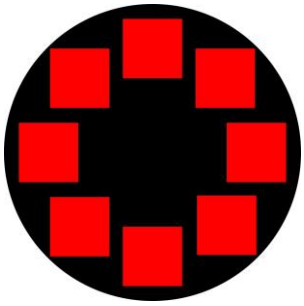


Figure 1. “Perfect” mark

We used the experiences of our previous researches. The point recognition procedure in the entire image area was split into several sections. First, we used the Hough transformation to look for the parts of the image on which we could run our algorithm to give a more accurate result. In each highlighted image part, we segmented the group of pixels which are valuable pixels for search. First, we defined the contour of all shapes (planes) within the circle identified by the Hough transformation using the "Image moment" method and then we calculated the centroid of the shapes. The center of the regression circle fitted to the centroids, regardless of the plane slides shape, which provides the exact position of the cutout image. As the cutout image position is known in the entire image area, the center of the detected point mark can be considered as known throughout the image area as well. The distortion of the circle identified by the Hough transformation (the appearance of the image as an ellipse) gives the opportunity to determine the angle of inclination between object plane and image plane. Based on the literature, we know that the Hough transformation is also suitable for the detection of ellipses [6], but we would like to try other methods during our research as a continuation of the research. Therefore, we hoped to increase the accuracy and we tried out our method.

II. PROGRAM BASICS

A. *OpenCV and Python*

The OpenCV (Open Source Computer Vision Library) is an open source software (BSD license) which could be useful with computer vision, image processing, or even machine learning. It has over 2500 optimized algorithms including filtering, contour detection, histogram drawing, Hough transformation, face detection and face recognition, motion classification, tracking of objects, panoramic image creation, and image-based search. Many contributors to the development (since 1999) as: Intel, Willow Garage, followed by Itseez. We can find Python, Matlab and Java programming interfaces available on Windows, Linux, Mac OS, Android and IOS versions.

Figure 2. OpenCV logo [6]

Interpreter language (not separated from the source and object code) are such languages, where the script could only be executed with a Python interpreter. The Python interpreter has been created for multiple types of PC and operating systems. In addition, several libraries were created, one of them is the OpenCV [6]. We chose this programming language because we studied this at university.

B. *Hough transformation*

First of all, we had to find the easiest recognizable shape and the most applicable transformation. In other word this is the circle and the Hough transformation. Paul Hough, whom the method was named after, designed the algorithm to detect simple, parameter-specific shapes (straight lines, curves) in binary images in 1959, and then patented it in 1961.

In addition to detecting straight lines, other geometry elements are recognized by the process. It is necessary that the complexity of the geometric element carries the number of descriptive parameters needed for storage and computing capacity:

$$r^2 = (x - x_0)^2 + (y - y_0)^2$$

The three parameters are:

- r - the radius of the circle
- (x_0, y_0) - the coordinates of the center of the circle

In practice, there is an edge image on which to find the circle of the known radius, which is formed in the Hough space so that each circle line point corresponds to a circle with potential centers.

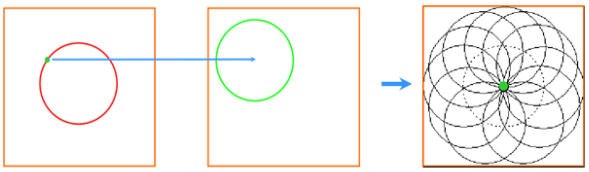


Figure 3. On the left side is the edge image, in the middle the Hough space with a detected circle, on the right there are maximum places in the Hough space → the centers of the detected circles [9].

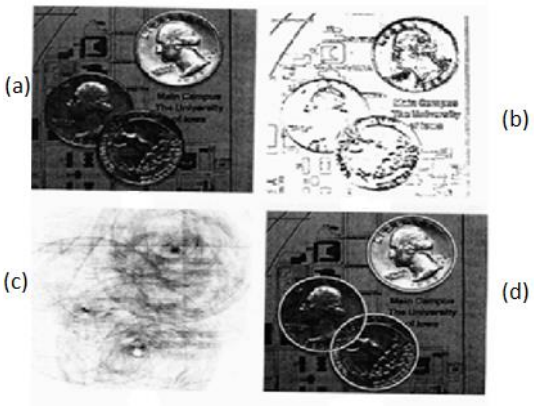


Figure 4. Example [9] (a) original image, (b) edge image, (c) 2D parameter space (searching for circles of a given radius), (d) detected circles in the original image

We used the Hough transformation to cut the useful area. Then we made a smaller image cut from the edge of the image, using the contour detection on this image, where we determined the true point (the center of the circle).

After we use the transformation. This gives us the three parameters of the circle. Use these to make a image cut as shown in the figure. We can do this with the following commands.

cropped = image[A, B]

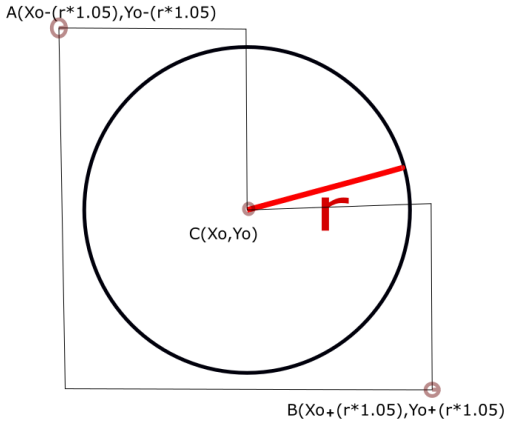


Figure 5. Image cutout using Hough transformation

By the 5% increase in the specified "r" value, we determined the size of the area to crop it to contain the circle line.

C. *Thresholding*

Image Segmentation is one of the most important basic problem of image processing, which deals with the grouping of pixels with similar properties in homogeneous areas. Many images processing tasks (such as shape recognition, surface descriptions, stereo vision, etc.) originate from such processed segmented images. Our choice of segmentation methods was "Thresholding".

The „Thresholding"

- One of the easiest segmentation methods.
- The best example is to split the portion of the image where we have valuable pixels. We need to modify the pixel values of the image to clearly distinguish the object and its background
- First, we separate the values from each other, and we can also set the pixel intensity if we want. For example, the most common case is to change the valuable pixels to white (255) and the background to black (0), but the threshold value and the pixel value can also be changed arbitrarily.

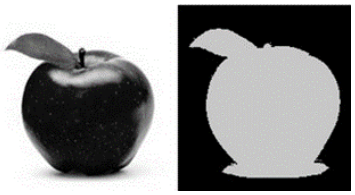


Figure 6. Segmenting the apple

D. *Image moment*

The "Image Moment" is given to determine the centroid of the shape with the following relation:

$$\{\bar{x}, \bar{y}\} = \{M_{10} / M_{00}, M_{01} / M_{00}\}$$

This formula can calculate the centroid of any shape, even in the case of a convex shape.

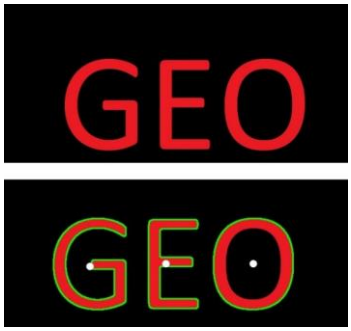


Figure 7. Search for "GEO" inscription centroid

E. *Tilt test*

The next step was to decide whether the object plane and the image plane deviates from the parallel position by more than 5 °. If this deviation does not exceed 5 °, then the circular regression may still work. If this value is exceeded, then we wont use it. In this case, the method of determining the inclination angle is used to determine the centroid of the mark, which also provides the point. The circle is distorted to an ellipse by varying degrees depending on the inclination of the planes. The semimajor axis of the ellipse is equal to the radius of the circle. The task was to pinpoint ellipse. We use contour detection. We applied the following relationship.

$$inclination = (\cos^{-1}\left(\frac{r}{R}\right))/\frac{\pi}{180}$$

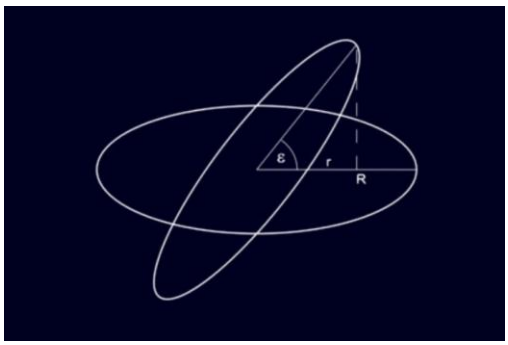


Figure 8. Distortion of the circle

F. *Adjustment computation*

The regression calculation deals with the alignment of simple shapes contour coordinate points. Only points with final coordinates can be included in the regression calculation. In practice, we use the least squares method.

The circular regression method can calculate the best fitting circle parameters sought from the points measured on the circle line. This requires at least three points along the arc. If we add more to the calculation, we will get the center and radius of the most fitted circle, and even their accuracy parameters.

G. *Point detection and Flow chart*

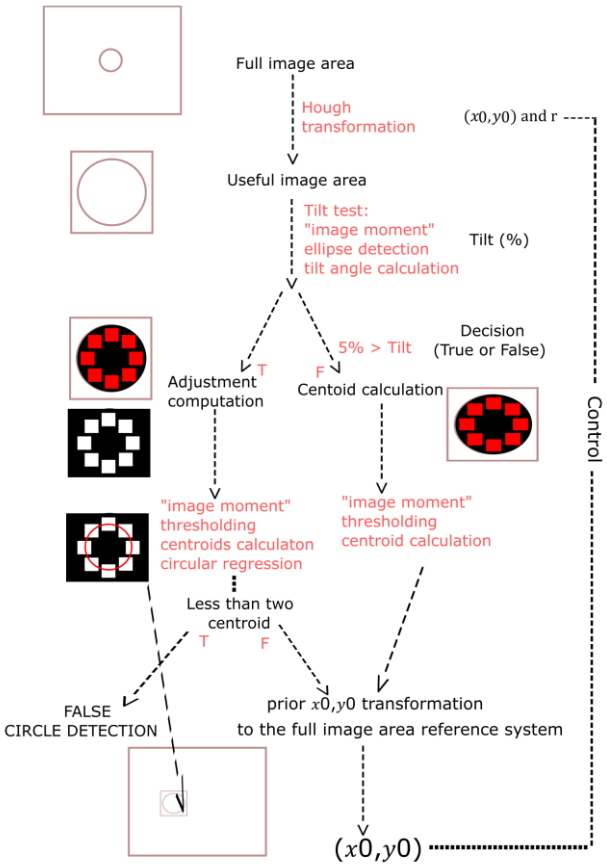


Figure 9. Flow chart

Figure 9 shows the flow chart of the point detection procedure. First, we used the Hough transformation to look for the parts of image on which we could run our algorithm. In order to decide whether or not the slope is over 5 °, the following operations are performed.

- Separate the background and the mark with Binary Thresholding. We created a image that consists of a total of two pixels of intensity. (Black & White Image)
- We inserted an ellipse into the distorted circle with image moment. This is used to determine the angle which between the object plane and image plane is different from the parallel.
- If this angle was greater than 5 ° (the circle was clearly ellipse distorted), the center of the ellipse was considered the center of the mark.
- If the angle value was less than 5 °, then calculate the centroid of the shapes in the circle (triangle, rectangle, etc.). Preparing the program in case any of the shapes in the circle are covered, we have separated and excluded shapes whose area differs by more than 10% from the maximum area shape. If only one or two shapes remain in the circle under the condition, then the signal can not be detected. The program goes to the next mark. In case of 3 or more shapes, a calculated regression circle

has been fitted to the calculated centroids. This will be the exact center of the mark.

In the final step, the defined image coordinates are transcribed in the reference system of the full image area.

The calculation of the difference between the center of the circle defined by the Hough transformation in the full image area and the center of the mark returned from the image offset gives the opportunity for verification.

III. DISCUSSION

The Table 1. Shows our result under laboratory conditions (the center coordinates are 500,500). Unfortunately, we was not able to test field conditions, so we can not report the results of this. All in all, we wanted to create a method that reduces or completely eliminates errors we have detected. In our opinion, we found solutions to these problems, so we could effectively close our research.

Table 1. Results were made under different procedures, under laboratory conditions

Process	**X**	**Y**	**ΔX [pixel]**	**ΔY [pixel]**
Hough transformation	504,600	502,200	-4,600	-2,200
Centroid calculation	500,497	500,491	-0,497	-0,491
Equalizer solution	500,000	499,999	0,000	0,001

According to the promising results under laboratory conditions, the next step is obviously to test the methodology for field conditions, which may be a major project in the near future.

REFERENCES

[1] M. K. Hu, "Visual Pattern Recognition by Moment Invariants", IRE Trans. Info. Theory, vol. IT-8, pp.179–187, 1962

[2] J Richard O. Duda and Peter E. Hart (1971): Use of the Hough Transformation to Detect Lines and Curves in Pictures. Artificial Intelligence Center. SRI International.

[3] Duda, R. O. – Hart, E. P. (1972): Use of the Hough Transformation to detect lines and curves in images. Graphics and Image Processing 3, pp. 11-15.

[4] H.K. Yuen, J. Princen, J. Illingworth and J. Kittler (1990): A COMPARATIVE STUDY OF HOUGH TRANSFORM METHODS FOR CIRCLE FINDING, Department of Electronics and Electrical Engineering, University of Surrey, Guildford, GU2 5XH. U.K

[5] Andrew W. Fitzgibbon, R.B.Fisher. A Buyer's Guide to Conic Fitting. Proc.5th British Machine Vision Conference, Birmingham, pp. 513-522, 1995. The technique used for ellipse fitting is the first one described in this summary paper

[6] Intel Corporation (2002): OpenCV Reference Manual

[7] Tóth Zoltán, Barsi Árpád, Analyzing road junctions by geometric transformations, INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY AND REMOTE SENSING (2002-) XXXVI:(I/W3) pp. 1-4. (2005)

[8] Bradski, Gary; Kaehler, Adrian (2008). Learning OpenCV: Computer vision with the OpenCV library. O'Reilly Media, Inc. p. 6.

[9] Kató Zoltán (2014): Alakzatok és minták detektálása, Digitális Képfeldolgozás.

[10] Epresi Konrád, 2017, Pontjelek automatikus felismerési lehetőségei a fotogrammetriában. Tudományos Diákköri Dolgozat, Székesfehérvár.

[11] K Epresi, B. Gáti, Z. Tóth, Active contours in geoinformatics,In: Orosz Gábor Tamás (szerk.), AIS 2017 - 12th International Symposium on Applied Informatics and Related Areas organized in the frame of Hungarian Science Festival 2017 by Óbuda University: Proceedings, November 9, 2017 Székesfehérvár, Hungary. 204 p. Konferencia helye, ideje: Székesfehérvár, Magyarország, 2017.11.09 Székesfehérvár: Óbudai Egyetem, 2017. pp. 187-190.

Different data storage built on ERP objects

A. Selmeci, T. Orosz
Óbudai Egyetem AMK, Székesfehérvár, Hungary
E-mails: selmeci.attila@amk.uni-obuda.hu, orosz.tamas@amk.uni-obuda.hu

***Abstract*— The ERP systems use tables for storing configuration and application data. The application data consist of master- and transactional data. When an HR expert would like to read employee data, the programs should read many corresponding master data tables and in some cases transactional tables as well. Afterwards the application builds up the employee object and the corresponding objects, like absences, qualifications, etc. Another example could be an invoice object, which contains the item objects and the invoice can be approved, rejected, parked, etc. The running systems are using object oriented thinking, but they convert the data from tables into object or vice versa.**

With our novel approach this conversion can be omitted. In our pilot environment we have defined some low level objects for test purposes. We plan to build some basic functions as well in the close future.

Keywords: **Object orientation, Data storage, ERP, Fast access, Object model**

I. INTRODUCTION

The ERP systems are well-formed, stabile, general solutions. The data models and data storage are different, but similar in that case that each of them are using RDBMS for the persistent layer and the business logic is defined in a procedural manner. The programming environment can be object-oriented, but the business elements, which are part of the daily process, are simple components, modularization units, but not real objects. We discuss in the chapters the object-orientation concept and its effect on ERP systems and present the object-oriented thinking. There is parallelism in data model relations and object relations, which are important to make a possible change from procedural to object based implementation.

II. OBJECT ORIENTATION AND RELATIONS

The object-oriented concept is deeply involved in software programming. However the paradigm can be treated on higher levels as well. The object-orientation is mainly built on some basic principles as pillars. These are the following:

- **Abstraction:** It is the simplification of the elements in the nature to be able to model them in an artificial environment. It means that we dismiss many unnecessary attributes, functions, which are not important for us. If we consider an object like Employee or CoffeeMachine, they have several properties and we can name some functions, which can be used to set, change the status of an Employee or CoffeeMachine. In abstraction we have to know which attributes of an object are important to be able to model the natural object and which functions are needed to simulate the natural behavior. On the other hand we don't have to know how these functions are working internally and we don't know who the attributes and functions are implemented (we do not know how the coffee machine work really internally).
- **Encapsulation:** It bounds the attributes and functions of an object together into one unit to be able to handle the internal state organically. The encapsulation is used to hide the state (property values of the object) and the implementation from the world. This helps to store specific information about the object and control the access to it.
- **Inheritance:** The objects can have several relationships among each other. One special relation between them is the inheritance. The new object can be derived from another object, so they can build a hierarchy. In this hierarchy some of the properties and functions of the original object are shared or inherited. Standard example can be a lamp and a street lamp, which is the descendant of the lamp object itself. The street lamp is a lamp as well, but it has some additional "features" (properties and functions).
- **Polymorphism:** This means several forms or shapes. This principle describes in object-oriented world that some objects having the same function with different implementation or a function occurs in an object multiple times with different signature (so-called overloaded methods). In programming the first one is a useful tool in many cases. We can imagine a booking company, where we can book hotels, flights, or cars for rental. Each of the booking procedures requires the same data, so the interface of the objects for booking is the same, but they execute internally totally different procedures. In the code we can call the same booking function and the environment can decide according to the given object, which implementation should be executed.

These principles are part of the core concept of object-orientation. We did not describe, but the similar objects can be classified and managed together. So we call the

same type of objects object-oriented class and the instances of such a class is an object [1].

We can also define other relations than inheritance between objects, like part-of or has-a. (The inheritance can be written as is-a relation. As we discussed the street lamp *is a* lamp.) The main relationships are:

- Association (has-a): a very loos relationship, but the most popular one. This is the standard connection between objects in the real World as well, like a doctor-patient, man-car.
- Aggregation (part-of, whole/part): special association type, where the objects belong to each other as part of the whole. E.g. wheels and car, employees and departments. The wheels and employees have their own lifetime. If the car is destroyed, the wheels can be used for another car, or if a belonging department is closed, the employee is still employed in the company.
- Composition (part-of, death): the objects belongs together, they are interdependent object, like brain is part of the body. (If the body object is deleted, the brain object should be destroyed as well.) It can be observed as special type of aggregation, but the lifetime of parts are dependent on the whole.

These principles and relationships can be used in data models as well for designing or defining tables in a relational database management system (RDBMS). The classes with the properties can be thought as entity types in data modeling. Each object is an entity from that type. E.g. people class or entity type, and man or woman as instances of class, objects or entities. The connection between these entity types can be defined with entity relationship model using connection types and cardinalities (see later). Here we should think about the severity of connection types and the available relationship between objects or classes. These differences should be handled and described during design. This is another form of abstraction. We can consider two level of abstraction:

1. Real World to data model
2. Data model to object model

The RDBMS implementation can be established using the association between classes and tables, and objects and records. This kind of representation is not object-oriented, it holds only the data model, but it can establish the relations as well. In the entity relationship modeling we can distinguish the relations with two main characteristics:

- Type or category
- Cardinality

The connection types are similar to the object relations, but here we can add more information:

- Hierarchical: an entity type is dependent on the existence of exactly one other entity type. In this case the ID of the referenced entity type is inherited ("used") by the dependent entity type. E.g. Sales office and offered goods. (The good is assigned to the given sales office. It cannot exist without the sales office.)
- Aggregating/Associative: an entity type is dependent on the existence of multiple other entity types. This type is used to resolve many-to-many relationships. The full ID of the referenced entity types is inherited by the dependent entity type.
- Referential: an entity type refers to another entity type but is not identified by the relationship. It is possible to change or eliminate the relationship, but as long as the relationship exists, it is non-optional. The primary key of the referenced entity type is passed onto the dependent entity type as a non-key field.
- Conditional-Referential: the same as a referential relationship, but the relationship is optional.
- Specialization: an entity type that represents a subset of the referenced entity type but also has additional attributes. In this case the referenced entity type is called the generalization, and the dependent entity type is called the specialization. Generalizations and specializations have the same primary key fields but are used to store different information in non-key fields.

The cardinality (n:m) describes the possible entities from the dependent entity type (target) from the viewpoint of the independent entity type (source). Referring to *n:m* relation the value of *n* is generally 1, which means that there is exactly one independent (on Figure 1. left side) entry.

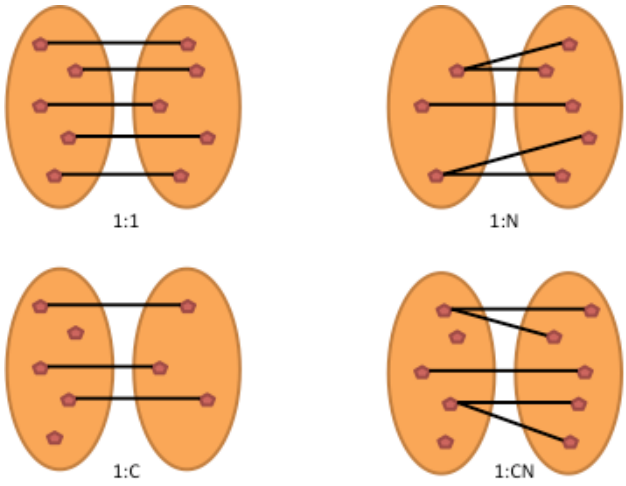


Figure 1 - General Cardinalities

The m value, as the Figure 1 shows can be 1,C,N or CN. Easy (but not really realistic) examples can be for 1:N the mothers – child assignment, because a woman is a mother when she has at least one child, but can have more as well. The 1:CN cardinality is the most popular one. As an example we can take person – car assignment, where one person can own one or more cars, but there are persons, who has no car at all. The 1:C cardinality is an interesting one. A technical implementation can be the ID and the description in a given language. It could be that there is no translation for the given language of some entities, but the others have. The 1:1 cardinality means that each dependent entity is assigned to exactly one independent entity, which would mean in RDBMS area two tables with same key having the same number of

entries, where the corresponding keys values are the same. It would be an extension of the first table.

These cardinality levels are connected with the relation type as well. We have discussed the cardinalities, where each of the dependent (rights) entities is connected exactly one independent (left) entity (n=1). If the n value is C, that means a dependent (right) entity can exist without an independent (left) entity. This kind of cardinality can occur only in case of referential connection type.

As we see the data model is almost similar to the object model in creating relations between entities. Certainly there should be a functional model in case of data model, but the objects are different. The object are not so flat type as the tables in RDBMS, but they contain the functions to be set or read the state of the corresponding data.

III. ERP OBJECTS AND MODELS

We can bring this abstraction from the very deep technical level to higher floor as well and we can define the objects / classes in ERP systems. In ERP system we can speak about objects of business. Some example can be employee, invoice, material, purchase order, etc. These objects in business level have more than only attributes (properties) and functions (methods). There could be a special element of object, the event. The standard, we can say static model knows who should be informed about changes. E.g. a price of a good has been changed, the invoice content should be changed before sending it to the customer (of course it depends on the business logic and policy as well). In the static model the price object should deliver information to each of the objects, which use this object. It means that the object should keep a list of connected object at least to fulfill this requirement. If we create a proper object model, this data mustn't be stored in the object. The events can help here, because the object just publish the state change via this event and any object interested in can catch and handle the event. It should only register for the event. This is a simple publish-subscribe method. In business world we can imagine a process where the employee is hired and some internal orders are created for the different group to prepare the office, laptop, phone, network connection, users in systems, etc. The equipment (like phone, laptop) should be assigned to the employee (association).

Within an ERP system we have to consider the different business modules, which contain several classes connected to each other and some (not object-oriented) interfaces connected to other modules. It is not easy to find the right granularity of objects and define the right functions and properties (abstraction) for them.

The older systems like SAP have their own data and functional models, which is based on the standard procedural thinking. There are many tables containing the master and transactional data with relations according to the RDBMS rules. The functions and many relationship types, categories (explained above) cannot be implemented in the RDBMS, or using an easier way, they are implemented in the applications.

The Figure 2 shows very simplified tables containing sales orders. Two tables represent the order itself with the item entries. These tables refer to the other two separate tables, which contain the material data and the customer list. There is no information on the assets or warehouse in this simple example. The only interesting is that we have more tables for an order. In the ERP HCM (Human Capital Management) or HR (Human Resources) area there are several different entries, which cannot be stored within one single table, because the different characteristics like addresses, positions, children, qualifications, etc. have varying structures. The varying structures cannot be stored within one single table if we do not want to mask the different record entries according the given structure in the table. (It requires a huge programming work during design time and processing resources during run time.)

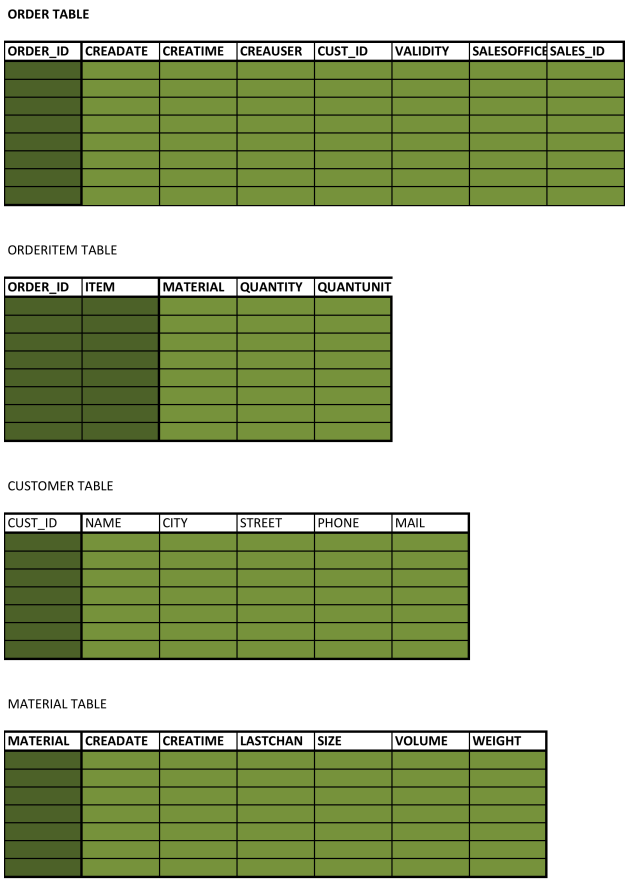


Figure 2 - Simple example from ERP tables

The newer ERP systems are not always built in an object-oriented thinking even if they are implemented using object-oriented programming. Some of them are having real objects from business, like sales order or employee. We referred to the old fashion SAP ERP system already as a system using procedural thinking with data and functional model. On the other hand SAP has implemented a higher level above the data and functional implementation layer. They are called so simple business objects. (These are not the same as the product called BusinessObjects®, which has similar name, but it is a Business Intelligent solution suite by SAP.) These are special programming units written mainly in procedural tools, but having object-oriented behaviors. The naming is not so coherent, because SAP does not use the class word to refer to the structure, but they have the name "business object type". (The instances are the business object.)

These objects have several types of components, like an ordinary object:

- Attributes: there are some attributes for identification, which are called *key fields*
- Methods
- Interfaces: Similar to the standard OOP interfaces containing only definitions without implementation. It is used to add the same attributes, methods, events to different business object types, so they can have the same "interface" for outside callers. The methods should be implemented within the business object.
- Events: for publishing the status change.
- Basic data: Technical data for internal usage.

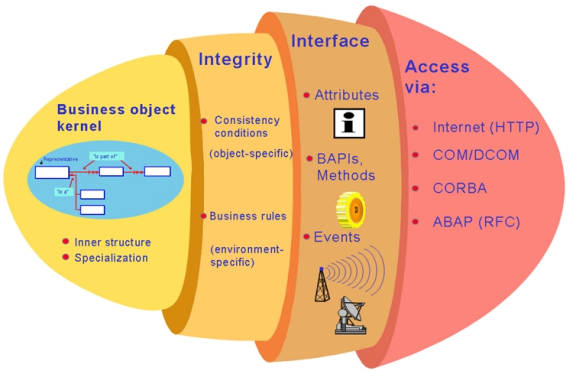


Figure 3 - SAP Business Object Type (c)

As the Figure 3 show, there is a multi-level encapsulation in this case. In the OO world there are visibility sections of an object. These handle the access to the attributes, methods, etc. from other objects (or outside). There are 3 main different sections: private (only internal), public (for everyone), and protected (private from outside, but public for inherited classes). With SAP business objects we have only private and public methods. The other components do not have access-level settings; they are "public" (the attributes are read only). The Figure 4 and 5 show some parts of the standard SAP Sales Order business object as an example. As earlier referred the attributes are divided to two parts, highlighting the key attributes. The Figure 5 shows the methods, where the little green light in the middle sings whether the method is public or not. The green ones are public, which could be called from outside the SAP system as well. These methods are called BAPIs [3]. So SAP did not really implement a real object oriented business model, but tried to define this layer offering an OO thinking and usage.

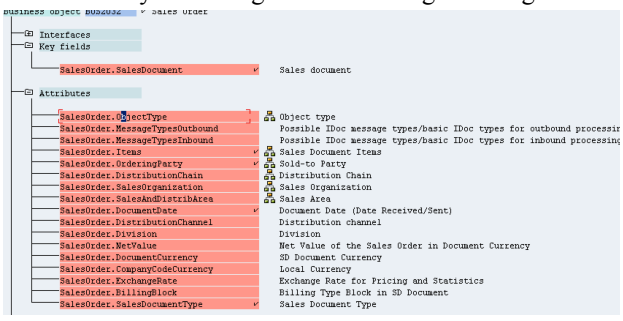


Figure 4 - Sales Order SAP object attributes (c)

The Figure 3 shows some access possibilities from outside. There are libraries for C++, Java and .NET to be able to build externally the object imaging and call the methods of the remote SAP business object via the local object [2].

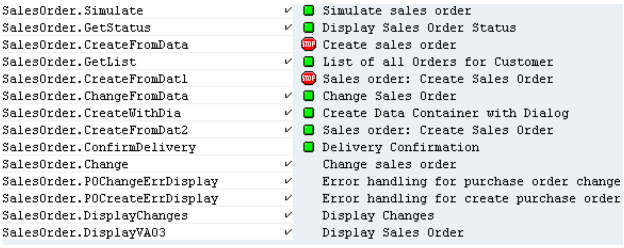


Figure 5 - Sales Order SAP object methods (c)

With these BAPI methods (signed with green on Figure 5) the external program can get a list of orders (list of the instances of this business object type), create one, confirm one or even change it.

This example helps to understand our purpose in this paper.

IV. DATA ACCESS

The ERP systems store data in RDBMS, which means they use the original data model, like on Figure 2. If we think about accessing the data we have mainly two kind of technique:

- Direct access: creating, changing or displaying (or sometimes deleting) entities by ID or key
- Sequential access: listing data (e.g. sales order list)
- Searching: subset of sequential access, but needs to read according to the attribute values

An ERP system is an OLTP (Online Transaction Processing) type system, so it mainly uses the direct access. The others are interesting only for value helps or consistency. You may say that there should be many reports and searches (queries) in an ERP as well. You are rights, but according to the nowadays technologies for reporting it is better to use data warehouse or analytics above big data environments, so we may skip those access requirements for a while.

Let us suppose that we would like to get the data of a sales order. The simplified database (on Figure 2) requires two reads. One read from the head and another from the item table. In case of SAP business object the standard way to read the status of a sales order from database accesses at least 5 tables. If we consider an employee the system should read more than 10 tables. If we check the opposite way, when we create an object (sales order, employee), so many or more tables are accessed directly.

It means when an ERP system would like to access a business object it should read several tables and build the object itself. This is the point where we see the possible object-oriented thinking for ERP systems. Logically we work with objects, but we always have to separate the content and store them in several tables, or select the corresponding rows from different tables to be able to set up the object.

If check again the parked sequential read problem we have to recognize that we always have functions to generate a list from instances of a class or object type.

Unfortunately these lists are not really useful for queries, because they do not contain the attribute differences, we cannot filter on specific attributes.

V. POSSIBLE WAY

According to our study at the first level we recommend to store objects directly. To store the object we have to have an object store. There are several possibilities to store object directly in a database. These are the Object Oriented database systems (OODBMS) or object-relational database systems (ORDBMS). Both are related to store and handle objects [5].

During our study we have faced some issues. The main problem is the searching. We cannot easily search for objects if we would like to filter on specific attribute values. There are no general tools to do that, because the objects are different. One may have read only attribute, which can be access directly, but other gives the value only via get methods. It could be even worse when more attribute values should be used for a complex search (e.g. those sales orders generated in a time interval are interesting, where the customer lives in a given city. This is a kind of join, but it could work only in ORDBMS environments as join. The other problem is the authorization or access control. According to a query the database can generate the list corresponding to the conditions, but it could contain those object references as well, which are not allowed to touch via the user. It goes beyond our paper currently.

Indexing could be a good solution, but for that we have to build up metadata and create index on them. The access control generates another issue of re-indexing on access change.

As we have wrote already we tried to step forward with the object oriented data storage. The general ERP system architectures have at least 3 layers: database, application layer and the presentation (or user interface) layer. Under the database system there are some infrastructure elements for real, technical data stores. As we see the database layer is handling complex data access, manages files, makes available multiple accesses, manages indexes, etc. [6]. We started in our study to work with an object-storage to make a persistent object layer without any database. Currently we have installed a single node Ceph (https://ceph.com/) object store, where we could figure out the way, how we can leave out the database layer.

We have defined some basic objects and pools. We can handle the data objects and the defined methods as well. Unfortunately the library does not support too much capability of object relationship, indexing or inheritance. Using some metadata can approximate all these functions. According to our study our choosen object store solution (Ceph) does not use metadata server of object store technique. We should consider other solutions as well.

VI. CONCLUSION

There are not too many OODBMS or ORDBMS solutions in the market. During the last par years the other NOSQL database solutions decreased the popularity of these database systems. We see that the ERP systems should be object oriented, because the information stored there, are mainly object based ones and not function based ones. For statistical and data mining purposes the RDBMS or object-oriented databases are not useful and the World is going to this direction with the big data platforms.

We have faced many issues concerning object-oriented databases. Some were listed in this paper, which are also referred in other papers. The usage of the object store as different data storage is a new direction of ERP systems, which open a new research area.

REFERENCES

[1] Igor Barbaric, "Design Patterns in Object-Oriented ABAP" *Galileo Press Bonn-Boston*, *2010*

[2] Rich Heilman, Thomas Jung, "Next Generation ABAP Development" *Galileo Press Bonn-Boston*, *2007*

[3] Selmeci, A., Orosz, I., Orosz, T., "SAP BAPI as a break-through and future communication enabler", 2011. Dec. AIS2011, ISBN:978-615-5018-07-7; 978-615-5018-22-0

[4] Priyanka J. Pursani, Prof. A. B. Raut, "*ITFOOD: Indexing Technique for Fuzzy Object Oriented Database*", March 2014, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 3

[5] Georghe and Bucharest, "*Comparison of RDBMS, OODBMS and ORDBMS*", Journal of RevistaInformatica Economică, 2007, Vol.4 , No.44

[6] A. Selmeci, T. Orosz, "*Usage of SOA and BPM changes the roles and the way of thinking in development*", 2012, In: IEEE 10th Jubilee International Symposium on Intelligent Systems and Informatics (SISY), Subotica Szerbia, ISBN 978-1-4673-4751-8

Novelty in storing ERP data

A. Selmeci*, T. Orosz**
Óbudai Egyetem AMK, Székesfehérvár, Hungary
E-mails: *selmeci.attila@amk.uni-obuda.hu , **orosz.tamas@amk.uni-obuda.hu

***Abstract*— ERP systems are at each company, nowadays as standard and basic applications. The data managed within an ERP system should be accessed differently according to the daily habits. To store and access data has been changed during the last years, and according to the requirements it should be changed further.**

Among others the ERP software architecture, the disk access, the bandwidth, the latency, the disk number and type are main factors in speed and reliability. In our paper we compare the different approaches from ERP point of view and show a new direction, which can be a manageable innovation for companies to better store ERP data.

Keywords: **Storage, SSD, FusionIO, Local storage, Architecture, ERP, Innovation, Sustainability**

I. INTRODUCTION

To install an earlier ERP system we had to design the disk layout accurately keeping in mind the behavior, functionality, and mechanism of many different hardware and software components. Some of these components should be mentioned, like the ERP solution itself, the underlying database management system solution and the operating system. The business data is stored generally always in database tables with primary keys. In some cases the relation between the tables are defined in the database level, but there are ERP solutions, where the table correlations are defined on application level having own dictionary for that. In both cases the data is read from (and stored in) tables, which are not always accessed by unique key attribute values, but several times by not key attributes. To have an acceptable response time from the database layer indexes should be created for the frequently searched attribute sets. The first database system based ERP systems required special attention to isolate table data and index parts to separate disk areas. The main reason was the speed, because the database can read the index and data parts simultaneously in that case. The data store of ERP systems as technique and method can be divided into three main areas:

- Business data representation in the ERP system (logical view)
- Database storage of the internal representation (technical view)
- Data or disk storage technology (physical view)

These different approaches cannot be handled as separated, unequally replaceable methods, but they are rather layers, which not in any case are really distinguishable. They have to be managed together as close as it could be. But different groups, companies develop the different layers, so it is not really possible to handle them together.

We show how the different layers envelope and how to follow them to achieve better performance in ERP environments.

II. GENERAL OVERVIEW

In the hardware world the physical layer was improved in the last two decades. The two main technical goals were generally the reliability (e.g. redundancy, high availability) and the response time, speed (e.g. load distribution, bandwidth). The early data stores used different RAID (redundant array of independent disks) technologies to approximate both goals together. The capacity and the speed of the disks increased and it was available to build and use larger data stores. The first RAID controllers were hardware components only, but afterwards some software configuration tools were offered to manage, design the disk distribution in a server. This direction founded the early storage system and solutions having specially connected disks with special-purpose hardware and an operating system offering many different features and services.

According to our study the storage vendors did not really follow the new hardware developments, options, but only the palette of services grew in a wide range.

Nowadays the storage systems offer several redundancy, high availability, data handling (like de-duplication), backup, fast copy, read cache and other options. These features are very useful and required in some cases, but the costs are much more higher and the value that they provide. On the other hand the speed as a goal is not in the main focus of the vendors, because they advice only faster disks or even flash drives (SSD-s), and of course more disks to distribute the load among the spindles. These directions are correct but with having many features, special connectivity layers, like Fiber channels and switches, the disks, the real data storing surfaces are far away from the database server or even application server CPU-s. The newer and faster hardware elements made easier to use bigger, faster and reliable 'local' disks in a server. The local disks or even PCIe storage cards (like FusionIO) provide data in a much higher speed limit, which are sometimes necessary. Our measurements proved that less CPU power is enough if the data storage layer can faster provide data for the ERP system (in case of commercial RDBMS it could decries license amount as well).

The top layer of the above mentioned three data handling approaches determine the mechanism of data management and data store in lower layers as well. Each ERP system contains several business modules and functions, which handles business objects, like accounting documents, employees, sales or purchase orders, etc. The business objects can be defined as object oriented entities

with status (defined via properties), object modification functions (declared through class methods), and of course published status modifications (managed by events) to inform other objects about status change (like employees were hired or even fired). The business objects on technical level in an ERP system are not always defined as real object oriented entities, but only logical correspondence and relationships are defined between data or even tables. Many ERP environments can manage them as objects in higher level, but each time these objects are stored in database tables separating the current status to several tables connected in the database (or as described above in ERP system) level via foreign key relationships. The object as a unit or even entity cannot be recognized on database level.

According to our study for ERP and other kind of OLTP (Online Transaction Processing) systems the object based data handling on logical and technical level would be more effective that the currently widely used table based representation. This direction goes forward to the database layer, where the object oriented database solution would handle better the ERP or OLTP requirements than the relational database management systems.

Our paper collects the availability of the offered technologies, which could be used for ERP systems and architectures as future environment and technical solution. We try to refer to data-warehouse systems as well pointing out the differences in architecture and requirements positioning the possible directions. We figure out the advantages and disadvantages of the different technologies, approaches and try to sketch a way, which leads to now opportunities of data store for ERP systems.

III. PROGRESSION OF PHYSICAL LAYER

The physical layer was developed a lot with solid-state solutions. Figure 1 shows some of the different possible storage usage. As we see close to the CPU we can only use smaller capacity, but by increasing the distance the capacity will increase as well.

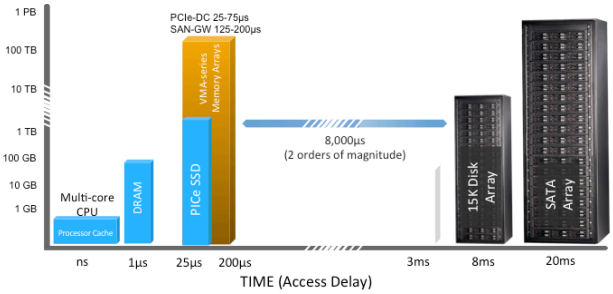


Figure 1 - Solid state storages (HPE ©)

If we consider a machine, there the I/O bus, where the data goes thru in case of local disks or DAS (Direct Attached Storage, see later as well) solution as the Figure 2 shows. The latency is almost the same, but the bandwidth and IOPS (IO per second) depends on the speed of the used disk or DAS containing the disks. If we use SSD instead of HDD (e.g. SAS, Serial attached SCSI) we recognize a magnitude in the speed. Technically the latency, IOPS and distance to the CPU can be used as KPIs for storage systems. [1]

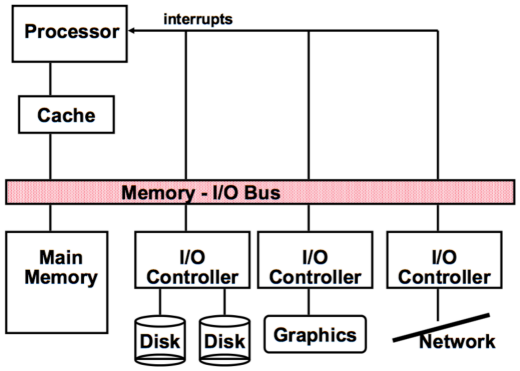


Figure 2 - I/O bus in general PC architecture

Certainly the main memory and the CPU caches are much more faster. (The currently used DDR3 and DDR4 RAMs have 192 and 204.8 Gbit/s bandwidth.) The memory speeds (10-100 times slower than CPU L1 cache) and bandwidths values are much higher than the rates of disks (HDD access time is 4-5 order of magnitude). If we think about using faster storage, than we should go closer to the CPU. One possible way is using the PCI express connections as shown on Figure 3 below (source: https://www.design-reuse.com). (As we can see on the Figure 3 the graphical card is on the PCI express bus as well. We can use the GPUs (Graphical Processing Unit) not only for gaming, but for calculations as well. See later in this paper.)

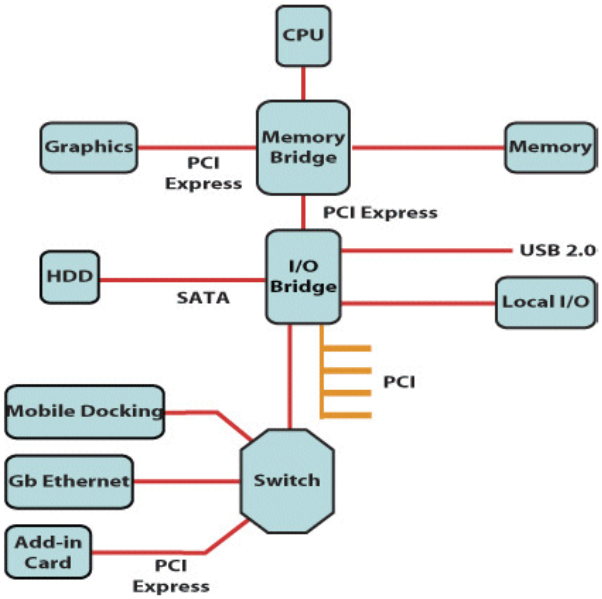


Figure 3 - PC architecture with PCI express

The PCI express solution is using switched topology. It helps to exploit the bandwidth and enables multiple extension opportunities. As the Figure shows the SATA and PCI connections are much further to the CPU than the PCI express bus. There are some products like the first fastest solution: FusionIO (SanDisk acquired in 2014 and in 2016 Western Digital acquired SanDisk), M.2 or similar. So with such memory (SSD) card we are closer to

the CPU, so the latency can be omitted and the bandwidth is what the PCIe can provide.

There were some measurements made by HW companies as well to see which solution give better performance. We should keep in mind that the speed and reliability are working in opposite directions. So if we use a faster solution, like PCI express card for storing data, the reliability and available may be suffered (see later).

As we write the disk and memory speeds are not comparable, but using SSD and PCI express we can narrow them to each other. These cards are also called NVM or NVMe (Non-Volatile Memory (express)) cards. In the future we suppose the use really NVMe cards instead of memory if the speed is just as good as the RAM has [3]. It is a very good opportunity to use such memory cards, because the used memory should not be recreated and rearranged if a system (like ERP) should be restarted. The processes should only reconnect to their memory areas. Some applications before stopping generate special memory dumps from the used important caches, buffers to be able to provide a faster start time. These content can help in it by loading the data directly into the place where they were before stopping the system. The old UNIX systems (like HP-UX; Solaris, AIX, etc.) are not really prepared for this usage. Therefor the several Linux distributions and MS Windows server solutions are more beneficial. (As we have tested with an Oracle database, the Mac OS can be used as well with such solutions, but technically these machines are not really designed for data center usage.)

The Figure 4 shows a similar approach where the internal storage areas and the storage system were measured and the calculation proofed that the storage systems are very slow compared to the internally used solutions.

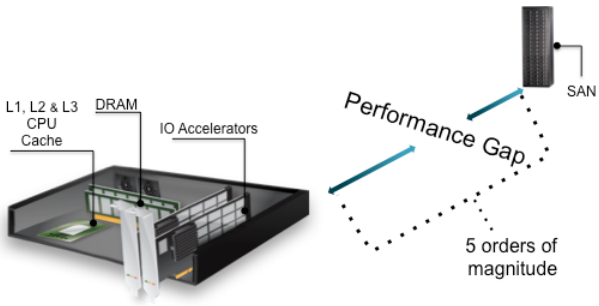


Figure 4 - Data distance from CPU (HPE ©)

IV. STORAGES AND ACCESSING THEM

As the Figure 1 and 4 show the further the data from CPU is, the higher the latency is. Not only the latency is a problem, but if we have a storage system and we put several SSD instead of HDD, though the SSD can be faster e.g. 30 times than the HDD, the whole speed will not be so grate. It is easy to think about. The storage is a computer with CPU, RAM and operating systems. It runs the program, which manages the data through the disks. Some components are built from hardware elements, but others are still software parts. The disks are attached to this internal CPU using IDE, SATA, PCI, and SCSI buses. The nodes of the storage if available have wide internal connection. The internal program should manage the disks and the defined layers of the storage. Some features like snapshots (creating logical copy from a volume) or de-duplication (store the same data only once and refer to the original) can also be available, but these will decries the overall performance as well. The faster elements, like SSD are connected on the same way, so the bus speed, internal calculation will derogate a part of the disk performance in the direction of outside world.

That is why the storage vendors offer PCI express accelerator cards into these storages. These cards solve as a cache in these machines offering read and write cache as well. Other usage can be in hierarchical storage management to provide different speeds for data storage and retrieval. In this case the storage CPU manages the data movements permanently between the different layers according to the measured usage statistics.

To access the data there are many options. Some of the database management systems under ERP used so-called raw devices (many of them can use row devices nowadays as well). It means that they write directly to the disks, they do not use file systems on them and they bypass the operating system as well. The file systems can degrade the disk access IO performance, which is very important in case of databases. The raw devices can be bound to block devices as well. Here is the point what is also interesting for us. We can use different file systems and access methods on disks. As the Figure 5 shows, there are generally three used storage environments: DAS, SAN and NAS.

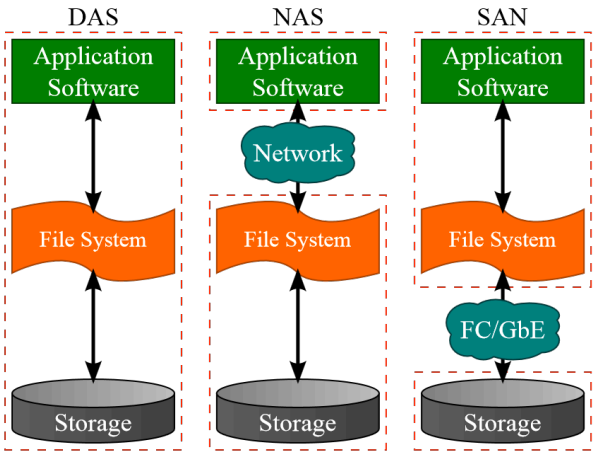


Figure 5 - Storage networks (source: abdullrhmanfarram.wordpress.com)

The difference is not only on the access, but also on the real storage on the disk layer.

- **DAS** (Direct-Attached Storage): disks are connected to server locally (internal DAS) or via SCSI or FC (Fiber Channel) building an external DAS. External DAS: bigger data amount, slower access (distance and protocols). Cannot be share between compute nodes.
- **SAN** (Storage Area Network): multiple hosts can use multiple storages (traditionally) via FC-switches (a kind of virtualization). Special SAN network should be built and the compute nodes should have SAN cards (in most of the cases FC-cards).

- **NAS** (Network-Attached Storage): IP based network and file sharing. Generally CIFS and NFS. Own operating system providing protocols, and storage features (snapshot, de-duplication, etc.).
- **IP-SAN** (Internet Protocol SAN): IP based SAN using iSCSI (block-level access)
- **CAS** (Content Addressed Storage): not frequently changed data, data objects (archived data) can be stored as object. Many benefits for object stores and prices, using redundant array of independent nodes (RAIN) communication via high speed interconnect channels. (Optical or data archives.)
- **VSAN** (Virtual SAN): this network technique is generally part of the VMware solution. It offers SAN access to disks managed and offered by V-Sphere.

These (NAS, DAS, SAN) storage networks can differently access the data on disks. It is important to see on Figure 5 that the red dashed line illustrates the computing unit. In case of DAS we have one unit, because the disk is attached directly to the server. In the other two cases the compute node (above mentioned as "Application") and the storage node are separated. The disks are attached to the storage node and there is a network connection between the storage and compute nodes. The main difference on the Figure 5 is the network place and type, because in case of NAS we have Ethernet network, but in case of SAN we have the fiber network.

There are many opportunities to connect the disks, virtual disks or storage provided disks to the compute nodes. The Figure 6 shows some example.

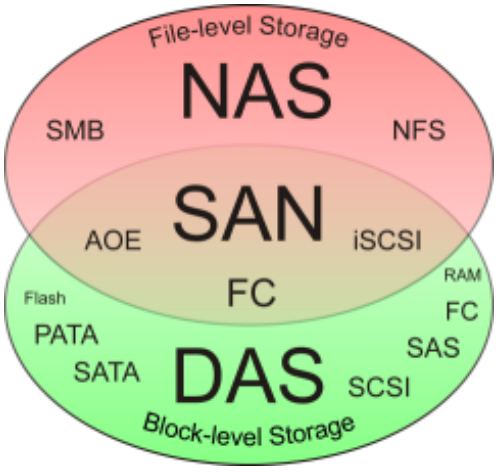


Figure 6 – Disk accesses on storage networks (source: mierebustelur.blogspot.com)

The importance of this Figure is in showing the access levels:

- Block-level access: the data is stored on disks as blocks and accessed by logical block addresses. Databases store the data in the database as well using blocks. The size of the database block and the storage block can play big role in performance (size is database activity dependent).
- File-level access: technically uses block access, but it is hidden, because the files are accessible by name and path.
- Object-level access: the data belonging to a logical or application level object are organized and managed together as one unit. (Not really used for general applications, but for object storage in clouds.)

We do not want to list the shown connection type, though some of them were already mentioned in this paper. They are not in the focus from perspective of our paper.

On the other hand, it is important to understand by the companies using ERP systems that the different data access can help in some special cases. Generally the file access is better if we have many smaller files, like pictures or log data, but block access can be helpful if the application reads binary data blocks. The block sizes should be comparable between the application and storage level. The ratio between them can increase or damage the performance. Theoretically we can say that the SAN is better for database access because the fiber channel should provide "light speed" access to the data and the block-access helps to read the required database blocks from the device. We have to consider that the block access is very difficult, because the data is not really tracked and can be easily damaged. That's why we use file systems to store and access data not using block addresses on compute nodes. NAS gives a different approach, we access the file systems of another computer in this case the storage nodes. It can be used in SMB or HOME environment if we have only pure network. But as the Figure 7 shows we can reach much better network parameters as well.

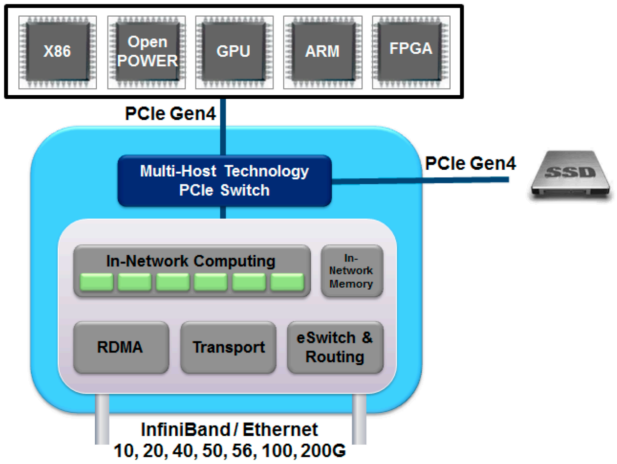


Figure 7 - Mellanox Infiniband card ©

There are some vendors, who could provide 100G or better Ethernet over InfiniBand. If we have so wide network access, the file-level access is much better than the SAN storage networks. First of all the FC card speed is limited around 16Gb, which is much less than 100 or 200Gb. (There are new 32-128Gb FC adapters (gen6, gen 8), but these cannot be connected to UNIX systems and the switches does not support each. [6]) Though these newer FC cards are attached to the PCIe bus, so they communicate to the CPU very fast. If we calculate all these, the result shows that a NAS could be faster even if the FC uses fiber cables.

We should also consider the Software Defined Storage (SDS) systems. These can bring some new ideas as well. If we listen to the voices of the storage arena we hear nowadays about convergent and hyper convergent storage techniques. These are very interesting and useful directions in case of ERP systems. The converged storage just tries to put the ideas together (but goes in another direction). In such a solution we integrate the compute node, the storage and the processing as well, so the latency can be small, the speed high. This technique is optimized and designed for virtualization and cloud platforms. The virtualization is not a bad idea in ERP environment. It could be used in many layers [2]. But we think according to our experience, that virtualization can decrease the disk performance in the database environment.

We can go further to see the hyper-convergent storage idea, where the storage nodes are connected to each other to build a virtual storage layer. In this case we can use DAS for faster storage nodes and InfiniBand network environment among the storage nodes. Each node can manage the disks using redundancy and we can have data copies among the nodes as well. Here we have the possibility to use other storage features as well, while we lose some speed (the features are always costly staffs). This solution is very tricky, because we can use even DAS for the storage nodes, but the communication to the compute nodes are made via NAS using the same InfiniBand network.

We have built such architecture under a production SAP ERP (HR) environment as shown in Figure 9.

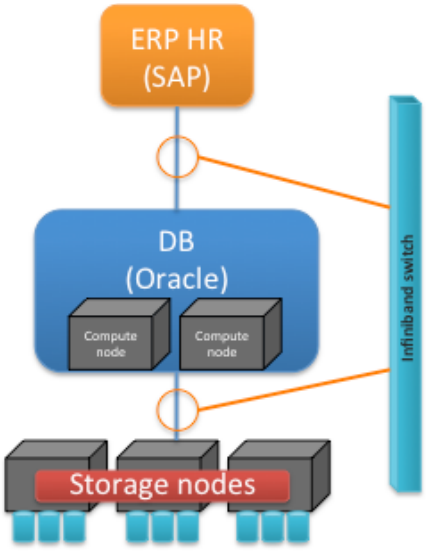


Figure 8 - POC environment with SAP ERP

The storage and compute nodes were Linux machines. We have to mention that the compute node is not the application level in this case, but the database engine (Oracle in this case) level. We have run the Oracle RDBMS on two nodes using ASM (Automatic Storage Management) volume manager and file system. The SAP system was installed on a similar machine running MS Windows Server as operating system. We had another environment built by the company experts. That configuration had a vendor delivered all flash SAN storage system and a huge, but virtualized DB server connected via FC cards. The SAP application layer was attached via standard 1Gb Ethernet as shown on Figure 9.

We have compared the present production environment with these two solutions using the same SAP ERP system. The result was impressive. Our small HW elements using SAS disks attached directly to each storage nodes and InfiniBand and NAS solution performed better (120%) then the present larger environment. On the other hand the new storage system fully loaded with SSDs and the preferred virtualized hosts could provide less performance (80%) than the original one. This information was a good example to the IT leaders to think about new technologies, services and solutions. It requires different approach on standard support and maintenance, because the workers should learn and manage a self-made environment, where they could take part on the implementation phase. In the other direction they learn and use products from vendors. They have ready solutions and limited compatibility. The new generation in ERP IT management should be able to work in a DevOps and LiveOps manner.

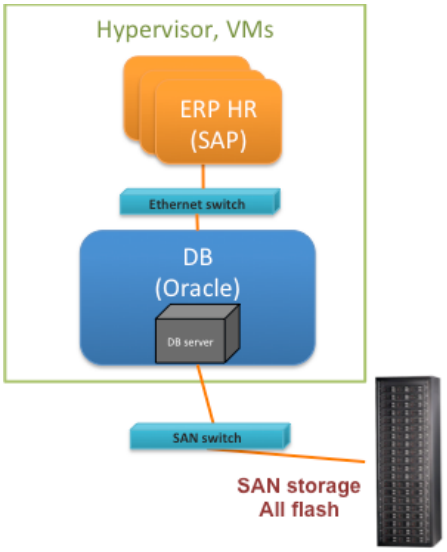


Figure 9 - Storage and VMs

V. OUTLOOK FOR OTHER SOLUTIONS

We have mentioned earlier that the GPU elements can be used as well. It is not really a storage area, but the graphical card designers work hard on providing better, quicker cards with larger amount of memory. Parallel they enhance the programming capabilities as well. The graphical cards are connected via PCI express buses, so they can communicate to the CPU efficiently. In an ERP environment we can think about other problems as well, like making statistics or planning according to the data. Some of these information need large amount of processing capacity, like ML (Machine Learning) or AI (Artificial Intelligence), others require larger dataset to be work on it. The graphical cards have several processors with smaller memories and shared memories. The problems, which could be parallelized, can be shifted to GPUs. Some special development is needed, but the cost and result will be better.

We have tried out another direction with ERP data as well. We have implemented the MAPD GPU based database system as it was released as community product

as well. We had only one graphical card in a server, but we could load many financial transactions and connected data to be searched. The result was amazing.

It can be imagined as an in-memory database having many processing units in one single system. It is useful for statistical analysis on read-only data sets. We plan to test in with ERP connection as well in the future.

VI. Conclusion

Almost the main point or component of an ERP system is the database. It should provide the data fast enough to be able to process it on application level. There are some special read-only data, which can be handled using other methods. The access time during write and read depends on the distance of the data and the CPU and the latency of the access. There are many technical, physical developments, where new technologies can and should be used under ERP environment.

The disk speeds are higher and the connection possibilities provide wider bandwidths as well.

We should consider not only use SSDs instead of HDDs, but relying on the practice, where more slower drives (or spindles) can provide faster environment, than less, faster and larger drives. We have also figured out that the fastest drives connected via PCI express bus to the CPU can provide speed, but the redundancy is needed for reliability in case of ERP systems.

The network speed and latency between the storage and compute nodes (where the database is running) are also important. Using wider bandwidth we can reach higher performance. That's why we recommend to use, but at least test the NAS architectures based on InfiniBand switches.

Our test environment presented that using software defined storage and having appropriate DevOps groups we can reach much better performance and reliability on lower price than the vendor provided solutions.

Our tests and paper has shown that the companies should think about using the new features under their ERP solutions.

References

[1] A. Selmeci, I. Orosz, Gy. Györök, T. Orosz: "Key Performance Indicators Used in ERP Performance Measurement Applications", 2012 SISY, Subotica Szerbia, pp. 43-48., ISBN:978-1-4673-4751-8

[2] A. Selmeci, T. Orosz, Gy. Györök: "Innovative ERP virtualization", 2013 SISY, Subotica Serbia, pp. 69-75., ISBN 978-1-4799-0303-0

[3] http://www.roceinitiative.org/wp-content/uploads/2017/01/RoCE-Accelerates-DC-performance_Final.pdf, ROCE 2017

[4] https://mierebustelur.blogspot.com/2013/06/macam-macam-storage-network.html, 2013

[5] https://abdullrhmanfarram.wordpress.com/2013/04/08/storage-technologies-das-nas-and-san/, 2013

[6] https://fibrechannel.org/roadmap/, 2018

Function point analysis by an SAP application

Katalin Nátz*, and Tamás Orosz**
* Szent István University, Budapest, Hungary
** Óbuda University/AMK, Székesfehérvár, Hungary
* natz.kata@gmail.com
**orosz.tamas@amk.uni-obuda.hu

***Abstract*— Nowadays from the beginning of the 21st century, the buying interest in SAP technology and Function Points Analysis (FPA) was very high. On the one hand, the software provides the IT landscape for business processes and enables the reusability of the software, which means most of all a big profit for customer. In addition, various solutions allow for the user.**

Of course, one can ask the question whether it pays to count the function points in the SAP systems. The answer is clearly yes. Above all, there are several reasons to measure the size of the SAP systems. For example, im-prove productivity of enhancing SAP. You can determine the purchase decision for purchased packages. You can determine the financial impact of an ERP implementation or improve your financial planning. This is why Func-tion Points (FP) has become the standard measure of software controlling and planning at many companies. There are, of course, big challenges to be overcome: time constraints, broad functional scope, highly integrated into the application, which requires the exact definition of application limits, etc.

***Keywords*: Function point, SAP technologies, Software measurement, Business processes, Financial planning**

I. INTRODUCTION

Applications and large systems like ERPs and the popular cloud platform systems are growing very fast. They are so complex and use such a huge amount of big data that they are becoming extremely difficult for the users to understand. The new coding tools already allow software developers to produce larger volumes of software to meet the ever-growing and complex user requirements. A clear and comprehensible method must be used for understanding, estimating and reviewing the projects.

FSM[1] methods are meant to measure the software size by quantifying the functionality provided to the users. In particular, the Function Point Analysis (FPA) was the first FSM method [1] and was originally introduced in the mid-1970s. Nowadays it is used by organizations worldwide. IBM's Allan Albrecht was the first to publish a method for functional sizing of software called Function Point Analysis (Albrecht, 1979, 1981) [3][9]. His name is inextricably linked with Function Point analysis. As an IBM employee, he developed Function Point analysis in the 1970s to assess the productivity of software projects [4]. Since its foundation in 1986, the International Function Point Users Group (IFPUG) has continuously maintained and developed the original Albrecht methodology for functional software (IFPUG CPM) [3][4]. In 2009, the new version CPM 4.3.1 of the IFPUG Function Point standard was published which has been in use since the beginning of 2010. [4]

The "function point" defines the number and functional size of the functions which are implemented or changed during the considered development, including the functions supported by this IT system. Function point analysis (FPA) can measure entire applications, even projects or releases. FPA is independent of the type of application or the type of implementation. It is a structured problem-solving solution, which breaks down systems into smaller components with the purpose of better understanding and analyzing them.

The appearance of the functional point technique has enabled the ICT community to increase the practice of software measurement with respect to the use of the "lines of code" (LOC) [5] approach. The FP count requires a clear, complete and detailed description level, e.g. design and application documents [2].

Meli and Santillo mention two possible cases where an alternative estimation meth-od compatible with the standard rules for FP could be crucial. The first case occurs when the development or enhancement project is in such an early stage that it is simply not possible to perform an FP count in accordance with IFPUG standards. The second case occurs when a survey of the existing software asset is required, but the required documentations and resources are not available for the detailed FP calculation. [2]

Because of these and other similar situations, the need for methods to estimate function points has been increased for organizations involved in the software business. This need is even more urgent for small businesses, which need usable and cost effective software.

II. FUNCTION POINT ANALYSIS (FPA)

To conduct a Function Point Analysis, we need the functional requirements provided by the project team. We differentiate between transactional functions and data func-tions. After identifying the transactional functions and data functions we have to determine the complexity and the functional size for each. Regardless of the imple-mentation technique used, it is possible to determine the size of the application. With the help of the results the average productivity can be estimated. If you know the experience of productivity, you can roughly calculate the project hours and the associated budget. [4]

A. Basics of FPA

FPA is in fact an evaluation process that defines the functional size of IT systems provided to users. During the measurement we do not evaluate the system itself, but to

[1] FSM: Functional Size Measurement

what extent it supports the implementation of business processes. In the course of the process, the developments implemented within the scope of the given project will be analyzed. Thus, the result of the FP analysis shows the functional size of the development.

Function points are the quantity of the number and scope of functional transactions supported by an IT system. In order to determine the function point value of an IT application or IT requirement, one investigates or determines the number of data functions managed thereby, as well as the input, output and inquiry functions available for the users. [4]

If one wants to determine the function points, the user view is decisive for the de-termination of the functional size. There are detailed rules in the standard for assigning the point value to transactions and databases. According to the number of elementary functions and the standardized values, the assigned FPs is accumulated.

In the FP analysis, we should also talk about the elementary processes. This is the smallest step that is meaningful to the user and constitutes a complete transaction. An elementary process is counted as the same compared with another one if it re-quires the same set of Data Element Types requires the same set of File Type Referenced and requires the same set of processing logic to complete the elementary process. [4][8]

B. Why or why not should we use the FPA method?

FPA offers many advantages over other software sizing methods [4][8][9]:

- It is independent of the programming language, development technology, plat-form or the knowledge of the project members.
- It connects directly with user requirements and features.
- Those who engage in FPA analysis play a consultative role in this case. The function points can help to create effective communication between the developers and users, and to give a better understanding to the non-technical user.
- It can be used on both "Waterfall" and "Agile" projects.
- It is applicable throughout the software development lifecycle.
- It provides a way to determine productivity and estimate the cost.
- FPA provides consistent, documented and repeatable measurement methods.
- FPA can highlight gaps in functional requirements, thus avoiding the early introduction of errors in the application.
- A goal is also to minimize the effort and severity of the measurement process.

C. FPA has also contras [1][7][11]:

- If there is no information about the complexity, it can be only use the rapid meth-od, which can corrupt the result. On the other hand, I have information on complexity, but it cannot be clearly counted, and hardly appreciated.
- Albrecht gives many advices , to identify logical internal files, but it can be very difficult to identify after that. IFPUG publishes many examples, but those are also very open.
- The applications have features and properties that can be almost the same for much software (such as SAP). Therefore, we can also make a big mistake in the survey, because the functions for the similar software are not the same in other companies.

D. Transactional functions

The transactions represent the functionality provided to the user for processing data by application. There are three transactions - external inputs (EIs), external outputs (EOs) and external inquires (EQs).

External Inputs: The main purpose of an input is to maintain one of several internal databases or to change system behavior. An input contains processing logic for processing technical data or control information that enters the application be-yond the limit [4][9].

External Outputs: The output is the presentation of information to the user. It contains at least one of the following forms of processing logic: mathematic computations, maintenance of datasets, generation of derived data, change of system behavior [4][9].

External Inquires: An inquiry is the presentation of information to the user, and it references a dataset to read technical data or control information, and it does not meet the requirements for an EO. It can be also a list box. It is important to mention that sorting and arrangement of data cannot be evaluated at all. [4][9].

In practice, the difference between EO and EQ is often difficult to impossible. A query should only be evaluated if it is clearly evident that the requirements for an output are not met, e.g. in a "simple" list box or multiple search.

E. Data Functions

Data functions show logical data groupings and databases that the user needs for his work. Data functions have two types - Internal Logical Files (internal data) and External Interface Files (external data).

Internal Logical Files (ILF): A dataset that is maintained within the considered application is classified as internal dataset: user-recognizable, logical groups of functional data, maintained by elementary processes of the application [4][7].

External Interface Files EIF): A dataset that is read-only but not maintained within the considered application and that is classified as ILF in at least one other application is classified as an External Interface File: user-identifiable, logical group of functional data that is not changed in the application / to be cared for. It is maintained by elementary processes of another application [4][7].

F. Non-functional categories [12][13]

Regarding to the analyzed sprints and the references we have specified a non-functional categories for the customer services company.

TABLE 1. NON-FUNCTIONAL CATEGORIES

Criteria	Comments/Examples for Collection
	• Security & Compliance: Effort to comply with the relevant requirements regarding security, compliance, legal, revision, etc. /Security requirements (confidentiality, in-formation security, data integrity, availability)
	• Stability: Effort for stability, availability, robustness (this also includes expenses for HWR measures) Reliability
	• Performance: Effort due to higher data volume, number of users, access and response times / performance and efficiency (response times, resource requirements, cost-effectiveness) scalability
	• Usability: Efforts to improve the usability and accessibility of apps, appearance and handling (look and feel), usability (comprehensibility, learnability)
	• Quality: Flexibility (support for standards), maintainability, changeability (analyzability, stability, testability, extensibility), portability (adaptability, installability, conformity, interchangeability)
	• ITSD: Effort to comply with the IT standards (patches, etc.), consolidation (code, documentation) and quality assurance, to comply with the maintenance contracts mandatory upgrades required
	• Incidents: Efforts for avoidance (business) errors / incidents, correctness
	• Migration: Migration effort (data migration and cleanup)
	• Support: Effort for support in requirement specification, user test and user training, process test and active operation validation
	• Launch: Effort and cost for launch
	• Operational Shares: Efforts to ensure optimized commissioning and efficient operation
	• Configuration: Efforts for the implementation of short-term application and product configurations
	• Retire: Effort for the shutdown of applications that are not functional (dismantling / conversion) can be assessed
	• HWR: Hardware Refresh
	• Other: Other non-functional efforts

G. Complexity of the transactions and data functions

Important for the determinations of the complexity are: for a transaction the number of used fields and datasets and for a dataset the number of included fields and field groups.

TABLE 2. COMPLEXITY FOR TRANSACTIONS AND DATASETS

Elements	Complexity		
	Low	Average	High
EI	3	4	6
EO	4	5	7
EQ	3	4	6
ILF	7	10	15
EIF	5	7	10

Both of these files include Data Element Type and / or Record Element Type. A data element type (DET) is a unique user recognizable, non-repeated field or attribute of ILF or EIF. A record element type (RET) is a user recognizable subgroup of data elements in ILF or EIF. When ILF / EIF are referred to as Transactional Functions for Processing Information, they are termed as File Type Referenced (FTR) [9].

TABLE 3. COMPLEXITY OF EI AND EQ

FTR	DET		
	1-4	5-15	> 15
0-1	Low	Low	Average
2	Low	Average	High
> 2	Average	High	High

TABLE 4. COMPLEXITY OF EO

FTR	DET		
	1-5	6-19	> 19
0-1	Low	Low	Average
2-3	Low	Average	High
> 3	Average	High	High

TABLE 5. COMPLEXITY OF ILF

RET	DET		
	1-19	20-50	> 50
1	Low	Low	Average
2-5	Low	Average	High
> 5	Average	High	High

TABLE 6. COMPLEXITY OF EIF

RET	DET		
	1-19	20-50	> 50
1	Low	Low	Average
2-5	Low	Average	High
> 5	Average	High	High

III. A CASE STUDY – FUNCTION POINT ANALYSIS OF A SCP APPLICATION

A. Function point analysis of the customer module of a SCP application

The analyzed company is a service company that deals with customer care and in-formation technology. Since 2016, the company has been developing an SCP application to analyze whether it is worth introducing and fulfilling customer expectations.

The company has been developing this software agile for 3 years so that flexible development is possible and unnecessary documentation is spared. In the development

should be justified, if the software was developed according to the expectations and can be continued. In addition, a possible KPI range should be determined if the data can be meaningfully derived from the data.

B. Steps by the counting [6]

According to IFPUG, following are the steps for Function Point counting:

- Determine the type of Function Point count and the application boundary
- Identify data functions and their complexity
- Identify transactional function and their complexity and non-functional part
- Compare the results in each year, then calculate a possible KPI for each category using budget

C. Counting of function points and non-functional categories

If a greenfield project is to be measured, that is a grateful task. You can get to know the new application very well. But there is also the challenge of finding all the functions and measuring them according to the regulations and to form a KPI with the data to be used.

After completing the application, the following information should be mentioned: The application is designed for customer service to store customer data and support product sales. The purpose of the application is to create a customer database, ar-chive the data and enrich it in a system that can be searched both offline and online. The enriched information should be exported at any time. The user should be in-formed about the activities at all times. The individual fields should be edited, maintained and deleted, but the history should always be available.

As far as the agile software projects are concerned, it can be said that it is a greefield area, you can find less analysis in this case. During the measurement, more than 260 sprints were analyzed and categorized regarding the functional parts.

TABLE 7. SPRINTS OF TRANSACTIONS AND DATABASES

	2016	2017	2018
EI	9	42	49
EO	11	26	21
EQ	9	37	35
ILF	11	6	3
EIF	2	4	2
WA	42	115	110

It is important to mention that it was the first time that the company developed agile. The experiences are still collected.

TABLE 8. COUNTING OF TRANSACTIONS AND DATABASES

	2016	2017	2018
EI	54	4	6
EO	77	5	7
EQ	54	4	6
ILF	165	10	15
EIF	20	7	10
SUMMARY	370	786	716

If we look at the results of the measurement we can see that all transactions and datasets belong to the category "high" regarding to the counting of RET, DET and FTR of functions and databases. In point 2.6 we can find the details for the complex-ity and the calculation of the elements.

TABLE 9. CALCULATION OF FP / EUR

	2016	2017	2018
EI	772 €	698 €	655 €
EO	972 €	999 €	954 €
EQ	949 €	912 €	955 €
ILF	461 €	846 €	848 €
EIF	643 €	738 €	821 €
SUMMARY	694 €	847 €	821 €

If the results are looked at in the FP / Eur calculation, it can be stated that the values are almost between 500 and 1000 FP / Eur. Compared to the FP table, it can be seen that the FP / Eur value is also higher with the increased FP count, which is probably related to the higher and functionality and complexity of SAP.

TABLE 10. SPRINTS OF NON-FUNCTIONAL CATEGORIES

	2016	2017	2018
Support	89	83	121
Configuration	226	95	201
Usability	26	27	68
Quality	4	4	4
Security	5	5	7
Incidents	12	9	32
Stability	8	2	2
Performance	4	5	5
Other	2	2	17
Summary	376	232	457

According to the non-functional list in paragraph 2.5, we identified 9 categories during the measurement. As it is a greenfield application, it is no surprise that the highest number belongs to the configuration.

TABLE 11. COUNTING OF NON-FUNCTIONAL CATEGORIES

	2016	2017	2018
Support	567	433	572
Configuration	1439	495	950
Usability	166	141	322
Quality	25	21	19
Security	32	26	33
Incidents	76	47	151
Stability	51	10	9
Performance	25	26	24
Other	13	10	80
Summary	2394	1209	2161

The valuation of non-functional parts without automation is a developing area. The nFP-s are rated here by the budget and the hours. Several evaluations could be used to deduct clear conclusions and make suggestions for the calculation.

TABLE 12. CALCULATION OF NFP / EUR

	Non-functional categories		
	2016	2017	2018
Support	4 005 €	4 178 €	3 387 €
Configuration	4 761 €	4 689 €	4 479 €
Usability	4 088 €	3 937 €	3 956 €
Quality	3 601 €	3 882 €	3 632 €
Security	3 850 €	3 739 €	4 168 €
Incidents	3 631 €	5 582 €	3 070 €
Stability	3 165 €	3 814 €	3 382 €
Performance	4 614 €	3 882 €	4 082 €
Other	981 €	7 128 €	1 530 €
WA	4 420 €	4 415 €	3 882 €

According to the table, it can be seen that most categories changes in a certain range. The exceptions are incidents and others; where in this case is about bug fixes or unreacted requirements.

IV. CONCLUSION

This article presents results of a case study conducted to analyze existing FPA meth-od in an SAP application where function points are used as a unit of measure to determine a KPI range. We have analyzed the application for 3 years in the user view.

All studies advise to use COSMIC FPA for size estimation. But as the documentation is often incomplete in an agile development, it should conduct many interviews, select the information and identify the data movements. In this case, there are such assumptions that no automated estimates can be made. [15]

This study does not directly provide a cost estimation refinement for the SCP pro-jects for the business scenarios. Instead, it provides an estimation matrix for the analyzed company that represents estimation strategies based on the situation, historical data availability, and knowledge base [14].

These studies show us that business processes are valuable resources for effort estimation. FPA method in the software projects could be a good candidate for the size estimation of SCP projects.

Since the SCP ideas are developing rapidly in addition to the SAP solutions, the FPA method in this area should also be objectively applicable to an FPA user. However, it is not easy to get accurate rule of thumb estimates, especially for new custom requirements. In the case study, the rule for the measurement was in a wide range, which could lead to estimation errors.

These conclusions cannot be generalized to all types of estimation scenarios for ERP systems due to the limitations of the case study. The main limitation of our study is that we only applied the function point-based estimation method to a specific type of ERP system, SCP. Another limitation is the sample. We only applied the method to one project. Further studies are needed, which also cover other cloud ERP systems and different types of ERP projects.

In the study, we found that FPA method is a good option for SCP projects. Critical parameters for this are the incomplete documentation, the decisive change and extinguishing scenarios that are not marked and the extra functionality that should be precisely defined and calculated with the COCOMO model. Further FPA-based studies should be carried out considering this model.

REFERENCES

[1] A.J. Albrecht: Measuring application development productivity, in: Proceedings of the Joint SHARE/GUIDE/IBM Application Development Symposium, 1979, pp. 83–91.

[2] R. Meli, L. Santillo: Function Point Estimation Methods: A Comparative Overview (1999), https://www.researchgate.net/publication/2462027_Function_Point_Estimation_Methods_A_Comparative_Overview, last accessed 2018/03/24

[3] CISQ Homepage, http://it-cisq.org/wp-content/uploads/2014/11/Automated-Function-Points-Specification-OMG-Formal-January-2014.pdf, last accessed 2018/03/24

[4] B. Pönsgen: Function-Point Analyse. Ein Praxishandbuch. 2nd edn. Dpunkt.verlag, Heidelberg (2012)

[5] N. Balaji, N. Shivakumar & V. Vignaraj Ananth: Software Cost Estimation using Function Point with Non Algorithmic Approach, Global Journal of Computer Science and Technology Software & Data Engineering Volume 13 Issue 8 Version 1.0 Year 2013

[6] IFPUG: Function Point Counting Practices Manual, Release 4.0. Couting Practices Committee, The international Function Points Users Group (2000)

[7] D. Longstreet: Fundamentals of Function Point Analysis, Total Metrics

[8] S. Palmquist, M.A. Lapham, S. Miller, T. Chick, I. Ozkaya: Parallel Worlds: Agile and Waterfall Differences and Similarities. Software Engineering Institute (2013)

[9] P. Vickers: An introduction to function point analysis, Northumbria University, https://www.researchgate.net/publication/250894239_An_Introduction_to_Function_Point_Analysis, last accessed 2018/03/24

[10] C. Dekkers: Counting Function Points for Agile / Iterative Software Development; http://www.ifpug.org/Articles/Dekkers-CountingAgileProjects.pdf; last accessed 2018/05/06

[11] B. Molnár: Funkciópont elemzés a gyakorlatban; http://www.mtaita.hu/hu/Publikaciok/FunkcioPont_1.pdf, last accessed 2018/05/08

[12] Gregor v. Bochmann: Non-Functional Requirements

[13] http://www.site.uottawa.ca/~bochmann/SEG3101/Notes/SEG3101-ch3-4%20-%20Non-Functional%20Requirements%20-%20Qualities.pdf, last accessed 2018/10/20

[14] https://www.scaledagileframework.com/nonfunctional-requirements/, last accessed 2018/10/20

[15] http://www.ifpug.org/Conference%20Proceedings/ISMA3-2008/ISMA2008-16-Maschino-can-SAP-be-function-point-counted.pdf , last accessed 2018/10/20

[16] Ömüral, Demirörs: Effort estimation methods for ERP projects based on function points; http://openaccess.iyte.edu.tr/bitstream/handle/11147/6734/6733.pdf?sequence=1&isAllowed=y, last accessed 2018/10/20,

Robotics Simulation Environment in an Educational Institution According to Demands of Industry 4.0

Károly SZÉLL
Óbuda University, Alba Regia Technical Faculty, Budai út 45. H-8000 Székesfehérvár, Hungary
szell.karoly@amk.uni-obuda.hu

***Abstract*—Simulation is an inevitable step for education and prototyping of robotics. The simulator allows you to test ideas in a secure and verifiable environment, preventing robot cells and the sorrounding systems from being damaged. This paper describes a robotics simulation environment that fulfils the requirements of educational and research work.**

***Index Terms*—Robotics, Simulation, Industry 4.0, Smart Manufacturing, AGV**

I. Introduction

Starting from the discovery of planets, industrial applications [1] to cleaning households, the application of robotics is diversified. In the simulation of robotics dynamics, informatics, control theory, sensors and actuators have to be combined, which makes modelling a real challenge [2].

With the development of processors and 3D graphics tools the possibilities in robotics simulations dramatically changed. The models may be more complex in computer simulations, and we can run real-time simulations connected to hardware (HIL) and control embedded systems from the framework.

We can compile a simulator from various open-source robotics libraries, but the architecture of the software environment play a decisive role in determining how these elements affect each other and how the overall efficiency of the system evolves. Robust simulation systems require versatility and scalability.

A robot simulator needs to provide a variety of tools at one time while the abstraction of the underlying systems. In addition, flexible control is required (portability and maintainability), generalized and scalable to different models.

Section II. discusses available open-source robotics simulation environments. Section III. introduces the main elements of V-REP. Section IV. demonstrates educational and training possibilities via V-REP. Section V. presents industrial application of V-REP.

II. Simulation Environments

In the era of industry 4.0 the tools for digitalisation gain greater relevance [3]. For robotics simulation we can find various software environments like Open HRP [4], Webots [5], ARGoS, Gazebo or V-REP. Due to their educational relevance the paper discusses the latter three.

ARGoS (Autonomous Robots Go Swarming) is an open-source multi-robot simulator (the main user interface can be seen in Fig. 1). The software is designed to simulate swarm robotics. It is both effective (performance) and flexible (customizable). We have the possibility to divide the simulated space into sub-spaces that are handled by parallel running physics engines. The architecture is designed to optimize the use of multi-core processors [6].

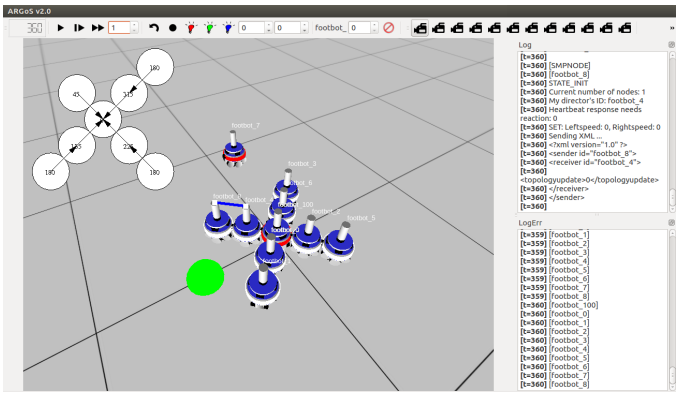


Figure 1. ARGoS[1]

Gazebo is a multirobot simulation, open source software environment capable of visualizing data, simulating remote environments, and supporting decryption of blackbox systems (the main user interface can be seen in Fig. 2). Gazebo is made in collaboration with Player and Stage projects [7]–[9].

Player and Stage projects started in 2001, popular in universities and industry as well. Player is a network device server, while Stage is a simulator for mobile robotics in 2 dimensions. The complement of the projects is the Gazebo as an outdoor simulator [10].

Gazebo was motivated by the spread of robotic vehicles in outdoor applications. Although the Stage is suitable for simulating interactions between indoor robots, there is a need for a simulator for modeling an outdoor environment where modeling of sensor feedback can be realistically modeled. That is why Gazebo has been developed to reproduce the dynamic environment with which a robot can meet in an outdoor environment. Based on the principles set by Player and Stage, Gazebo is open source. Thus, the Gazebo community is still active today to continuously develop the environment. This is an efficient, scalable and simple software environment that can be used for education as well.

[1]http://iridia.ulb.ac.be/supp/IridiaSupp2017-008/index.html

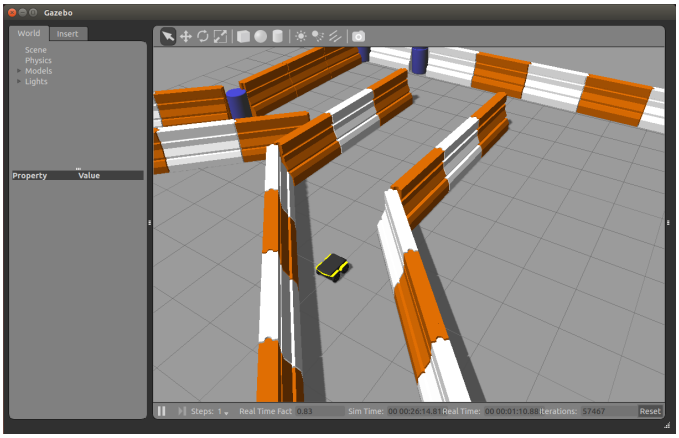


Figure 2. Gazebo[2]

V-REP (Virtual Robot Experimentation Platform) [11] (the main user interface can be seen in Fig. 3) is a versatile and scalable general purpose simulation framework. This is a portable and flexible framework that allows direct control of different control techniques. By reducing the complexity of the simulation model, simulations and simulation models have become more accessible to a much more comfortable community. Further improvements include built-in functionalities, as well as a variety of programming options [2].

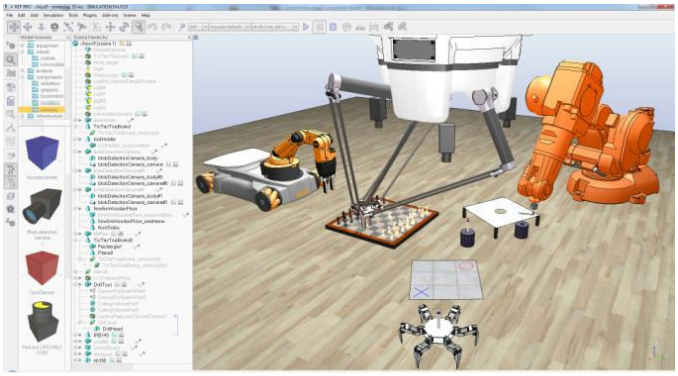


Figure 3. V-REP[3]

Main advantages of the three simulation environments based on speed, model detail and available features [12]:

- **V-REP** is very rich in features, includes editor for scenes and models, and many built-in models. It is available for educational purposes for free.
- **ARGoS** is a special option for swarm robotics.
- **Gazebo** is an intermediate solution between V-REP and ARGoS.

V-REP is the most complex of them. However, it offers a number of useful features, so it is an excellent choice for detailed and good looking simulations running on high performance machines. There are several types of physical engines available, a comprehensive model library is available, the user can change the environment while simulating the environment, video recording is available and there is an excellent documentation. V-REP offers excellent optimization options to simplify the model's network, which means that computing resource requirements are relatively easy to reduce without using third-party 3D modeling software. In addition, V-REP automatically generates new threads for multiple CPU core so that it can reach the full CPU resource if needed [12].

Based on these advantages the paper discusses the application of V-REP in the necxt sections.

III. V-REP

V-REP is easy to learn, supported by tutorial projects. The main simulation loop is a simple Lua script ("main script") [13]. The user can choose from the following programming techniques:

- Embedded scripts (Lua)
- Add-ons
- Plug-ins
- Remote API clients
- ROS nodes [14]
- BlueZero nodes

Controllers can be written in C/C++, Python, Java, Lua, Matlab or Octave.

The V-REP simulation contains several objects that are assembled in a tree-like hierarchy. The following types are supported:

- Joints
  Link two or more elements. They can be in force mode or inverse kinematics mode.
- Shapes
  V-REP uses meshes of triangles for visualisation and rigid body simulation.
- Proximity sensors [15]
- Vision sensors
  Integrated filtering and image processing features allow complex filters to be created.
- Force sensors
- Graphs
- Cameras
- Lights
  Directly affect cameras and vision sensors.
- Paths
  They allow definition of complex spatial movements, such as for controlling a welding robot to move along a certain track.
- Dummies
- Mills
  Customizable convex volume elements that simulate cutting operations such as milling, laser cutting, etc.

Objects of the scene are seldom used by themselves, they work with other objects. V-REP offers a number of computing

[2]http://www.clearpathrobotics.com/assets/guides/jackal/_images/gazebo-jackal-race.png

[3]https://ai2-s2-public.s3.amazonaws.com/figures/2017-08-08/b127edeb5bb407920e3fc8315f574bdf7351758b/1-Figure1-1.png

modules that can work directly on one or more objects. The following are the most important computing modules:

- Kinematics module
  Performs forward and inverse kinematics calculations for branched or closed mechanisms as well [16].
- Dynamics module
  Performs rigid body dynamics via Bullet Physics Library [17], Open Dynamics Engine [18], Vortex Dynamics [19] or Newton Dynamics [20].
- Collision detection module [21]
- Mesh-mesh distance calculation module
- Path/motion planning module [22]

IV. Education

Practical experience is always a pivotal point in university studies. To raise a real industrial environment within the walls of the university is not only expensive but also can be dangerous or impossible (depending on the university's capabilities, possibilities) [23]. It is not avoidable but it can be well prepared and reduced to the lowest possible extent by the use of simulation environments like V-REP. Besides the classroom studies the digitalised environment can be used for home work, blended learning and e-learning purposes. It is actually the "digital twin" (based on Siemens terminology) of the represented industrial environment. In Fig. 4 a classroom modell of a sorting machine can be seen, while in Fig. 5 the V-REP modell can perform the same tasks.

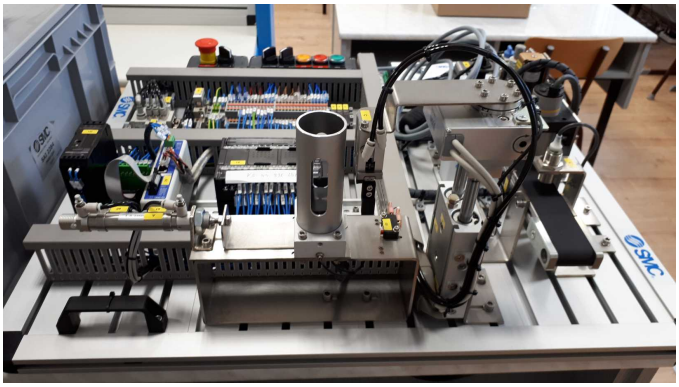


Figure 4. Sorting machine

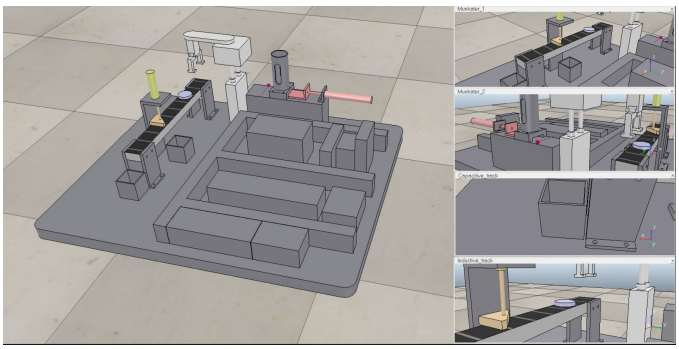


Figure 5. Simulation of sorting machine

The simulation environment amongst others gives the opportunity of education of the following subjects:

- robotics
- automation
- programming
- process planning
- control theory
- logistics [24]

V. Industrial Application

V-REP provides a good starting point also for industrial research. Obuda University as a part of a consortium takes part in the development of an automated forklift fleet based on automated guided vehicles (AGVs). The main goals of the project:

- Raise automated intralogistics systems to a higher level of service by maximizing added value.
- Development of vehicles and navigation solutions suitable for both indoor and outdoor application.
- Comprehensive integration of the AGV system into the corporate IT environment: from ERP to onboard management.

The tasks of Obuda University in the consortium are the following:

- Development of vision based navigational strategies.
  Support for positioning.
  Identification and avoidance of obstacles, security features.
- Optimal sensorfusion in positioning.
  Selection of parallel sensor systems.
  Development of an optimal filtering algorithm.
  Support for outdoor and indoor navigation capabilities.

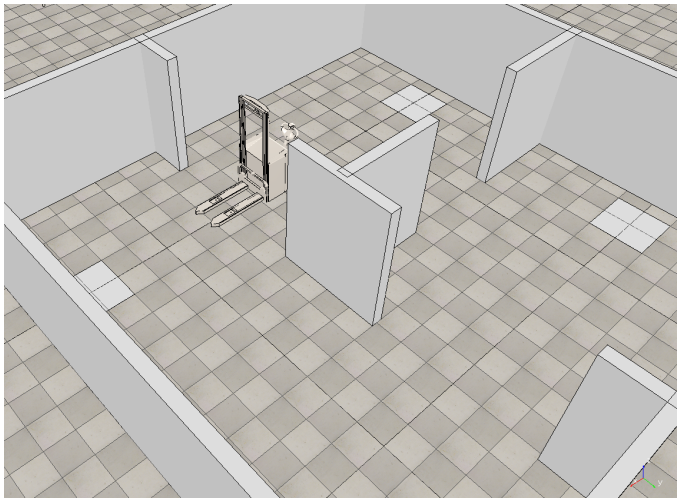


Figure 6. Layout for AGV simulation

VI. Conclusion

V-REP provides a wide range of possible applications and therefore it is suitable for univerity studies and research activities:

- fast algorithm development
- system verification
- rapid prototyping
- safety double-checking
- remote monitoring systems
- robotics related training and education
- hardware control
- factory automation simulation

Several projects have been developed both for educational and industrial purposes.

VII. Acknowledgement

The author wish to thank the support to the Arconic Foundation.

References

[1] J. Cselényi, B. Illés, and L. Rádai, "Control strategy of the palette movements on the storing-levels in block-storing multi-level car parking systems," in *Miskolcer Gespräche 2003, Die neuesten Ergebnisse auf dem Gebiet Fördertechnik und Logistik"*, 2003, pp. 169–174.

[2] E. Rohmer, S. P. N. Singh, and M. Freese, "V-REP: a Versatile and Scalable Robot Simulation Framework," in *Proc. of The International Conference on Intelligent Robots and Systems (IROS)*, 2013.

[3] G. Györök and B. Beszédes, "Fault-tolerant software solutions in micro-controller based systems," in *AIS 2017 - 12th International Symposium on Applied Informatics and Related Areas*, 2017, pp. 7–12.

[4] H. Hirukawa, F. Kanehiro, and S. Kajita, "OpenHRP: Open Architecture Humanoid Robotics Platform," in *Robotics Research*, R. A. Jarvis and A. Zelinsky, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, vol. 6, pp. 99–112.

[5] O. Michel, "Cyberbotics Ltd. Webots TM : Professional Mobile Robot Simulation," *Int. Journal of Advanced Robotic Systems*, vol. 1, pp. 39–42, 2004.

[6] C. Pinciroli, V. Trianni, R. OGrady, G. Pini, A. Brutschy, M. Brambilla, N. Mathews, E. Ferrante, G. Di Caro, F. Ducatelle, M. Birattari, L. M. Gambardella, and M. Dorigo, "ARGoS: a modular, parallel, multi-engine simulator for multi-robot systems," *Swarm Intelligence*, vol. 6, no. 4, pp. 271–295, Dec. 2012. [Online]. Available: http://link.springer.com/10.1007/s11721-012-0072-5

[7] R. T. Vaughan, B. P. Gerkey, and A. Howard, "On device abstractions for portable, reusable robot code," in *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*, vol. 3, Oct. 2003, pp. 2421–2427 vol.3.

[8] B. P. Gerkey, R. T. Vaughan, and A. Howard, "The Player/Stage Project: Tools for Multi-Robot and Distributed Sensor Systems," in *In Proceedings of the 11th International Conference on Advanced Robotics*, 2003, pp. 317–323.

[9] B. P. Gerkey, R. T. Vaughan, K. Stoy, A. Howard, G. S. Sukhatme, and M. J. Mataric, "Most valuable player: a robot device server for distributed control," in *Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No.01CH37180)*, vol. 3, 2001, pp. 1226–1231 vol.3.

[10] N. Koenig and A. Howard, "Design and use paradigms for gazebo, an open-source multi-robot simulator," vol. 3. IEEE, 2004, pp. 2149–2154. [Online]. Available: http://ieeexplore.ieee.org/document/1389727/

[11] "Coppelia Robotics V-REP: Create. Compose. Simulate. Any Robot. http://www.coppeliarobotics.com/." [Online]. Available: http://www.coppeliarobotics.com/

[12] "V-REP, Gazebo or ARGoS? A robot simulators comparison :: LENKASPACE ::." [Online]. Available: http://lenkaspace.net/tutorials/programming/robotSimulatorsComparison

[13] "The Programming Language Lua." [Online]. Available: https://www.lua.org/

[14] M. Quigley, B. Gerkey, K. Conley, J. Faust, T. Foote, J. Leibs, E. Berger, R. Wheeler, and A. Ng, "ROS: an open-source Robot Operating System," p. 6.

[15] F. Marc, O. Fumio, and M. Nobuto, "(11) Collision Detection, Distance Calculation and Proximity Sensor Simulation using Oriented Bounding Box Trees(Sensing and Data Fusion,Session: MA1-A)," 2004.

[16] C. W. Wampler, "Manipulator Inverse Kinematic Solutions Based on Vector Formulations and Damped Least-Squares Methods," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 16, no. 1, pp. 93–101, Jan. 1986.

[17] "Bullet Real-Time Physics Simulation | Home of Bullet and PyBullet: physics simulation for games, visual effects, robotics and reinforcement learning." [Online]. Available: https://pybullet.org/wordpress/

[18] "Open Dynamics Engine - home." [Online]. Available: http://www.ode.org/

[19] "Simulators and Simulation Software." [Online]. Available: https://www.cm-labs.com/

[20] "Newton Dynamics About Newton." [Online]. Available: http://newtondynamics.com/forum/newton.php

[21] S. Gottschalk, M. C. Lin, and D. Manocha, "OBBTree: a hierarchical structure for rapid interference detection." ACM Press, 1996, pp. 171–180. [Online]. Available: http://portal.acm.org/citation.cfm?doid=237170.237244

[22] J. Kuffner and S. LaValle, "RRT-connect: An efficient approach to single-query path planning," vol. 2. IEEE, 2000, pp. 995–1001. [Online]. Available: http://ieeexplore.ieee.org/document/844730/

[23] G. Györök and B. Beszédes, "Adaptive optocoupler degradation compensation in isolated feedback loops," in *IEEE 12th International Symposium on Applied Computational Intelligence and Informatics (SACI 2018)*, 2018, pp. 167–172.

[24] J. Cselényi, B. Illés, and L. Rádai, "Tömbtárolásos automatizált gépkocsiparkoló rendszerek tárolási szintjén történő mozgások irányítási stratégiájának kiegészítése a rendszer blokkolódása esetén," in *Doktoranduszok Fóruma : Miskolci Egyetem Gépészmérnöki Kar Szekciókiadványa*, 2003, pp. 212–218.

Parameter analysis of disc fitting method

Gábor Nagy*
*Óbuda University, Alba Regia Technical Faculty, Institute of Geoinformatics
nagy.gabor@amk.uni-obuda.hu

***Abstract*—The Disc Fitting method is an point cloud processing algorithm which can calculate the elevation of the ground surface in a position. This result depends on two parameters: an radius (R) and an quantile (q), because the Fitting Disc method selects the points of the cloud that are nearest the examined position that R, and fit a plane (a disk) to this area whose three sectors contain q quantile of the points under the plane. This article describes some study about the accuracy of the Disc Fitting method with different R and q parameters.**

***Index Terms*—point cloud, LiDAR, Digital Elevation Model**

I. Introduction

The airborne LiDAR technology is a very productive method of spatial data acquisition. The created point cloud is a huge data set which contains millions (or some cases billions) of points and provides a very detailed representation of the terrain, but each point means a small bit of the information: the laser beam reflected from something in these points.

The elevation models describes the terrain surface. The most of the points of a LiDAR point cloud are located over the ground surface on different natural and artificial objects, for example trees and bushes. There are varied processing method to create digital elevation models, for example [3], [9] . This article studies the Disc Fitting method [7].

II. The Disc Fitting method

The Disc Fitting method fits a plane to a point cloud and a horizontal position so, that the points located nearest than an R radius to the position are selected, and this circle is divided to three sectors. The q portion (for example 50% in $q = 0.5$) of the points are under the fitted plane (the part of this plane in the R radius circle is a disk) in each sector.

All sectors have an control point in the center of these sectors (the horizontal coordinates are fixed), and the elevation of the plane in these points are defined the plane. The parameters of the plane can be estimated by an iterative algorithm. The initial values are the q quantiles of the elevations of the points in the sector. The iteration steps modify the elevation of the control point of the sector so, that the q quantile of the points will be under the plane (the two other control point do not change). These iteration steps are repeated around the three sector until the difference between the before and the after elevation of the sector's center is lower than a limit.

The three parameters of the equation of the plane (H_0, a and b in the $H = H_0 + a\Delta X + b\Delta Y$) can be determined from the three elevations of the sector's centers (and backwards). The plane has an elevation in the inspected position (H_0) and two slope (a and b) in the two horizontal axes. The Disc Fitting

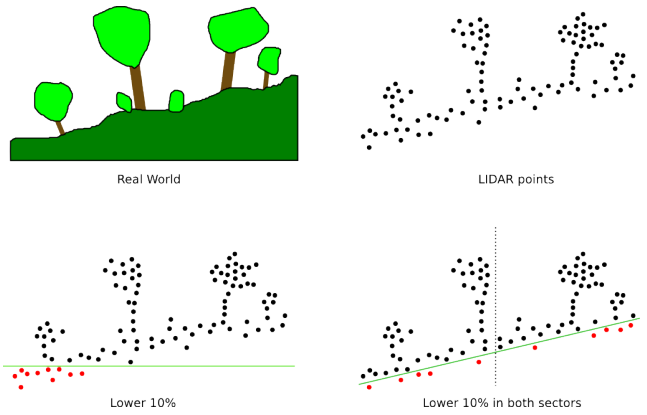


Figure 1. The principle of the method in two dimension. In this case, the area is divided to two sectors, because the regression line has two parameters. The 10% of the points ($q = 0.5$) are located under the regression line in all sectors.

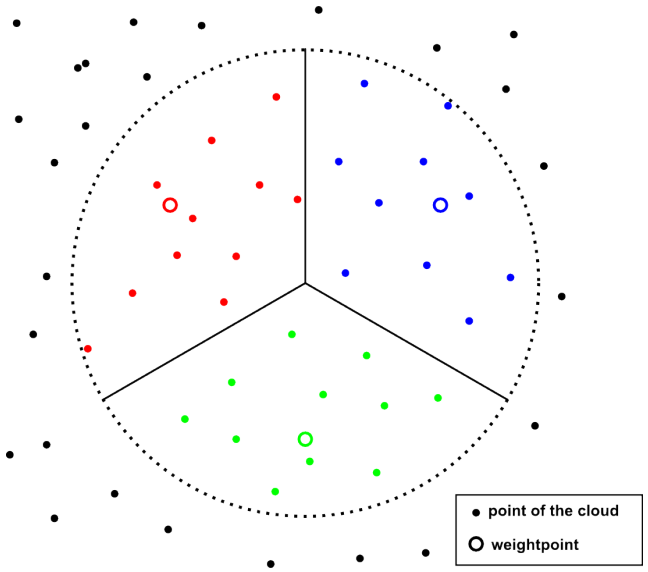


Figure 2. The sectors of the method in three dimension. Disc Fitting is applied in the R radius of an inspected horizontal positions, the regression plane is calculated from the points of this area. The circle area is divided to three sectors denoted by different colors. The q portion of the points are located under the regression line in all sectors.

method can determine an elevation in any horizontal position, and the slope (and aspect) of the terrain on this position.

Disc Fitting method can be computed in any horizontal position. If these positions are the points of a grid, an GRID type digital elevation model will be made. This calculating process is well applicable with parallel computing and distributed computing, because the calculation is independent in each grid point. The method can work without any filtering process (neither before nor after).

The mathematical background of the Disc Fitting method is an quantile linear regression. This linear regression can be generalized to any dimension. In an N dimension space, the regression has $N-1$ independent and 1 dependent variables (in case of Disc Fitting, the two independent variables are the horizontal coordinates, and the dependent variable is the elevation), and the area will be divided to $N+1$ sectors, because the N dimension plane has $N+1$ parameters. [6]

III. The effect of the parameters

Disc fitting method has two important parameters: the radius of the disc (R), and the quantile of the linear regression (q). In wooden areas a bit of the points are located on the terrain surface, therefore the method needs very low q value. The very low q value needs many points, and many points need large R (radius) value. The large R means large disk, and the terrain surface is modeled by a plane in this large area, but the terrain surface is not sure a plane in a large area, the small details of the surface may be lost.

IV. The study area

I wanted to compare the Disc Fitting method with different R and q parameters and other LiDAR point cloud processing methods. I surveyed the ground surface in 195 points. This study area located in Iszkaszentgyörgy, near Székesfehérvár. I have an airborne LiDAR point cloud from this area which was made in May of 2008.

The survey made with total station and GNSS technology in summer of 2015. The elevations of the surveyed points was accepted the elevation of the ground surface in the horizontal positions of the points. The elevation was calculated from the point cloud by different methods: the Disc Fitting methods with various parameters and the TopoSys Filled Digital Terrain Model (FDTM) which was created by the LiDAR surveying company.

V. The result by different R and q parameters

The differences between the elevation of the surveyed ground surface and the surface of the examined elevation model are calculated in each test points in each models. The distribution of the differences can be illustrated in a violin plots [1]. (Figure)

The violin plot is a good tool for illustrating the distributions, but one diagram can show only some R and q parameter-pairs. I created other plots for showing various R and q pairs. Each parameters generated by geometric series. The first element of the R sequence is $a = 1\,m$ and the common ratio

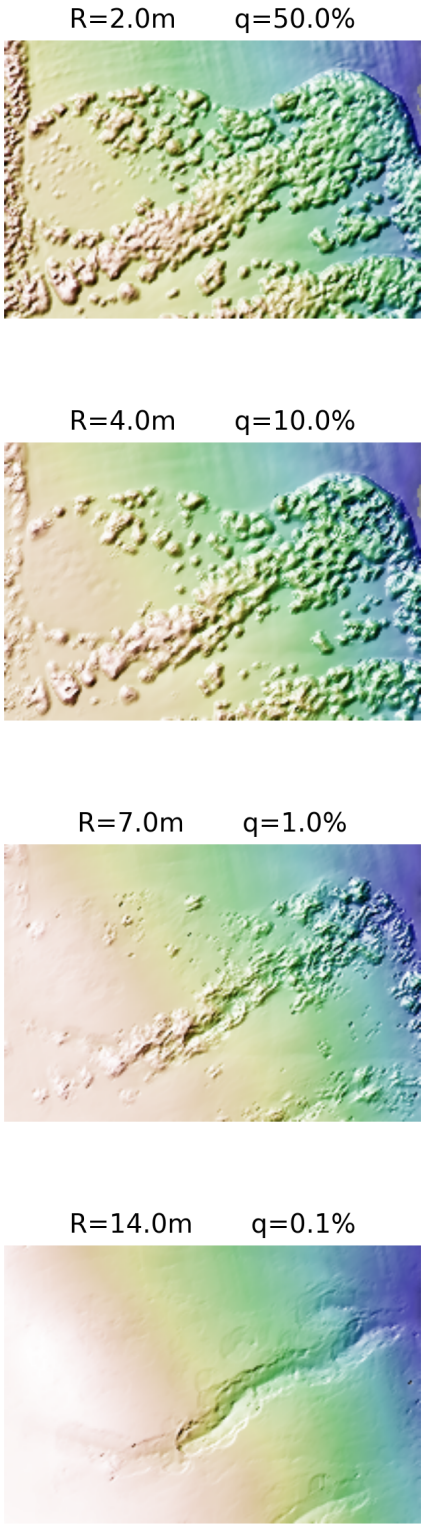


Figure 3. The processed surfaces with different R and q values. The $R = 2,\ 4,\ 7,\ 14\,m$ and $q = 0.5,\ 0.1,\ 0.01,\ 0.001$ (from the top to the bottom).

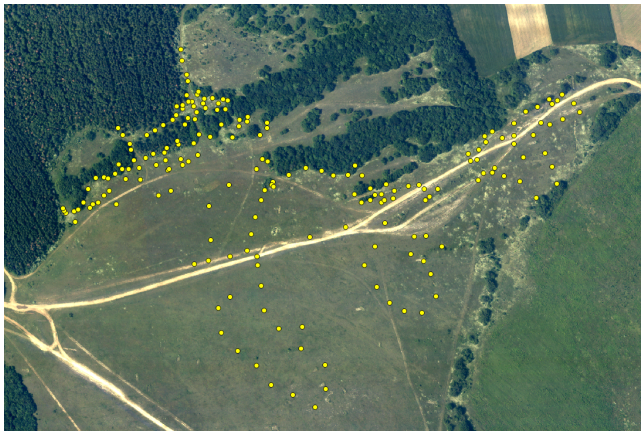


Figure 4. The surveyed points in the test area.

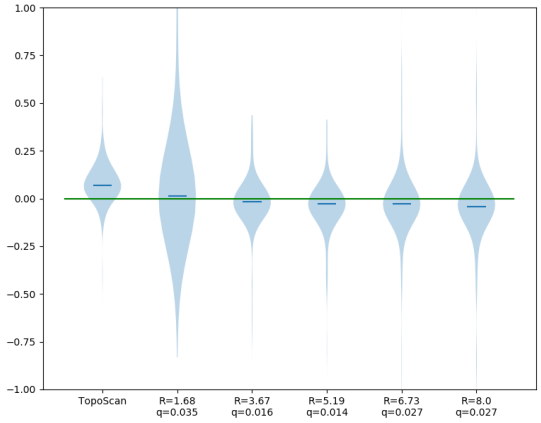


Figure 5. The distribution of the differences with some R and q values. (The first graph is the TopoSys FDTM.) This violin plot made by Matplotlib [2], [5].

is $r = \sqrt[8]{2}$. The both of first element and common ratio are $a = r = 0.875$ in the q sequence.

A Python [8], [4] program calculate the average of absolute values of differences in each R and q pair. These values are shown by a color scale in the Figure.

VI. The thickness of the point cloud

The thickness of the point cloud is defined by the difference of the elevation with two different quantiles. In this article the thickness of the point cloud is defined as the Disc Fitting elevations with $q = 0.875$ and $q = 0.118$. The $R = 4\,m$ in both of these cases.

The test area may be divided to two about the same size set: the thin areas where the thickness is less than $0.4\,m$, and the thick areas where the thickness is greater than $0.4\,m$. The plot of the Figure 6. can be created separately in the thin and thick areas (Figure 7). The optimal parameter pairs are different the thin and the thick areas.

The Figure 5. also can be created separately. The difference of the distributions can be seen in the Figure 8.

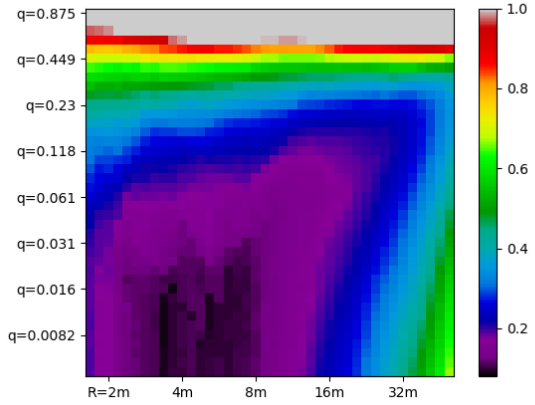
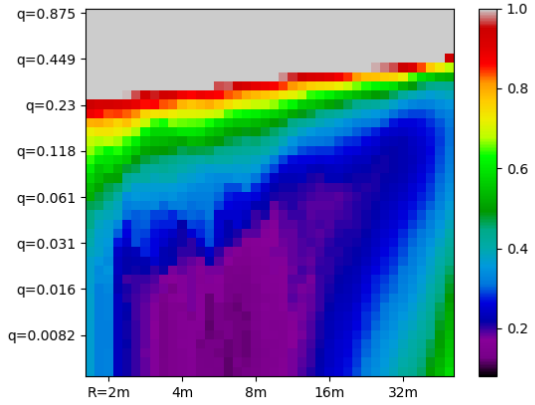


Figure 6. Averages of the absolute values of the differences in different R and q values. The plot made by Matplotlib [2], [5].



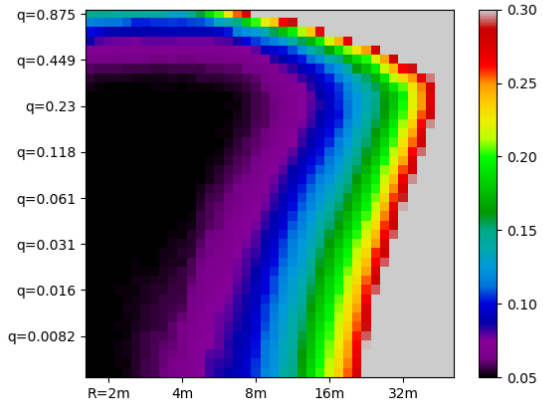
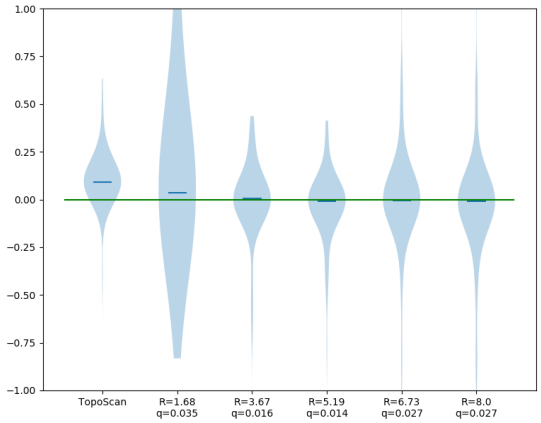


Figure 7. Averages of the absolute values of the differences in different R and q values in the thick (up) and the thin (down) areas. The plot made by Matplotlib [2], [5].



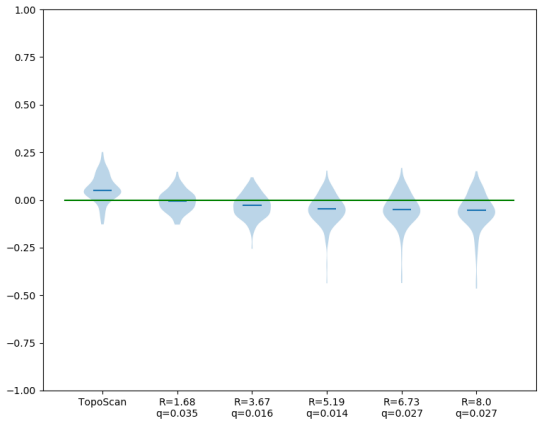


Figure 8. The distribution of the differences with some R and q values in the thick (up) and the thin (down) areas. (The first graphs are the TopoSys FDTM.) This violin plots made by Matplotlib [2], [5].

VII. Using variable R and q parameters

As we have seen before, different parameters (R and q) given good results in different point cloud thickness. A two-step method can be applied: the firs step calculates the thickness (from two Disc Fitting), and the second step calculates the elevation of the terrain surface with parameters depended by the thickness.

In my research I use a simple method, but this simple method provide good results. The thickness is calculated between $q = 0.875$ and $q = 0.118$ with $R = 4\,m$. In the thin areas (where the thickness is less than 40 centimeters) the $R = 1.68\,m$ and $q = 0.035$. In the thick areas (where the thickness is greater than 40 centimeters) the $R = 5.19$ and $q = 0.14$.

The violin plots are created with this thickness depended parameters in the Figure 9.

VIII. Conclusion

The Disc Fitting method can create good quality digital elevation models from LiDAR point clouds, if we use suitable

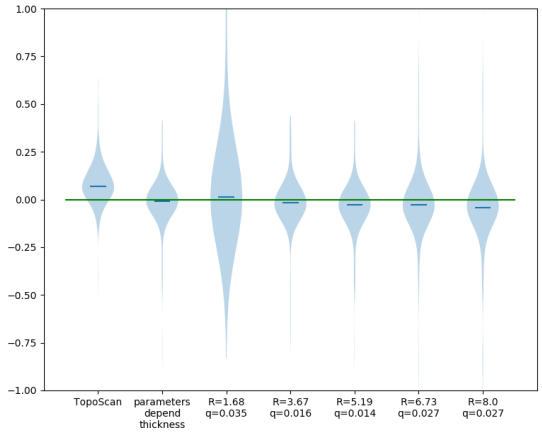


Figure 9. The distribution of the differences with some R and q values. The second graph is the distribution of the thickness depended method. (The first graph is the TopoSys FDTM.) This violin plot made by Matplotlib [2], [5].

R and q parameters. An area may be combined different parts, different parameters may be suitable in these parts.

This topic need more research in further test areas and sample data. More test points can provide more sophisticated models for the ideal R and q values.

References

[1] Jerry L Hintze and Ray D Nelson. Violin plots: a box plot-density trace synergism. *The American Statistician*, 52(2):181–184, 1998.

[2] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9(3):90–95, 2007.

[3] Xuelian Meng, Nate Currit, and Kaiguang Zhao. Ground filtering algorithms for airborne LiDAR data: A review of critical issues. *Remote Sensing*, 2(3):833–860, 2010.

[4] K Jarrod Millman and Michael Aivazis. Python for scientists and engineers. *Computing in Science & Engineering*, 13(2):9–12, 2011.

[5] G. Nagy. Matplotlib in the geospatial analyzes. In Gábor Tamás Orosz, editor, *12th International Symposium on Applied Informatics and Related Areas (AIS 2017)*, pages 90–93. Óbudai Egyetem, 2017.

[6] Gábor Nagy. Sector based linear regression, a new robust method for the multiple linear regression. *Acta Cybernetica*, 23(4):1017–1038, 2018.

[7] Gábor Nagy, Tamás Jancsó, and Chongcheng Chen. The fitting disc method, a new robust algorithm of the point cloud processing. *ACTA POLYTECHNICA HUNGARICA*, 14(6):59–73, 2017.

[8] Travis E Oliphant. Python for scientific computing. *Computing in Science & Engineering*, 9(3), 2007.

[9] George Sithole and George Vosselman. Experimental comparison of filter algorithms for bare-Earth extraction from airborne laser scanning point clouds. *{ISPRS} Journal of Photogrammetry and Remote Sensing*, 59(1-2):85–101, 2004. Advanced Techniques for Analysis of Geo-spatial Data.

Automated Car Parking Facilities in the Age of Autonomous and Connected Vehicles

Levente Rádai PhD
Alba Regia Technical Faculty, Óbuda University, Székesfehérvár, Hungary
radai.levente@amk.uni-obuda.hu

Abstract: **The automated car parking facilities can give an environmental-friendly and time-friendly solution for parking problems in busy European old town centers. But the the fear of the dependency of automation and operating failure of active actuating systems prevent its spreading over the world. Although the fear and the complexity of these systems, we may to apply these facilities to provide parking opportunities in city centrals. But is this demand changed by the newly appeared electrical and/or autonomous and/or connected vehicles, which can park themselves? Can they co-operate to create more favorable semi-automated facilities? How can we utilize the smart mobile devices to get a better parking experience? The article collects the solution options for smart cities to utilize the nowadays available automated car parking technologies.**

Keywords**: Smart cities, Parking facilities, Parking experience, Semi-automated car parking facilities, Autonomous cars, Connected cars**

I. INTRODUCTION

Nowadays, the automated, mechanized car parking system have got a good challenger in the field of comfortable car parking solutions. One of the biggest benefits of robotic car parking facilities is that the driver needs only to locate a free receiving cabin nearest to the destination, drive in and can immediately leave the car, and the system take care about the load, storage and retrieval of the parked car.

The autonomous, self-driving cars can challenge this feature, because we do not need to locate and drive into a parking facility, but we can stop with the car at the destination, jump out – which is not so difficult in jammed city traffic – and order the autonomous car to find a parking place somewhere nearest.

In Europe, this feature of autonomous cars is disabled because of the European highway code does not allow an autonomous car to run without a responsible driver. But we can utilize this self-driving ability to establish a cheaper automated, mechanized car parking facilities, which use the self-driving car motion ability to replace one horizontal motion axis in the 3-dimensional car storage system.

Figure 1. Generic cuboid types of multi-storey robotic car parking facilities

For this, the paper analyzes the generic type of multy-storey robotic car parking facilities, optimization tasks, cost-functions and related general control strategies and develop optimal solutions for system simplification by controlling self-driving car in parking facilities. At last, but not least, some general security and social issues will be discussed related to remote control ability of self-driving or semi-self-driving cars with distance holding and lane assist.

II. OVERVIEW OF GENERIC TYPES OF MULTY-STOREY ROBOTIC PARKING FACILITIES

Generic multi-storey robotic parking systems are 3-dimensional cuboid or cylindrical storage systems served by 3-axis transport system. This paper deals primarily with cuboid shaped systems, where the transportation can be performed by high-rack loading and retrieval machines or elevators and some type of horizontal transport system in a distributed service model.

a) Standard aisle type high-rack shelf storage with automated loading and retrieval machine

In these systems the car storage is performed by high load capacity shelf storage racking system in aisle layout. Cars are placed onto car storage pallets or are "grabbed" at the tyres by self-adjusted roller arms. Automated high-rack loading and retrieval machines are carrying the pallet or type grabber units in each aisle (storage corridors).

b) Multi-storey Robotic Parking Facilities with Aisle Type Distributed Machinery

In this type of robotic parking facilities, vertical and horizontal machinery can operate separately. Elevators lift the cars on pallets or standalone to the right floor, then satellite cars transport horizontally and place the car or the pallet into the right storage place. Satellite cars can be equipped with "pallet-grabber" or "tire-grabber".

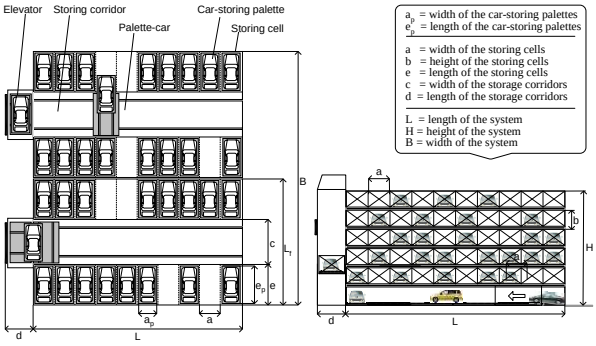


Figure 2. Multi-storey robotic parking facilities with aisle type distributed machinery: vertical elevators and horizontal satellite cars [2]

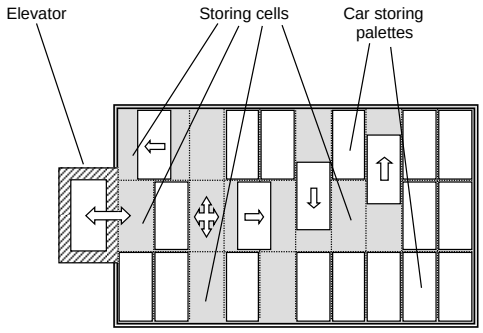


Figure 3. Multi-storey robotic parking facilities with block-storage type motion-store matrix machinery

In case of using pallets, empty pallet-handling should be performed.

c) Robotic Parking Facilities with block storage multi-storey machinery

In the block storage robotic parking systems, cars are placed onto car storage palettes. These palettes are elevated with the car up to the right floor, then is loaded out from the elevator to the storage floor. On the storage floor "transport&store" cells are installed in more rows and columns side by side near to each other. Each "transport&store" cell is mounted with a 2-axis driving mechanism, which can transport the palettes in longitudinal and transverse direction.

The advantage of this storage system is, that nearly 100% of utilization can be reached and only a few free cells are required to the movements.

Disadvantages, all the cells should be mechanized, which results higher installation and operating costs and the control of this system is much complex.

III. REQUIRED OPTIMUM CRITERIA IN OPERATING ROBOTIC CAR PARKING FACILITIES

In operating car parking systems, the key performance indicator is the waiting time of the driver for park-in and park-out or in other words drive-in and drive-out. The immediate drive-in requires only the availability of free receiving cabin. The occupation depends on the time duration of park-in process by the driver. After the driver left the cabin, the machinery must free up the receiving cabin by transport the car into the storage system. On the other hand, the procedure of the retrieval in case of park-out request is longer, because the car should be retrieved

Figure 4. Receiving cabin in robotic car parking facility using car storage pallet

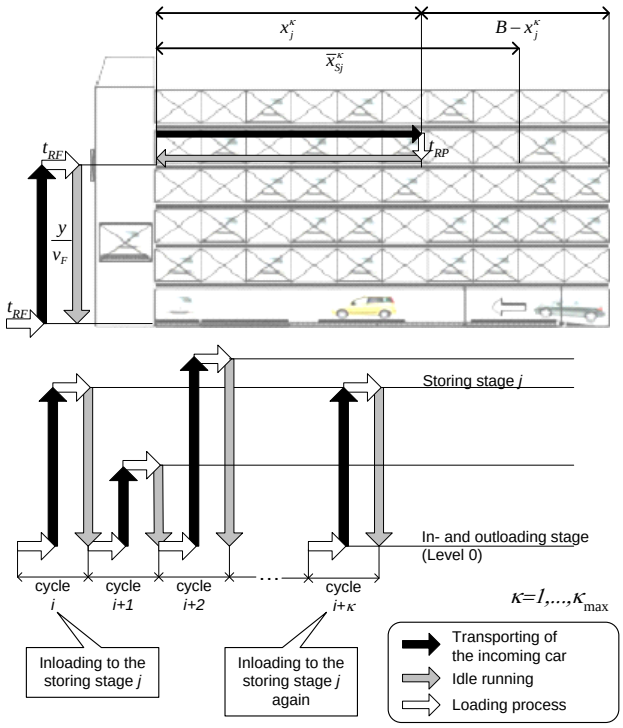


Figure 5. Periodical transport cycle in multi-storey robotic car parking facilities served by distributed machinery

from the storage system. So, the driver must wait. Therefore, the mayor task is to minimize the driver's average waiting and service time by the right service strategy. On the other hand, the operating costs of the entire facility should be at an adequate level. Related to this, the right balance of utilization of the driving units should be performed regarding to the planned maintenance cycles [4].

IV. OPERATING STRATEGIES AND OPTIMIZATION TASKS

Based on the optimum criteria, the operating strategies should support the waiting queue optimization, the transportation of the incoming cars and the retrieval of the outgoing cars.

Basic strategy is to free up the receiving cabin as soon as the machinery can arrive to catch the incoming car from the receiving cabin.

Additional strategy is the right sorting order in the storage system to make the loading and retrieval time shorter. And at last, but not least, equipment utilization and usage optimization should be performed for the right maintenance processes. At this point, there are different strategies for loading and retrieval machine, for multi-elevator systems and multi-satellite car systems and for block storage floors.

V. SIMPLIFIED ROBOTIC SOLUTIONS FOR CAR PARKING AUTOMATION

There can be found simplified robotic solutions, where classic Ramp Access Garages (RAG) are equipped with remote controlled Autonomous Guided Vehicle (AVG) running on concrete surface altogether with other driver-controlled cars. This option is a cost-efficient enhancement

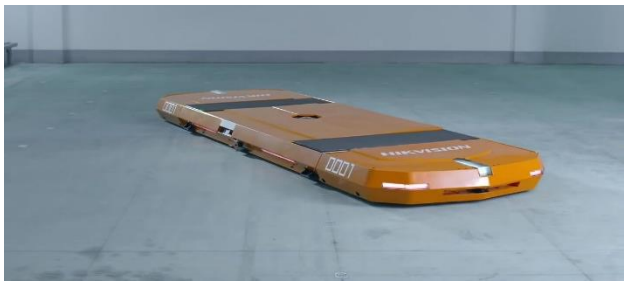


Figure 6. Remote controlled car lifter Autonomous Guided Vehicle (AGV) working on classic Ramp Access Garage (RAG) concrete surface

of parking facility giving a hybrid solution both for drive-in parking and a comfortable luxury parking.

The system is very scalable, because the number of the car lifter AGV can be adjusted to the demand for automated park-in-park-out.

Disadvantages are that the volume utilization optimization is not available, and the lifting method of the car is not gentle with the car chassis.

VI. Role of Self-Driving Cars in Parking Facility Automatization

The role of autonomous, self-driving cars is that it can replace one motion axis and the required drive system in the robotic parking system. For this, remote control performed by the robotic parking system.

In non-automated parking systems, this motion axis can be utilized to reach block storage in row format, which can raise the volume utilization up to 25% in first approach. Disadvantages are the high sorting demand.

In semi-automated parking systems, where the drive-in ramps are replaced by elevators, this utilization growth can be extended by additional 25% in first approach.

In fully automated systems one motion axis can be replaced, therefore pallet installation, pallet or tire grabber unit and tire grabber or pallet handling operations can be left, which factors result a much simpler system with less installation and operation costs.

But it requires the self- or semi-self-driving ability of the car, and the availability to control the car by the facility control system. On other hand, the complexity and requirements for right operation and reliability of control system is much higher, and industrial standards required to reach the car manufacturers.

Figure 7. Loading cars by tire-grabber in robotic parking system (without applying car storage pallet) [5]

VII. Optimal Solutions

Optimal solutions are the standard high-rack shelf storage with automated loading and retrieval machine, block storage with elevators, and multi-storey parking systems served by Elevators/Loading-Retrieval Machine and Satellite cars.

At all solutions, the self-driving ability is applied to enter and leave the lifting or transport unit of the machinery, so the car can roll on or roll off the surface of lifting unit loading and retrieval machine or the satellite car. On any error, car transporter AGV must called to place the stuck self-driving car and driver must be notified.

This easy operation takes only approximately 5-6 meters at up to maximum 1 m/s speed controlled directly by the car onboard self- or semi-self-driving capability.

Steering by the car is not necessary because of the linear motion, therefore semi-self-driving capability is enough using distance sensor and lane-assistant, like e.g. Toyota Auris Safety Package.

VIII. Issues to be Solved

Externally controlled parking mode should be developed using standardized protocols and interfaces. This standard should contain the specification of the external command process and protocol.

The externally controlled movement should be enabled by the driver at leaving the car. If the eternally controlled movement is started it should be reported to the driver and to the owner in real-time by the car itself. For this, wireless coverage should be provided, which is not easy in steel racking system.

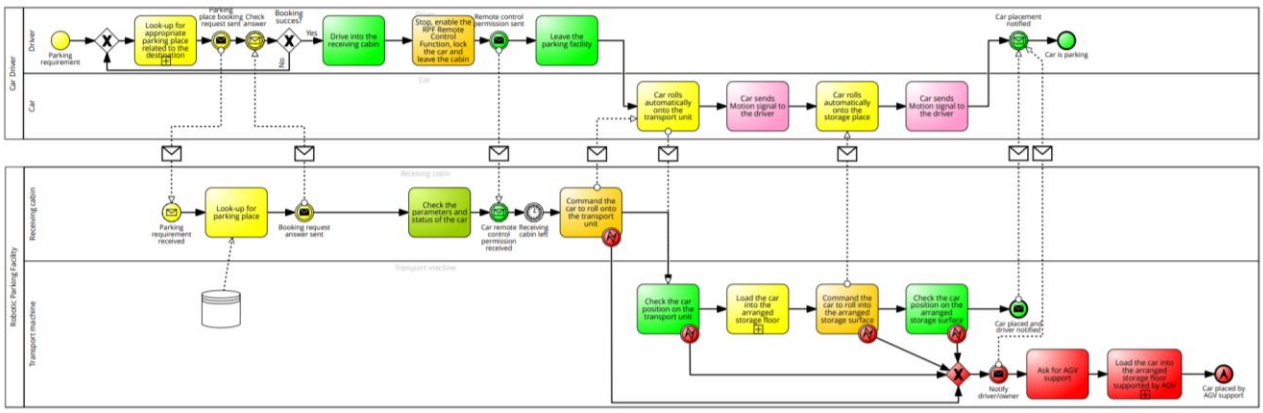


Figure 8. Drive-in and loading process in collaborative robotic parking facilities

Token-based check signal should be applied to check the car's real position. If the signal lost, an automatic call of police should be triggered.

Because of the European highway code does not allow driverless drive of cars, exception for robotic facilities with unmanned storage area should be legalized.

IX. SUMMARY

The article presents a solution for automated, mechanized car parking facilities utilizing the self-driving ability of autonomous cars, which use the self-driving car motion ability to replace one horizontal motion axis in the 3-dimensional automated car storage system. The research overviews the generic types of robotic parking facilities and the basic control strategies and optimization tasks, selection of optimal system solutions by controlling self-driving car in parking facilities, and several security and owner issues to be solved. Optimal roll-on-roll-off solutions are:

- standard high-rack shelf storage system with automated loading and retrieval machine in each aisle,
- multi-storey car storage systems served vertically by elevators or loading-retrieval machine and horizontally by satellite cars,

Further benefit is that this function can avoid the entire system from overload in case of high utilization by splitting the motion requirement between the parking system and the car driving system. And it decreases the facility installation costs by replacing one motion axis and its drive system.

On the other hand, a lot of related issues should be solved such as legal, social safety and property safety issues. The further research needs to solve the industrial standardizing of the communication and regulation solutions and additionally, trust of driver and owner society should be gained by reliable, useful pilot projects.

REFERENCES

[1] D. Monahan (2015) „De-Mystifying Automated Parking Structures", Walker Parking Consultants

[2] L. Rádai, J. Cselényi, B. Illés (2006) "Models and Methods for Determining the Inloading Performance in Small and Medium Size Automated Car Park Systems Served by Elevators and Palette-cars", *MicroCAD 2006 International Scientific Conference*, University of Miskolc, Hungary, 2006

[3] "No need to put in a garage by the driver, a system will automatically carry the robot to the parking space fully automatically", in: *Gigazine*, 28th January, 2018.

[4] K. Széll, G. Manhertz (2016). "Development of a web-based training module in robotics", *International Symposium on Small-scale Intelligent Manufacturing Systems (SIMS2016)*, Narvik, Norvégia, 21-24.06.2016, pp. 1-6., ISBN:978-1-5090-3074-3, 2016.

[5] "Multiparker solutions", Otto Wöhr Autoparksysteme GmbH, 2014

[6] K. Széll, A. Czmerk, P. Korondi (2014). "Friction with Hysteresis Loop Modeled by Tensor Product", *AUTOMATIKA 55* (4) pp. 463-473., 2014.

Dangerous Driving Remote Sensing based on Mobile Society

Levente Rádai PhD
Alba Regia Technical Faculty Óbuda University, Székesfehérvár, Hungary
radai.levente@amk.uni-obuda.hu

Abstract: **I am disappointed driving a car in Szekesfehervar. It is good to see, that there are a lot of cars, more than 500 cars pro 1000 citizens, which shows, that people have got money, so the economy in Szekesfehervar is really prospering. But most people do not care of traffic rules, actual traffic situations and optimal traffic. Some of them are only impolite and aggressive, but lot of them cannot really drive. Can we help on this situation? Can we detect dangerous drivers not harming the GDPR? Could we warn the conscious drivers for dangerous drivers in the close traffic environment? Seeing an alarming ambulance car, my daughter said once, that we cannot know, when an accident will happen. I reflected as alighting, that of course, we can! The article presents an initiative mobile application based on android society.**

Keywords**: Smart cities, Smart traffic, Driving society, Driver responsibility, Connected cars**

I. INTRODUCTION

My experience is, that the driving moral is terrible in Hungary and in Eurasia east from Hungary. There are terrible drivers driving too fast and/or ignoring the driving rules and/or too slow and/or dummy. Furthermore, there are impolite and dummy drivers who do not or do not want to recognize the optimization opportunity in a congested concurrent traffic situation, e.g. traffic situation in Figure 1.

As Google drives the traffic indication on Google Maps based on Android Mobile Device anonym position data submission, there could be setup a social dangerous driving detection appliance named to CDDDA or RDDDA. Other major example for anonym driver tracking works in Waze Mobile Navigation Application.

A sensing and warning system, as a part of an Intelligent Transportation Systems (ITS) could support not only the conscious drivers and pedestrians by warning them for reckless drivers in the certain traffic zone, but it could help to broadcast the appearance of alarmed ambulance, firefighter, police or disaster management vehicles to avoid the collision with personal vehicles and with other alarmed vehicles (Fig. 1.). In Hungary, approximately 500 ambulance car accidents happen annually. The internet is full of articles and videos about careless drivers ignoring the siren on the public roads.

Moreover, these applications can support the Usage Based Insurance (UBI) [1], or can "*gauge subjectively gauging recklessness from a passenger's perspective"* [2].

An appliance like this can use the following data collecting technologies for rating the driver's behavior and broadcast real-time warning information:

- connected mobile devices,
- dedicated IoT devices,
- connected cars.

II. METHODS FOR DETECTING DANGEROUS DRIVING BEHAVIOUR

A comprehensive literature review shows that sensing the dangerous driving is a general topic in many countries and in many fields.

The major device for dangerous driving behavior detection method is the smartphone. Typical solution is the yaw angle detection, which can be based on low accuracy accelerometers and gyroscopes. The benefit of this application is, that there are no requirements about the orientation of the device, but it should be developed different algorithm for each of behavior, moreover there could be interferences with normal motion features. [1]

The difference in developed systems are the applied sensor-fusions and the driver behavior and risk analysis methods and algorithms. The gyroscopes are not able to detect the reckless driving directly, there are a lot of circumstances to correct the measurement in real time.

Figure 1. Accident of two official ambulance car in a rural major city in Hungary

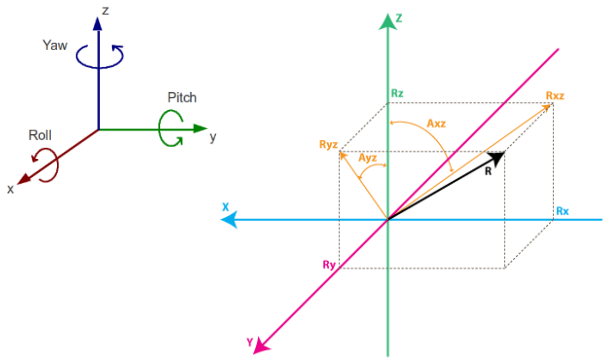


Figure 2. Yaw, pitch and rolling detection with gyroscopes [3]

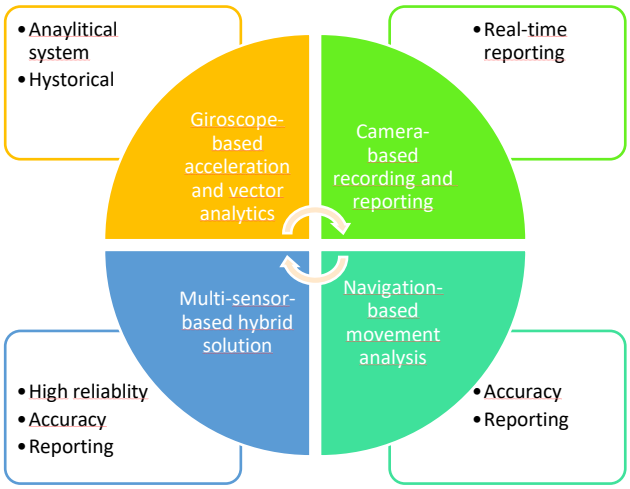


Figure 3. Short analysis of detection possibilities of Reckless and dangerous drivers

On the other hand, the capability of mobile devices hosting these embedded sensors is limited in energy consumption and computing.

Most of the application targets only the device onboarded in the car. The available devices are motion detection camera, gyroscopes, and GNSS-based navigation devices. These devices can detect the reckless driving or sense it by several analytical process, but there are accuracy limitations. [4] There are camera-based solutions aiming the detection of drivers influenced by the alcohol or drugs [5] or recording speeding and ignoring red-light. And there are wearable equipments development for detection sleepy drivers. [6]

III. CONNECTED APPLICATIONS

Ref. 5. examined the recent results and applications not only the sensor-based, but network-like applications mobile device for supporting intelligent traffic management systems.

The key of these intelligent traffic management applications is the connectivity, which can provide the warning information to the transport society integrated into widely used applications running during driving or transportation.

There are several navigation applications, which provides traffic information on the map used for navigation. In next subchapters, there is a short comparison about these applications used primarily and widely in Hungary.

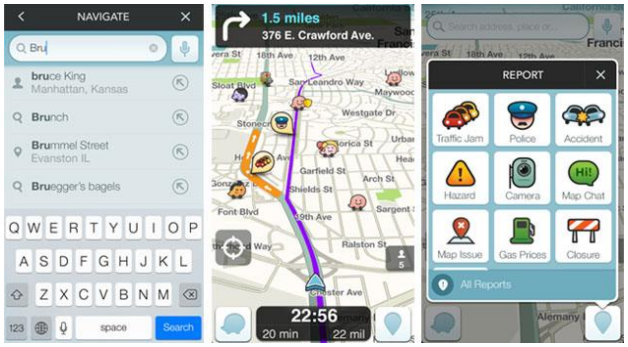


Figure 4. Navigation Screen and Event-reporting UI in Waze Mobile Navigation Application

a) Waze application

Waze application is a generally used mobile navigation application extended by "Wazer" society coming from event reporting and registration opportunity.

Benefits are its popularity, and the anonym and identified end-point tracking, which can be used for traffic flow tracking and tracing. Based on the traffic flow information, enhanced route rating is provided, which helps to optimize the route based on estimated travelling time. Moreover, it can warn the drivers for approaching at a traffic jam.

Re-routing is also available function, but the system does not sense the growing or changing traffic load, (or only does not take it into the calculation), therefore, mostly at the start of traffic jams and at the end, it may propose longer, but stuck bypasses. This gives a worse route than the original stuck route. And typical periodical traffic load cannot be analyzed.

Moreover, based on user-friendly event reporting with minimized interaction requirement, it can warn the drivers for accidents, broken down cars and reported road controls, or building sites, road constructions and closures, as well. Some of these events can be managed centrally by Waze moderator.

b) Traffi Hunter Mobile Application

The Traffi Hunter is a road control forecasting application to prevent drivers from penalty about especially speeding, road-toll and truck weight control. It is officially not forbidden but helps the drivers to get away with penalty of highway code fouls. It is equipped with user-friendly reporting user interface for real-time submission of speed control and road controls. Moreover, there is a database about fixed speed traps and published monthly road control schedules.

c) Google-Maps Navigation Application

Google-Maps is the general map application of Google therefore, it is basic application for Android-based mobile devices. But it is available on desktop application. It collects end-point devices, as well, so traffic flow tracking and route rating is also available. Based on this, the typical traffic load can be checked for a certain period of week, but only on desktop. (Fig.5.)

The other benefit of Google-Maps is its worldwide usage, which give the application developers and service providers to insert into the maps their locations as a Point-of-Interest (PoI).

Figure 5. Traffi Hunter Mobile Applcation

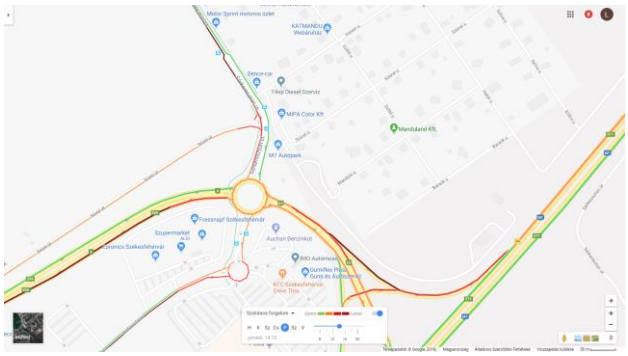


Figure 6. Periodical Average Traffic Load in Google-Maps Web Application

Moreover, other applications like site and facility structural maps, mobile parking place discovery and location- and time-based payment solutions can be mapped on Google-Maps.

IV. Structure and Functionality of Complex Warning System

As a summary, Waze and Google-Maps applications can collect and provide driver behavior information in real-time, but there are a lot of missing components:

- Driver rating based on international highway code:
  - collection driving behavior-related data,
  - conversion collected data to smart data,
  - submit smart to the central system,
  - collection of pedestrian density data.
- Warning system using driver rating data:
  - central system warns drivers/pedestrians in certain traffic environment for occasional dangerous drivers,
  - warn drivers to high density of pedestrians/traffic jams.
- Collecting additional traffic safety intensive data:
  - tracking the position of alarmed ambulance, firefighter, police and military cars,
  - mash-up of real-time weather station reports.
  - (Tracking the weather conditions from car onboard systems, which requires connected cars.)

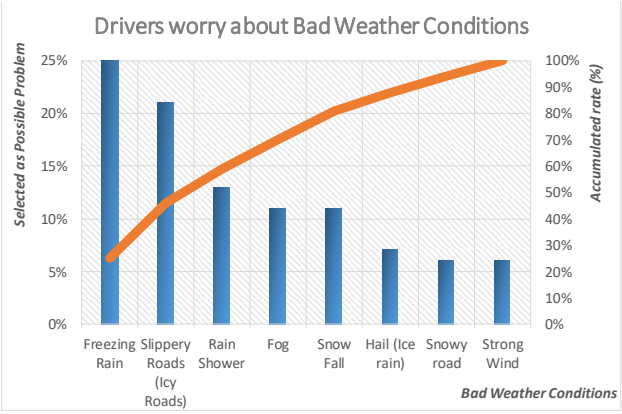


Figure 7. Most Frequent Bad Weather Conditions Driver's Worrying about

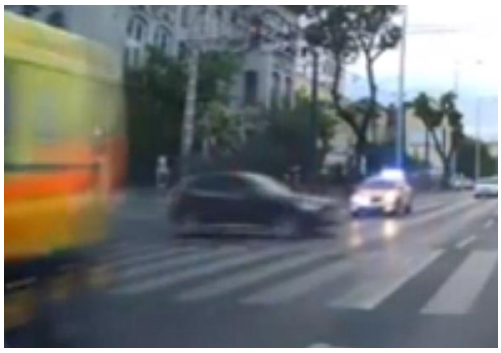


Figure 8. Inattentive driver crossing on Ambulance Convoy hitting nearly both ambulance vehicle

- Warning system using additional data:
  - approaching of alarmed ambulance, firefighter, police or military vehicles,
  - approaching to bad weather conditions, especially freezing rain, slippery roads, hail, fog, rain shower as a short survey shows .

The tracking of alarmed vehicles has got high importance, because not only the firefighters and sometimes police vehicles move in convoy, but there are more and more cases, when an ambulance vehicle is followed by a medical car in much more cases, and the transporting society is not prepared for this in many cases. Most of the drivers seeing an ambulance car crossing a road, think that the way is free, and causing serious dangerous traffic situation. (Fig.8.)

V. Most Frequent Traffic Incidents in Székesfehérvár and in Hungarian Public Roads and Detection Possibilities

Analyzing personal traffic data, the following traffic incidents are most frequent in daily city and intercity traffic:

a) Ignoring right hand rule

At known locations of crossroads with same priority, ignoring the right-hand rule, namely passing the cross section without braking or no stop for a car approaching in from right, can be detected. This event can be weighted if there is a car approaching from right, the weight of ignoring is bigger. If there is detectable braking by the vehicle on priority road, the weight of ignoring it is the biggest.

An additional weighting factor should be added to rate the behavior at crossings with objects in eclipse or hard to see. This feature should be applied at any situation where eclipsed objects are involved into the crossing situation.

b) Ignoring STOP sign protecting priority road

At known locations of Stop signs, stopping the car can be detected. Ignoring the Stop sign can be weighted by the certain traffic flow, namely if there is a car approaching to the road crossing on the priority road, the weight of ignoring is bigger.

If there is detectable braking by the vehicle on priority road, the weight of ignoring it is the biggest.

c) Ignoring priority against a car approaching on primary road

Similar to ignoring stop plate, at known locations of Priority is mandatory signs, approaching vehicle on priority road can be detected. In this case the score is smaller, than at ignoring Stop sign.

d) Ignoring priority in roundabout type street crossings

It is similar to the previous cases, but score can be weighted by approaching speed and the presence of other road users.

e) Ignoring Pedestrian Crossing

Basically, pedestrian crossing zones must be approached by car with 30 km/h in Hungary. In this case, the score is depending on the approaching speed and the presence of pedestrians and its number.

In this case, weighting factor should be added to rate the behavior at crossings with objects in eclipse or hard to see or near to nursery, elementary and high schools and playgrounds and parks.

f) Ignoring the right direction in one-way road

The detection can be based on movement direction and triggered by turning or leaving the one-way road on the entrance side, but the weighting should be based on distance and on stuck cars moving into the right direction.

g) Ignoring alarmed Ambulance, Firefighter and Police cars

Detection is based on motion vector, and scoring based on speed, interference and crossing the route of alarmed vehicle.

h) Dangerous take-over at frontal oncoming car

Detection is based on motion vector, and scoring based on speed, and interference with frontal oncoming car (distance measurement). GPS accuracy could be enough depending on absolute speed.

i) Crossing closing line

If closing line data is available on the road section, the closing line crossing detection is available. It is hard to detect because of average GPS accuracy, can be corrected based on comparison positions of previous and following vehicles proceeding into the same direction on the road.

j) Ignoring speed limit especially at bad weather conditions (Absolute speeding)

This behavior can be detected by GPS signals and can be weighted by the certain bad weather condition and based on the presence of other road users who can be endangered.

k) Ignoring bad weather conditions especially freezing rain, slippery roads, hail, fog, rain shower (Relative speeding).

This behavior can be detected primarily relative to the other vehicles on the same road. Rating can be based on speed, following distance and takeover attemptions.

Figure 9. Slow down! My dad works here, too!

l) Road construction and building sites

Continuously refreshed database can determine the locations of these sites, and the prescribed traffic limitations such as speed limits, lane closures, lane shifts and other restrictions, which can be updated and revised by car onboard camera-based traffic plate recognition.

Driver's behavior scoring can be weighted by the endangering level of the road users and site workers.

As it previously mentioned, additional weighting factor should be applied at rating the driver's behavior at crossings signed with objects in eclipse or hard to see or near to nursery, elementary and high schools and playgrounds and parks. This information can be added, managed and revised centrally and by road users.

VI. ISSUES ABOUT DRIVER'S PRIVACY AND RELIABILITY AND RESEARCH AND DEVELOPMENT TASKS

In case of a driver's behavior scoring system, drivers do not want to be identified themselves, and get a negative ranking. Therefore, the system should identify the driver's anonym, without name, if we want the system to make popular in road user society. At this point, we lose the opportunity to use the system to support the work of judicial experts and insurance companies in case of evidentiary issues.

So, one of the development tasks is to develop the system to even-handed, socially-trusted, consciously usable and useful warning system. In this case, the drivers want to participate in the communal usage of the system and position in the ranking in best position meeting with its conscious thinking about driving safety.

Technological requirements are above the accuracy and real-time information providing are the minimal battery consumption, CPU usage and mobile network traffic load by the application, and the integration with popular navigation systems.

VII. SUMMARY

The paper summarizes the foundations of driver behavior analysis and traffic warning system. The analysis of possible traffic incidents, its detection possibilities and issues about driver's privacy shows that there are two major tasks to develop:

- reliable, energy- and communication-efficient, real-time detection of inattentive, reckless behavior of drivers,

- even-handed driver's behavior rating system weighted by the presence of affected road users, the danger level of the situation and the complexity of road crossing site.

The success of such a system is dependent on trust of road users especially running seamless and at low power and network consumption in background on mobile devices of non-driving road users, pedestrians, bicycle, roller, Segway and motor riders, and participants in public transport. In case of bicycle, roller, Segway and motor riders, a useful extension of the warning system can be the headset- or smart-watch-based warning signals.

REFERENCES

[1] Fu Li, Hai Zhang, Huan Che and Xiaochen Qiu, "Dangerous driving behavior detection using smartphone sensors," 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), Rio de Janeiro, 2016, pp. 1902-1907., doi: 10.1109/ITSC.2016.7795864

[2] A. S. Zeeman and M. J. Booysen, "Combining speed and acceleration to detect reckless driving in the informal public transport industry," *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013*), The Hague, 2013, pp. 756-761. doi: 10.1109/ITSC.2013.6728322

[3] How to Estimate Yaw Angle from Tri-Axis Accelerometer and Gyroscope, Robotics Stack Exchange, Robotics Beta, 5th October 2014.

[4] Stratis Kanarachos, Stavros-Richard G. Christopoulos, Alexander Chroneos, "Smartphones as an integrated platform for monitoring driver behaviour: The role of sensor fusion and connectivity", *Transportation Research Part C: Emerging Technologies*, Volume 95, 2018, Pages 867-882, ISSN 0968-090X, https://doi.org/10.1016/j.trc.2018.03.023.

[5] Prasad, Vidur. (2017). Traffic Camera Dangerous Driver Detection (TCD3). *Electronic Imaging*. 2017. 67-72. 10.2352/ISSN.2470-1173.2017.9.IRIACV-270.

[6] Fürész, J., L. (2017) Éberség figyelő rendszer tervezése és megvalósítása, Student Scientific Conference, Óbuda University, 2017

[7] K. Széll, G. Manhertz (2016). "Development of a web-based training module in robotics", *International Symposium on Small-scale Intelligent Manufacturing Systems (SIMS2016)*, Narvik, Norvégia, 21-24.06.2016, pp. 1-6., ISBN:978-1-5090-3074-3, 2016.

[8] K. Széll, A. Czmerk, P. Korondi (2014). "Friction with Hysteresis Loop Modeled by Tensor Product", *AUTOMATIKA 55* (4) pp. 463-473., 2014.

[9] Huang, Yong-Ren. (2011). Real time dangerous driving status detection. *Proceedings - 4th International Congress on Image and Signal Processing, CISP 2011*. 1. 10.1109/CISP.2011.6099989

New Ways in Application of Robotic Car-Storage Systems using Mobile Parking Application

Levente Rádai PhD
Alba Regia Technical Faculty Óbuda University, Székesfehérvár, Hungary
radai.levente@amk.uni-obuda.hu

Abstract: **Car parking facilities are basically storages for temporarily not utilized vehicles. But APF can be defined as an interface between the live traffic and people who would like to travel by car. There could be different ownership samples of these cars, but if they are out of utilization they should be parked, stored somewhere, as close to the loading point as it is possible. There are some requirements in case of these facilities, mainly they should be discovered and be occupied in the shortest time and should have got high reliability to provide appropriate car safety and fast drive-out opportunity. The article shows how mobile devices can help the discovery and the use of these facilities by optimizing the park place management.**

Keywords**: Robotic Car-Storage Systems, Mobile Parking Application, Location-based Waiting Queue Optimization, Discrete-event simulation**

I. INTRODUCTION

The demand for parking space does not decreased in last decades. Although there are directed city development programs and projects, such as P+R parking systems, railway and tramline public transportation developments, bike and car sharing systems, self-driving cab systems, there is a continuously raising traffic load in cities. As we proposed in publications of last decade, the automated, mechanized, multi-storey car parking houses and facilities in underground, integrated into office, commercial, residential buildings and standalone high-building constructions on hole-sites can provide optimal solution for provide parking capacities. The compact design of these facilities helps to find the feasible solutions in the certain traffic environment.

We should think in several problems and hazard appearing at the application of these facilities from both driver and facility side:

Figure 2. Typical Robotic Parking Facilities in steel or concrete structure [1]

- overload of machinery and service systems in the facility and in it's traffic environment,
- lamer user/driver (lack of experience, skill),
- possible malfunctions of control system and machinery,
- potential power-off failure,
- incompatible design of the machinery, which can result car damage.

And dangers from "malicious economy" and groups:

- hacking of control system for car-theft and cash-leaks,
- malicious tracking and following people,
- trial for causing asset damage, casualties and catastrophe
- observing and tracking certain persons.

Parallelly, drivers generally have got less time, they are busy and therefore they are much inpatient to look after an appropriate parking place. They want parking possibilities in much less time, which is impossible in busy city centers. Parallelly there are more and more cars.

Therefore, parking facilities have to provide much higher performance and much better cycle time in car storage and retrieval. With the help of information provided the mobile applications, we can make a better and reliable service for parking cars.

The drive-in points appropriate to a certain destination can be founded by driver's navigation device more reliable and the drive-out procedure can be performed much faster and by less effort. Regarding to the topic, there are more mobile apps supporting these issues partially [6], but comprehensive optimization is not in general use.

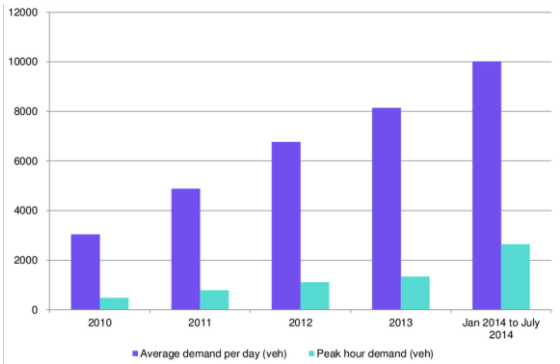


Figure 1. Typical demand growth for parking place in China [2]

Figure 3. Entrance and exit of Robotic Parking Facilities [3]

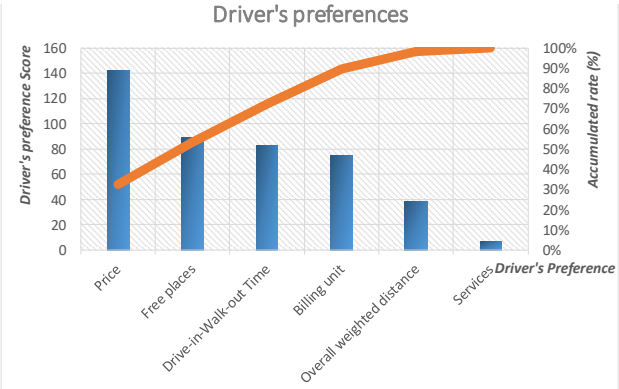


Figure 4. Driver's preferences for decision support of parking place selection in a certain multi-target traffic environment

A comprehensive literature review shows that the robotic parking has started to come in general use in North America and Far-East [4]. In Western-Europe, as well as in Hungary, the largest spread of RPF was before the last global economic crisis, but in last decade, the growth has been stopped.

Although in Hungary, the last reference to robotic parking system was in the last decade, it is reasonable to support the application of these systems.

II. STANDARD PROCESS OF DRIVE-IN AND DRIVE-OUT IN ROBOTIC CAR-STORAGE FACILITIES

In "classic" robotic parking facilities, driver looks after the appropriate entrance door of the RPF. If there are more, than one entrance door, the control system of RPF should direct the driver to the appropriate door

In this "classic" case, driver gets assistance only in selecting the right door, but must locate the parking facility by itself. If the facility is full, driver must drive to another parking place. There are some mobile parking applications help the driver to find the free parking facilities.

Driving into the entrance cabin, driver must stop, fix, lock and leave the car. There cannot be any livings in the car. Afterwards the car will be checked about its size. If the loading is available, the driver gets the parking card or RFId tag and can leave the entrance cabin.

The system starts to load the car into the storage facility.

After driver arrives to get its car, the system identifies it, driver can pay the bill, and the system starts the retrieving of the car.

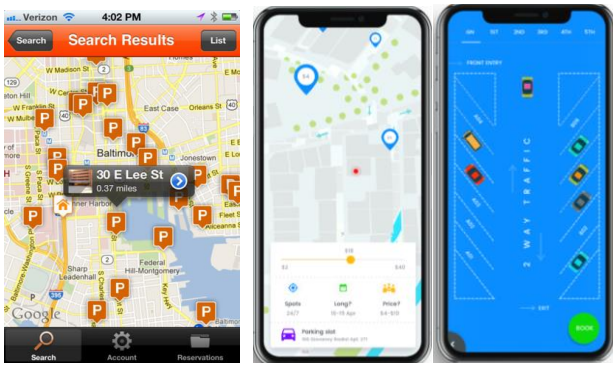


Figure 5. Some example for Mobile Parking Applications [6], [7]

III. PARKING PROCESS BASED ON CO-OPERATIVE MOBILE APPLICATION

The following chapter details the process of driving experience using integrated mobile parking application

Driver selects the real destination.

Mobile Navigation Application locates the nearest parking lot/facilities.

Mobile Parking Application (MPA) collects the parameters of potential parking destinations and creates a priority list based on multi-parameter optimization weighted by driver's preference:

- free parking place in the facility,
- distance between real destination and parking facility entrance weighted by indoor-outdoor distance rate.
- time requirement based on drive-in-walk-out and walk-in-drive-out time,
- price and billing unit,
- category and additional services, e.g. car wash service, etc.

Driver selects the appropriate parking place from the list or let the MPA to select it based on the user (driver's) preferences.

Based on this selection, MPA sends booking request to the information system of the selected parking facility (PFIS) and waits for approval.

	A	B	C	D	E	F	G
1	**Personas Overall Thoughts** P1:An official group of people who use car everyday. P2: People who take rented cars occasionally. P3: Cab drivers for whom car parking has become a major problem. P4: Heavy vehicle drivers and owners.	P1	P2	P3	P4	P5	P6
2	Had the participant got the meaning of the word "Car parking"?						
3	Whether the needs of the user fulfilled?						
4	Is the user willing to pay for the app?						
5	Has the user got how the navigation was organized?						
6	Were the users getting parking slot whenever they need?						
7	How the slot booking is managed for the peak hours?						
8	Was the app easily accessible for elderly people?						
9	Would the app be given top priority?						
10	What changes the app has made in the user's work life?						

Figure 6. Example research results for user preferences about Mobile Parking Applications [7]

FPIS approves or dismiss the booking request:

- if approved, the mobile navigation system selects the modified destination and recalculate the route.
- if the selected FPIS refuses the booking request, the driver or the MPA must make the booking on the next parking place in the list, and step back to the approval request.

IV. Car Retireving process

In automated, robotic parking facilities, the harder task is the retrieving of the car on demand by the driver. Depending on the size and layout of the facility and the performance of the machinery, it takes generally 1-10 minutes to retrieve a car from the storage system. If there are more than one car to be retrieved, this time requirement can raise rapidly.

Therefore, the driver should estimate the planned parking time to help the control system of the RPF to sort the cars by planned retrieving time to shorten the waiting time and the waiting queue of the drivers.

This method can be extended by the retrieving order function of the Mobile Parking Application. If the driver decided to park out, can send the Retrieving Demand Message (RDM) to the control system of RPF and the system can put this order into the optimal position of the waiting queue of the retrieving task.

During this period, the MPA can inform the RPF about the driver's walking position and based on this further waiting queue optimization can be performed by the FPIS.

Moreover, in multi-entrance parking facilities, the FPIS can select the optimal entrance for the driver's position and the MPA can inform the driver which entrance should be used to get back the parked car.

V. General Model of the Information System

The general model of the information system of mobile parking contains the client devices at the drivers, which can be simple Android, IOS or Windows mobile application, or a standard web application, like SAP Fiori applications on the one hand. On the other hand, we should calculate with the connected car developments, where the certain car manufacturers and the OEM suppliers of car onboard multimedia systems install this functionality on the car multimedia system. These systems can be specialized on high level, therefore standardization of data models and

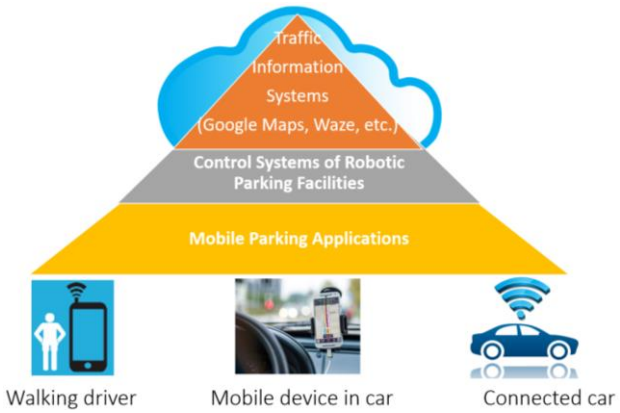


Figure 8. General Hierarchy of the Information System Mobile Parking Applications

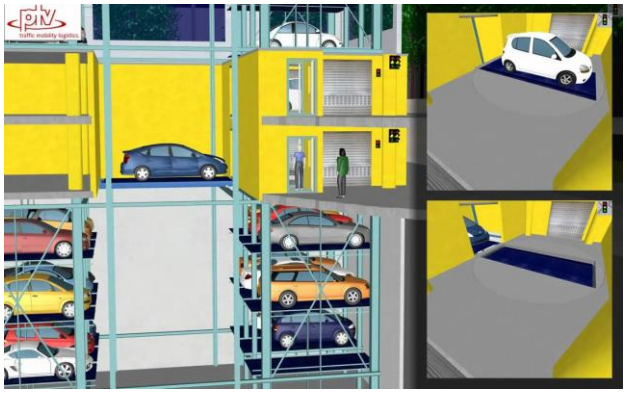


Figure 7. PTV Vissim Automated Parking Simulation [11]

communication interfaces is required, and this brakes the spread of the parking application.

VI. Performance analysis of Robotic Parking System extended with Mobile Parking Application

For examination the throughput capacity of a robotic parking system, generally we use computer-based discrete-event simulation methods. In education and research, for robotics and complex mechatronic systems, there are different simulation tools from commercial professional tools [8] to individual, specialized software developments [9], [10].

For parking facility simulation, there are examples by VisSim [11], which is primarily widely used for traffic environment simulation. We use individually developed simulation tool (Fig. 8.), which was developed primarily for research of movement algorithm of block-storage automated parking systems. [12] Based on the developed movement algorithms, we can test the performance of the RPF helped by mobile parking application.

VII. Examination of the system

Evaluations using Discrete-event simulation provides comparable results about system service performance using optimized serving strategies and movement algorithms.

These strategies are full-sorting allocation strategy as the general car sorting strategy for block-storage RPF and level

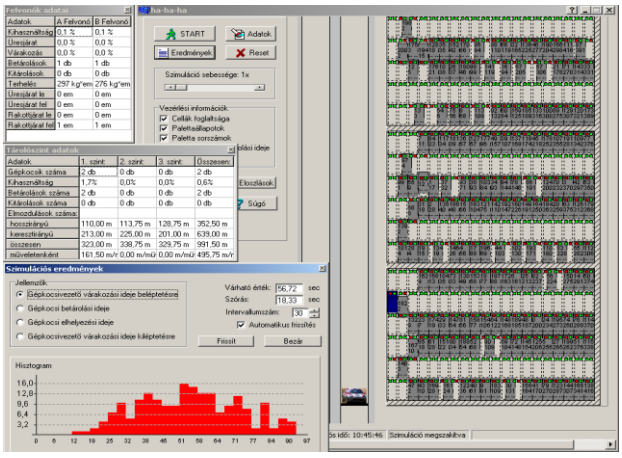


Figure 9. Discrete-event simulation evaluation of the material and control information flow of block-storage, multi-level car parking system

selection optimization based on waiting queues of incoming cars and cars to be retrieved and estimated parking time at the elevators.

As in Chapter I detailed, the distribution of the parking demand and parking periods is much more dispersed, than even before. As the proposed system helps to determine the waiting queue optimization at the retrieving, we can calculate with the best effort performance refined by possibility of occurrence of deviation from arrival time provided by the mobile app. This deviation is expected to late arrival of the driver to the drive-out zone based on traffic examination of parking houses in specific traffic environment.

On the other hand, the processing order of incoming and outgoing cars is also optimized by up to 20%, thank to know the incoming cars much earlier, than they arrive to the RPF entrance.

I. SUMMARY

The article presents and summarises a model for mobile parking application integrated into the control system of robotic parking facilities.

It presents the process structure and the general model if information system and examine the estimated performance growth in multi-level, block-storing RPF using discrete-event simulation method.

In further steps we would like to examine the performance in dependency of traffic rate with simulation methods. In additional we would like to analyse the decision-support model and develop fuzzy-based algorithms for selecting the appropriate parking place in a certain situation.

Based on the models and methods to be elaborated, the software design of co-operative mobil parking application and the standardized interface and middleware components can be developed.

REFERENCES

[1] "Create Space for Your Project", "RPS 1000 automated parking garage in steel or concrete structure", Robotic Parking Systems, https://www.roboticparking.com/robotic_parking_cad_files.htm, 2018

[2] Lv, Jie & Guo, Jianmin & Zhang, Yibin. (2017). "Research on determination of the scale of parking space on High Speed Rail Station, using East Ji'nan Station as an example." *IOP Conference Series: Earth and Environmental Science.* 81. 012132. 10.1088/1755-1315/81/1/012132.

[3] "More comfort with Wöhr Car Parking Systems", *Parking Network,* 20th August, 2014

[4] "Global Smart Parking Systems Market to Grow at 22% CAGR

[5] L. Trego, "Automated parking systems market to grow at CAGR of 11.42%", *Autonomous Vehicle Technology*, BNP Media, Troy, Michigan, December 20, 2017

[6] J. Shueh (2015). "5 Mobile Parking Apps to Help Cities Ease Traffic Congestion", *Government Technology*, e. Republic, Inc., 6th February, 2015.

[7] J. Vino (2017) "UI/UX Case study: My Car parking", *Muzli - Design Inspiration*, 7th December, 2017, Url: https://medium.muz.li/ui-ux-case-my-car-parking-1ca00506245f

[8] K. Széll, (2018) "Robotics Simulation Environment in an Educational Institution According to Demands of Industry 4.0", *AIS2018 13th International Symposium on Applied Informatics and Related Areas*, 8th November, 2018, Óbuda University, Alba Regia Technical Faculty, Székesfehérvár, Hungary

[9] K. Széll, G. Manhertz (2016). "Development of a web-based training module in robotics", *International Symposium on Small-scale Intelligent Manufacturing Systems (SIMS2016)*, Narvik, Norvégia, 21-24.06.2016, pp. 1-6., ISBN:978-1-5090-3074-3, 2016.

[10] K. Széll, A. Czmerk, P. Korondi (2014). "Friction with Hysteresis Loop Modeled by Tensor Product", *AUTOMATIKA 55* (4) pp. 463-473., 2014.

[11] "PTV Vissim: Automated Parking Simulation", PVT Group, 4th February, 2011

[12] L. Rádai (2008) "Theoretical Foundations of Multi-level Block-storage Automated Parking Systems", PhD Thesis Work, University of Miskolc, Department of Materials Handling and Logistics

Human resources in a start-up company

About the relevancy of the human resources management and about the computer added HR processes in a start-up company

Ildikó Petőné Csuka

Alba Regia Technical Faculty of Óbuda University

csuka.ildiko@amk.uni-obuda.hu

Abstract

Some people have the opinion that HR or HR Department is not necessary at a small venture like a starting start-up company. However even the smallest companies should also invest in HR. If a strong HR function is developed by the company, employees may feel valued. Founders should explain to employees what the company does and whom it does it for, and why the company is in this business. If employees see through the processes of the company they are able to work for the company efficiently. In the interest of transparency and to share business information with all of the employees, start-up company should apply an ERP-system. In the work the author makes a suggestion to introduce S/4 SAP HANA system to see the business processes of the company. The work emphasizes the importance of processes of human resources management especially in the case of start-up companies. The author brings into focus the Personnel Development and Talent Management from the processes of human resources management which are very important recently either by companies functioning in the Hungarian economy or in other European countries. Employers shall select and hire those people who are engaged to the company's vision and if it is so, then founders and managers shall value, honour them for example with personnel development and talent management (career planning). The work shows those functions of SAP HCM (Human Capital Management) which are able to help in retention of employers hereby raising their commitment to the company.

I. STRATEGIC HR ISSUES OF START-UP COMPANIES

There are **three strategic HR issue** which start-up companies should pay attention. The success of the business depends on them:

A. Personality matters

First the HR specialists should understand the personalities and personas of the company's customers. Marketing, customer support, sales and product management plans should be designed with the focus on the company's diverse customer base.

B. the company's values and vision must be articulated for employees

Employees work better when they can make decisions on their own. If they have got the skills, knowledge and training needed in their position and they understand the purpose of their job within the organization, employees do their best. As the start-up company grows, the number of employees have been increasing and the founders must rely on a team to represent the company to partners, customers, suppliers, creditors, and investors. If the employees can not explain why the company exists, what its vision is, then the company's message will never spread. Therefore start-up founders need to spend time making sure that their employees share their vision of a better future empowering them to be effective „flag-bearers". Managers must pay attention and resources into managing their employees so that employees should feel important and valued. Through the processes of personnel development and career planning employees can feel the care and that they are important for the company.

C. requirement of transparency

Start-up founders have to be aware of that how much information they should share with staff. Employees are expected to be informed about the business processes and have all of the information they need to work effectively, but are reluctant to share confidential information [1]. Employees shouldn't be burdened with excessive communication. The application of ERP-systems are the best choice to solve the problem of information-sharing [2]. Employees and managers can come to better decisions with the use of ERP.

II. S/4 HANA SAP

One of the best ERP-system applied in a company is SAP-system. The introducing of S/4 HANA from SAP by a start-up company is a good choice to show the business processes for the employees. It is important that the employees understand the goals and the operation of the company.

SAP S/4 HANA is a business suite (**S** in the name means both: simple and suite), and one of the biggest innovation in the history of SAP. SAP S/4 HANA delivers simplifications (customer adoption, data model, user experience, decision making, business processes, and models) and innovations (Internet of Things, Big Data, business networks, and mobile-first) to help businesses to run easier in the digital economy [4]. SAP S/4 HANA provides access to all of data of the computer systems, including the Internet, and it can be available to all users.

It insures opportunities to run the companies their day-to-day business in real-time using industry best practices [5].

A company has a lot of business processes in the different functional areas (asset management, production, sales, customer service, financial accounting, management accounting (controlling), human resources management and so on). The author deals with the processes of the Human Resource Management in SAP in her work.

With the help of SAP University Alliances the latest SAP technologies can be integrated into teaching. SAP University Alliances (SAP UA) is a global curricula program in educational institutions of 113 countries. The author uses the case studies created exclusively for SAP UA global curricula with the Global Bike (GBI) data set.

III. HCM PROCESSES OF SAP:

a) Organizational Management

b) Personnel Administration

c) Recruitment

d) Personnel Development

e) Talent Management

f) Performance management

g) Personnel Controlling.

The author brings into focus the process of Personnel Development and Talent Management in her work. Personnel Development and Talent Management has been getting more and more important among the tasks of employers because of the rising shortage of labour force. Before covering of the processes of Personnel Development and Talent Management the enterprise structure in SAP ERP HCM must be made clear through the exposition of the Organizational Management.

A. Organizational Management

Organisational Management represents the structural and personnel organization of a company, consists of organizational units and shows the organizational structures and hierarchies and also an employees responsibilities. It consists of three parts:

1. Enterprise Structure
2. Personnel Structure
3. Organizational Plan

1. **Enterprise Structure**

Each employee needs to be assigned to an enterprise structure which represents formal and financial structures in the company and consists of company code, personnel area, and personnel subarea. **Personnel Area** represents a company area differentiated between personnel administrative, time management, and expensive organizational aspects. **Personnel Subarea** is the part of a personnel area and represents a company area differentiated between personnel administration, time management, and payroll accounting aspects.

2. **Personnel Structure**

Personnel structure means an employee's position within the company and it is defined by employee group and employee subgroup. **Employee Group** is an organizational unit for which personnel related regulations can be specified for examples: active employees, external employees and retirees (see Figure 1). **Employee Subgroup** is an organizational unit within the employee group, for which personnel related regulations are specified for examples: industrial employees, commercial clerks, laboratory assistants).

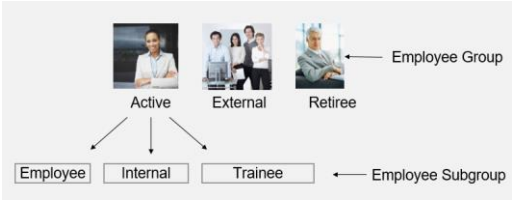


Figure 1. Employee Group and Employee Subgroup

Source: GBI tutorial, SAP UA [3]

3. **Organizational Plan (OP)**

Organizational Plan exposes in which function an employee is acting. OP is object-oriented: each element is represented by an object which can be assigned to each other. **Organizational Units** can be grouped according to functional and regional aspects. **Position** is assigned to organizational unit and filled by employee (for example: HR manager in the HR Department). **Person** are employees and fill positions. Cost Center derives from Controlling (Management Accounting) and can connected to positions or organizational units**. Job** represents the general description of tasks which an employee must perform, is assigned to position (see Figure 2).

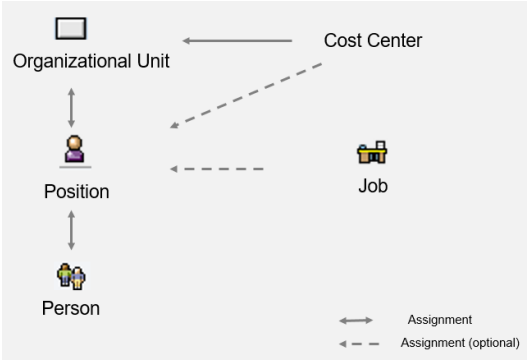


Figure 2. Organizational Plan

Source: GBI tutorial, SAP UA [3]

Thereinafter we can come to the point of Personnel Development and Talent Management.

B. Personnel Development and Talent Management

The purpose of **personnel development** is to qualify employees according to their actual and future requirements and to optimize the benefits of employees. The personnel development in a company consists of the comparison of actual and future requirements.

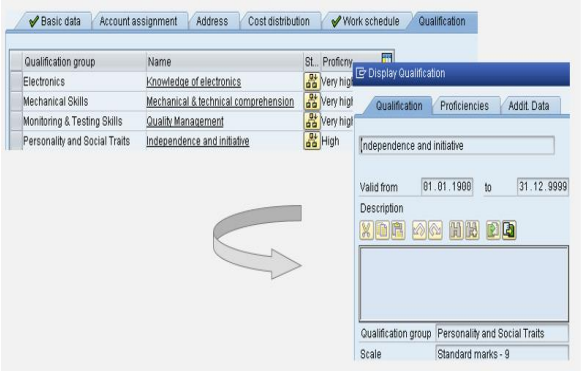


Figure 3. Requirements (qualifications, knowledge and experiences)

Source: GBI tutorial, SAP UA [3]

The process of personnel development helps the HR department with planning and administration of teaching and development actions, as well as enables to heed qualifications of employees and to reveal shortages in qualifications of employees. To show the deficits between the requirements of a position or job and the qualifications of employees the system makes a **profile matchup**.

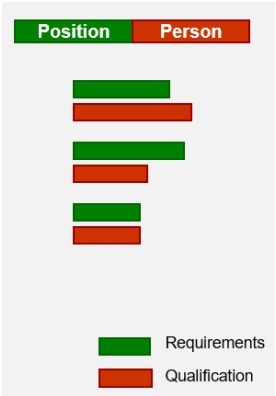


Figure 4. Profile matchup

Source: GBI tutorial, SAP UA [3]

The results of the profile matchup can be three kinds:

- both characteristics are identical,
- under-qualification: requirements are higher than the qualification ,
- over-qualification: requirements are lower than the qualification.

The personnel development system can generate development proposals and plans. Moreover, employees can be booked directly for training courses or other events (in the case of under-qualification). After a training course or a development action, acquired qualifications are absorbed to the qualification profile of an employee from the personnel development system (both are identical). In performance management, benefits and developments of employees can be evaluated.

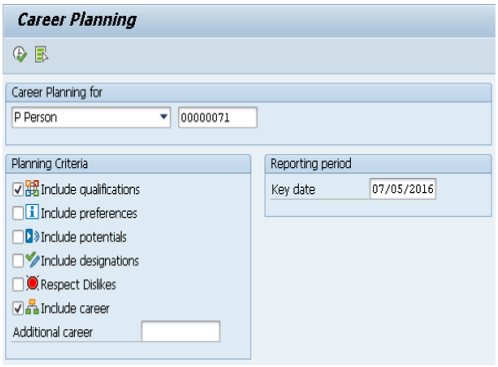


Figure 5. Career Planning

Source: GBI tutorial, SAP UA [3]

The **talent management** is concerned with career planning as well as further training of employees and is integrated in personnel planning. The talent management consists of analyses and applying of careers, individual development plans and following plans.

Positions and tasks for an employee are able to be allocated **in career planning** (Figure 5.), that the person can fulfill during his career in the company. Development actions are originated, which are important for the employee to reach a station. This is arranged when qualifications of an employee get adjusted with requirements of a station in the career. The career of employees is planned and development possibilities are made up. Individual development plans get derivate out of the career planning. **Development plans** describe a summary of further training and general training events, that an employee can use in the career, so they are developed for a special employee.

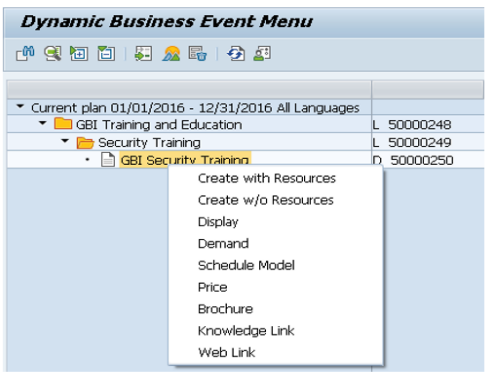


Figure 6. Creating an event

Source: GBI tutorial, SAP UA [3]

Figure 6. shows the way of creating an event. The missing certificate will be added.

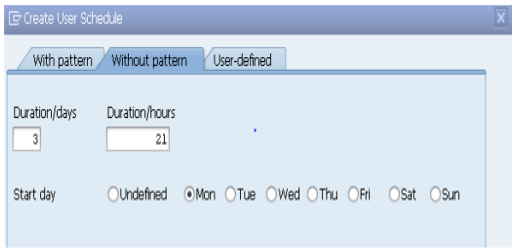


Figure 7. Creating of user schedule

Source: GBI tutorial, SAP UA [3]

After entering the Name of the event its location is able to be chosen. Before saving the event the minimum number of attendees, as the optimum and as the maximum number of attendees must be entered. In the upcoming screen *Create User Schedule* (Figure 7), the duration (days and hours in total) must be also entered in the Without pattern tab. This training (in Figure 7) will start on a Monday.

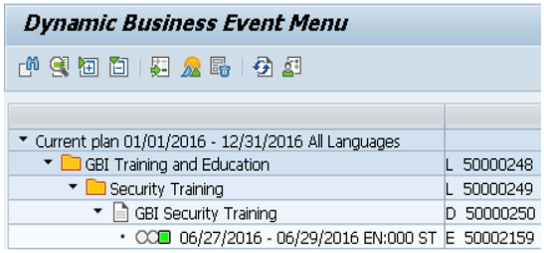


Figure 8. View of the created event

Source: GBI tutorial, SAP UA [3]

If the folder for the Security Training in the Dynamic Business Event Menu is opened, the created event is able to see. Afterwards, booking of the event come (Figure 9).

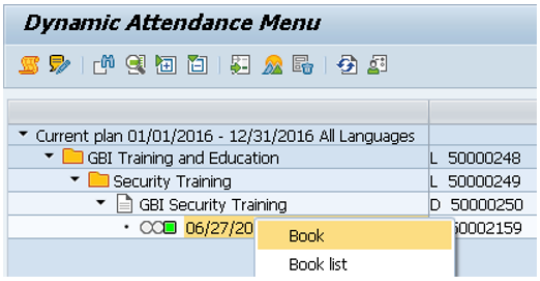


Figure 9. Booking of the event

Source: GBI tutorial, SAP UA [3]

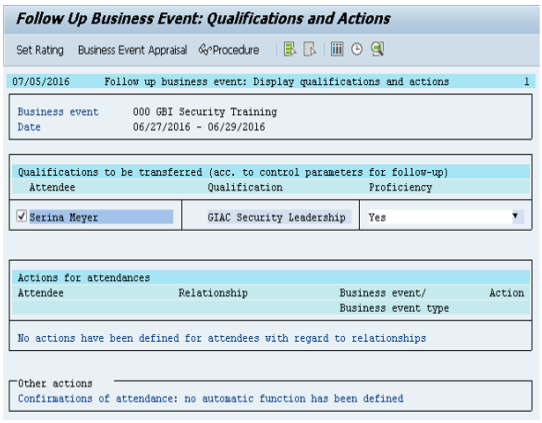


Figure 10. Follow up business event

Source: GBI tutorial, SAP UA [3]

There is an opportunity to follow up the booked event and if it is necessary to mark it as fixed in the SAP system. Additional if it is necessary you can correct the period of time for the actual plan version (Figure 10).

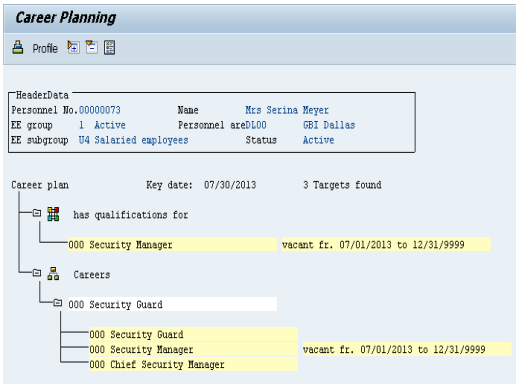


Figure 11. Career planning

Source: GBI tutorial, SAP UA [3]

Person must be selected and the personnel number of the employee must be entered. As the reporting period, the current date must be chosen (as the Key date). The planning criteria shall include qualifications and careers.

A case of personnel development and career planning was able to see and the way how the HCM-modul from SAP could support the previous HR processes. It makes easier for HR experts to manage and help employees in their path of carrier.

IV. CONCLUSION

On the basis of research companies with strong HR function are better at: hiring employees, developing employees, and retaining employees. Other important requirement is the understanding of the business processes of the company by the employees. Founders and managers have responsibility to share the vision, mission, purposes of the company with the employees and empower them to make decisions on their own in order to feel that they are

appreciated. Employees understand the mission and goals of the company if the information is shared with them, so they can give the mission forward for the partners of the company. Employees work better if the present situation and the future is cleared. They are able to understand the processes of the company through the ERP-system. The author wrote about the worldwide SAP-system, about its up-to-date, latest version: S/4 HANA. The work brought into focus the processes of HR, which processes are as important in start-up companies as other small, medium or big companies. The computer added HR processes by SAP (HCM-modul) consist of seven processes: Organizational Management, Personnel Administration, Recruitment, Personnel Development, Talent Management, Performance management and Personnel Controlling. The work deals with the Personnel Development and Talent Management empathically because these processes are indispensable for a recent successful company in our economy with labour shortage. Employers must retain their employees, in this way the company can save (recruitment) costs and can be much more productive if the employees feel connected, important, valued and empowered. It is worth to invest into introducing HCM-modul of SAP HANA because of the transparency and the better opportunity of organizing of Personnel Development and Talent Management. When a start-up company have been growing and the number of employees has been rising, the bigger organization requires other skills, knowledge, tasks from the employees and managers. They must develop themselves personally. The work of HR Department supported by SAP HANA is definitely more efficient and successful in the field of Personnel Development and Talent Management (Career planning) than without HCM-modul of SAP software.

REFERENCES

[1] M. E. Porter: How Competitive Forces Shape Strategy. HARVARD BUSINESS REVIEW, 1979.

[2] J. Hetyei: ERP systems in Hungary in the 21. century. Computerbooks, 2009

[3] S. Magal, Stefan Weidner, Jeff Word: GBI tutorial, SAP UA

[4] T. Dr. Orosz: Enterprise Resources Planning SAP business processes and programing, Óbuda University 2013.

[5] A. Selmeci: SAP technology architecture Óbuda University 2015.

Humanoid robot communication

D. Ódor*, L. Gugolya**
* Óbuda University – Alba Regia Technical Faculty, Székesfehérvár, Hungary
odor.dominika95@gmail.com
gugolya.laszlo@amk.uni-obuda.hu

***Abstract*—This article gives you a insight of how it works to speak with a humanoid robot. And how the speech recognition works in those machines. I am going to show the limits of those technologies.**

I. INTRODUCTION

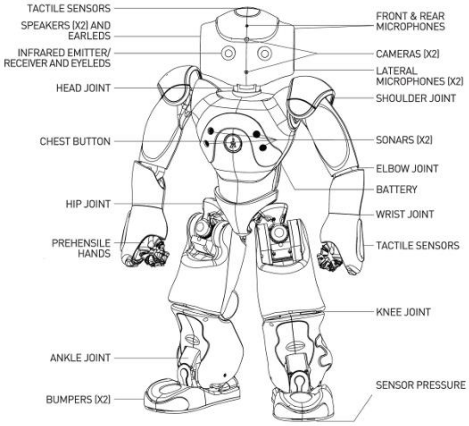
My goal is that the reader of the following lines understands how the speech recognition of the robots works. Which system better for the exercises and what the problem whit the construction of those robots. This theme is all based on my experiences.

A. The humanoid robots

I tested those humanoid robots. One was Nao version V5[2], and the other one was Pepper V1.8a version[3]. Those two robots are manly different in their physique appearances and slightly different in their software parts.

I am not going to introduce all the parts of those robots, because most of them are irrelevant in my theme.

The Nao robot is a bit older than the Pepper robot.



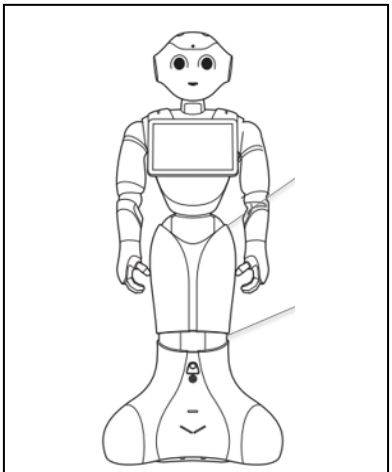
1. Figure Physical appearance of Nao

This was maybe the first humanoid robot that could interact and communicate with people on a normal way.

Pepper, the other robot is a little newer than Nao. It was designed differently, and they used better parts to make it.

The creators of those robots made a special platform to make the programming easier for individuals.

The old software of those robot uses a modular platform, but if somebody wants to make something more complex, then it is possible to do Python programming or use them mixed. Users can modify the modules too.

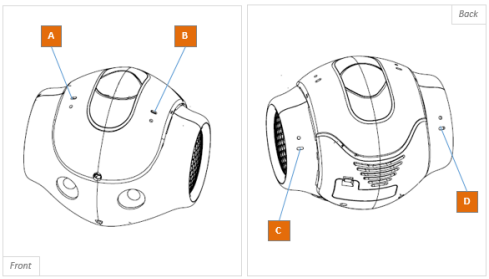


2. Figure Physical appearance of Pepper

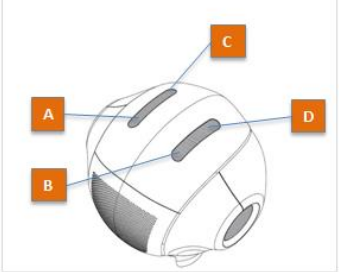
As you can see in the table below, in Nao they used a different TTS – Text To Speech – engine than the Pepper. But in the newer robot they changed it to the world-wide used Nuance.

Table I.
Nao and Pepper compare

Robots	Nao	Pepper
"Age"	9 year	4 year
Used speech recognition system	ACAPELA	NUANCE
Locate of microphones	on the front and back of	on the top of its head



3. Figure Location of the microphones on Nao

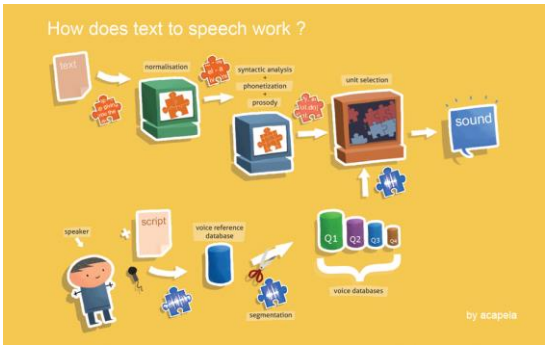


4.Figure Location of microphones on Pepper

II. HOW SPEECH RECOGNITION WORKS

A. *The process*

For the better understanding you should look at the picture below.



5. Figures The process of speech recognition

The first step to reproduce the natural sound of any language, a people records different contents which contain every possible sounds of the chosen language.

Then those recordings are segmented and sorted into an acoustic database. After this the system are creating new databases where it puts in the further sliced recordings, like sentences, phrases, words and etc.

When we reproduce words, the TTS system begins a special analysis that transfers written text into acoustic text. Before the system plays the sound, it sets the proper toning and pronounce.

The last step when we can hear the results.

B. *AKAPELA*

The Acapela[1] group did this speech recording service that was used at Nao. It is hard to decide if it is a well written program or not, because there can be other issues too that can manipulate the results.

In the description of Nao it is written that people need to speak with it from 1 meter distance and a bit louder than the normal.

The problem that makes things more difficult for Nao to understand the sentences, can be the cooling system in the back of its head. Maybe it makes so much noise that it causes misunderstandings or do not even understand the person who tries to communicate with it.

It is used nowadays too for education robots, toys and for people whit voice damage. They can create a new or similar to the original sound.

C. *NUANCE*

This platform is more popular than the other. It is used in iPhones for the Siri API and other in smart devices.

In Pepper case it works more reliable than the ACAPELA in Nao. It understands much more precisely. But if we compare the two robots in this aspect, it can be that Pepper understands better because its microphones are better and more sensitive than Nao's. But this can work hectic too, because the creators put there too the cooling system into Pepper's head. So, the cooling fan can bother the recognition here too.

But if we compare the two system, then I think that NUANCE is more advanced than the ACAPELA. We can even choose from more languages.

III. SPEAK WITH A HUMANOID ROBOT

As I already mentioned there is specific software that is used for those human-like robots. Its name is Choregraphe [4]. With this, users can create a unique dialog or smaller speeches.

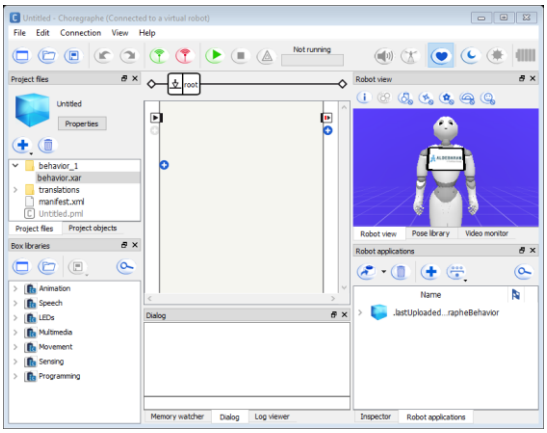
But before a person begins to do something, it is very important that those robots need a network via they can communicate with their TTS system. If it is not possible then we cannot speak with our robots, because those robots make a real time speech recognition.

So, if there is no network signal then we have robot doll that are not able to any interaction with people.

Another important thing if we speak other language we must strive to say the words with more regular pronouncements.

So let's start. On the next picture you we can see the surface of the Choregraphe software.

As we can see, it is a pretty complex but easy to handle program. It has a virtual robot part too, where we can test

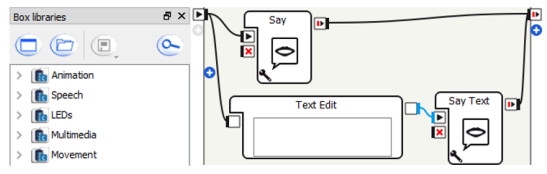


6. Figures Choregraphe

the programmers that the users wrote. So we do not need to wait until we can use the real humanoid robots, because there are many segments we can try at those virtual machines.

This would be the ideal software if we could test every part of the written application we made and then we should not even have to own a humanoid robot. But sadly this is not the case. There are many functions that we cannot test on the virtual robot. Some of those are the speech recognition and the tablet functions. For those actions we need the real ones.

But out of this, it has a very user friendly environment, because this in this application we are capable to make programmers in a modular way. The creators of this program gave the us an environment that the beginners can easily use after they get to know the basic knowledge.



7. Figures Choregraphe windows

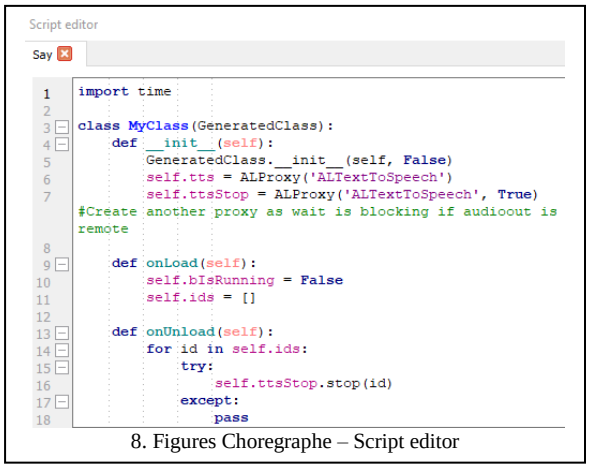
On the left picture we can see the Box libraries menu, where we can find different modules that the users can use for their plans. There are many categories that covers most of the parts and functions of the robots. With those, anybody can make a simple or a little more complex application.

As you can see on the right figure, there are some of these modules, which can make our robots say something. Both variations do the same thing. In the case of the upper one we need to open its setting on the left side and write there the text we want, on the other hand in the second solution there is a Text Edit box where we need to write the text and this box must be connected to the other Say Text box.

There are not any significant difference in the two solution, but for some reason the first one is not as reliable as the second one. There were many times when the robot did not say it. So in many cases I used the other one, because that has always run.

Out of the text we can change many other things too. Some of those the pronunciation, the voice shaping, the speed, the volume and etc. Some of those setting can be modified if we open the settings of the Say boxes, but most of these must be written in the text by the user. Before I show this method I am going to introduce the readers another, a more professional method for the programming.

In this case we will not use the modules, because we will write our codes in Python. The humanoid robots and the Choregraphe too compatible with this language. And if we open the script of a box, we can see that it will show us a Python code.



8. Figures Choregraphe – Script editor

If we do not want to type the whole code ourselves, then we can copy or modify the current code of a box. With those we can create a more complex program than before.

By the way we do not want to use the Choregraphe then we need to make sure to write in the code a connection part with the robot.

```
from naoqi import ALProxy
tts = ALProxy("ALTextToSpeech", "the ip of your robot", 9559)
```

We need to call the ALTextToSpeech module if we want to make our robot to say something. The next thing we need is our robot IP address. The last step is to call the say function.

```
tts.say("Hello world!")
```

After this we play our code, then the real robot are going to say the text. If this is not happening, then check if you have internet connection, the IP address of your robot and your code.

How can we modify the voice of our robot? As I have already mentioned there is a solution for this. The developers made special tags that we need to put in our text in the right place for this. I am going to introduce some of the basic tags.

•\\vct=value\\

With this tag we can change pitch of the voice. We can set the value in percent between 50 and 200. The default value is 100. For example:

```
# Say the sentence with a pitch of +50%
tts.say("\\vct=150\\Hello my friends")
```

•\\rspd=value\\

This will change the speed of the speech. Just like the previous one we use percentage as value, but there we can modify it between 50 and 400. The default there too 100. Example:

```
# Say the sentence 50% slower than normal speed
tts.say("\\rspd=50\\hello my friends")
```

•\\pau=value\\

We can guess what this tag will do. With this we can put pause inside our text. Its value needs to be given in millisecond. Example:

```
# Insert a pause of 1s
tts.say("Hello my friends \\pau=1000\\ how are you ?")
```

•\\vol=value\\

In this case we can modify the volume. The value can be given in percentage between 0 and 100. Exapmle:

```
# Say the sentence with a volume of 50%
tts.say("\\vol=50\\Hello my friends")
```

•\\rst\\

With this we can reset all the settings we did so far.

```
tts.say("\\vct=150\\\\rspd=50\\Hello my friends.\\rst\\ How are you ?")
```

There are many other tags that we can use during our programming, but they can be used just on the Pepper humanoid robot, because that machine uses the Nuance system. And just with this system can we change some options. This can be the reason why the SoftBank Robotics used Nuance for the new released Nao robots.

One of the most important options is that we can change any words pronunciation. With this we can set how the robot will say the selected words. Although the robot uses a library for its speech recognition via the internet but if we change the specifications of the word then that will be stored on the robot so those settings will remain even if the robot is switched off.

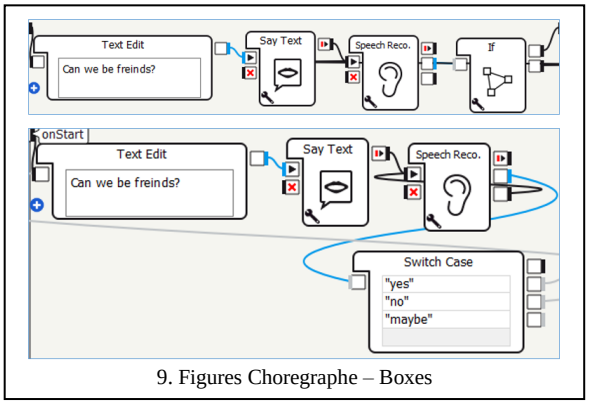
For this we use the addToDictionary function of the ALTextToSpeech modul. Example:

```
tts.setLanguage("English")
# Say Emile in english
tts.say("My name is Emile.")
# Add Emile to dictionary
tts.addToDictionary("Emile", "\\toi=lhp\\E'mil\\toi=orth\\")
# Test it
tts.say("My name is Emile.")
# Delete the word Emile
tts.deleteFromDictionary ("Emile")
# Re-test it
tts.say("My name is Emile.")
```

As you can see on the figure above the first step is to set the language which we want to modify the word. The next step is that we make the robot to say out the current text. Then we change the pronunciation of the word, Emile. After this we test it. If we want that this modification will be temporarily then we need to delete it with the deleteFromDictionary function. At last let's test it if we get back the original word.

It is all good, that we can make our robot to say monologues, but what we want is that we can make conversations with our robot. For this task we have two solutions.

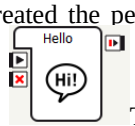
First we can use the Say modules in the Choregraphe and we can complete it with a Speech recognition and an If or Switch case boxes.



9. Figures Choregraphe – Boxes

On the top picture I used the „if" version. In this scenario we must set the condition operator and the value we want our robot to understand. We can add multiple words or sentences if we separate them with „ ; " this sign. But with this we can just make it able to understand our words but after this it is up to us, how it should treat it or what it should do.

But if we use the Switch Case box then this problem would be automatically solved, because each word has its own output that we can connect anywhere we want.

Although we want to make a proper conversation we need from them many copies and many so questions, that we can guess what the answer will be. But for this problem the developers created the perfect module. And this one is the Dialog box. 

This box is already renamed by me.

If we open its script it is a bit different than the other boxes.

```
concept:(greetings) ^rand[hi hello "hey there"]

concept:(drink) [red white] wine
concept:(alcohol) [beer ~drink]

u:(~greetings) ~greetings
u:(do you have _~drink) yes, I have $1
u:(I want to drink something) do you want ~alcohol?
```

There we will not see any code, just the ones we use for making the dialog. The one I used for this example is a simler one. There we can see some lines that I will describe. As we can see on the picture there are 3 concepts where each contains different words. If the words are between in [] those containers, then it means we can say any of them and the robot will understand it. If there is a phrase then we need to put between quotation mark as we can see by the greetings. Buti f there is a word out [] those signs then that means that word is fix.

The real conversation starts under those lines. The wavy sign with the word „greetings" means that when the user say one word from the greetings concept the robot will understand it and says a greeting from the same concept, and it will be a random one. After this if we ask the robot the next question and say a drink from the possible answers then the underline means our robot saved our answer and in its sentence the „$1" means that it will say out our choice.

With this I showed many possible ways that will lead us to having a simple conversation with our robots, but there so many other possibilities that we could use during those tasks that I could write a whole book from them. But this is a preview from what we are capable to do on a humanoid robot. And if it is piqued someone interest then they should take a visit at the website of the Aldebaran.

IV. SUMMARY

My opinion in this theme that we are on the right path to make a humanoid robot that can understand everything that a person says to them. There are multiple TTS and voice recognition systems outside of the mentioned ones. It is good that we have those awesome services, but it is equally important that the hardware part be functional without bothering each other.

I do not know what the developers' solution will be, but in short time we will found out, because they are already making a new type robot, the so-called Romeo. It will be used like a caretaker for old people, so they must solve those problems.

REFERENCES

[1] http://www.acapela-group.com/voices/how-does-it-work/
[2] http://doc.aldebaran.com/24/family/nao_h25/index_h25.html#nao-h25
[3] http://doc.aldebaran.com/25/family/pepper_user_guide/repair/brake_release_keys.html
[4] http://doc.aldebaran.com/2-5/index.html

Applying Data Mining Algorithms on Demographic Data with Power BI

Tsvetanka Georgieva-Trifonova*, Tihomir Trifonov**, Márta Seebauer***
* "St. Cyril and St. Methodius" University of Veliko Tarnovo, Department of Mathematics and Informatics, Veliko Tarnovo, Bulgaria
** Vasil Levski National Military University, Veliko Tarnovo, Bulgaria
*** Óbuda University, Budapest, Hungary
cv.georgieva@uni-vt.bg, TihomirTrifonov55@gmail.com, seebauer.marta@amk.uni-obuda.hu

***Abstract*— Analyzing business results along with demographic data is a useful approach to discovering important summaries, hidden dependencies, and predictions to support making successful business decisions. This paper examines the features of Power BI for performing clustering; correlation analysis; descriptive analysis by using decision tree. Illustrative examples about data, available on the website of the National Statistical Institute in Bulgaria are presented. The results confirm the benefits of applying these analyzes to demographic data.**

I. INTRODUCTION

Business companies store and continuously generate data from different domains related to their business: ads, sales; customers, users, feedback reviews; accounting data; text content, video data, photos, graphics, others. Data accumulation strategies may include tools for:

- analyzing customer needs and preferences to provide a better understanding and targeting the customers;
- managing and searching for consumer feedback, including from social networks;
- support ads and ad campaigns targeting, etc.

The gathered data provide huge opportunities for companies, but various researchers question whether the maximum benefit from them is extracted in fact. If the archiving data work strategy is presented as a human needs hierarchy defined by Maslow, the resulting pyramid is shown in Fig.1. According to [13], the level of media companies falls into the middle, between the stages of "Information" and "Knowledge".

Figure 1. Maslow's hierarchy presented the stages of working with big data [13]

In order to reach the "Wisdom" stage, companies need to take advantage of the work and data analysis capabilities. For this purpose, they have to invest in technology and training. The expected result is to improve processes such as those related to customer service and products.

The present research is motivated by the fact that the combination of demographic data with data from a particular business activity or customer feedback and the implementation of data mining algorithms allows to obtain aggregated information, forecasts, and dependencies, useful to support making adequate business decisions.

The advantages and disadvantages of implementing data mining algorithms through Power BI are summarized. Execution of clustering, correlation analysis, and descriptive analysis by tree decision on demographic data is demonstrated.

This paper is organized as follows. In Section 2, an overview of existing research to implement demographic data mining methods is made. Section 3 summarizes the advantages and disadvantages of applying data mining by using Power BI. Section 4 is dedicated to experimenting with demographic data downloaded from the National Statistical Institute's website in Bulgaria to derive the benefits that Power BI offers.

II. REVIEW OF EXISTING RESEARCH TO IMPLEMENT DEMOGRAPHIC DATA MINING METHODS

Clustering methods are successfully applied when analyzing demographic data [2, 6, 10, 11].

In [1], the methods of Holt and Holt-Winters for multiplier and additive seasonality are applied to obtain monthly forecasts of the average monthly salary (in BGN) of employees in Bulgarian tourism. The forecast is based on data from the National Statistical Institute on the average annual salary of employees under labor and employment relationships by economic activities and forms of ownership in the hotel and restaurant sector.

A correlation analysis is applied in [9] to explore the relationship between demographic data of Twitter readers (education, location) with their credibility perception. In order to improve the effectiveness of e-learning, a classification is applied to define a set of student profiles in [14]. For the classification, survey data are collected, some of which relate to demographic information, such as sex, age group, university, faculty, specialization, school year.

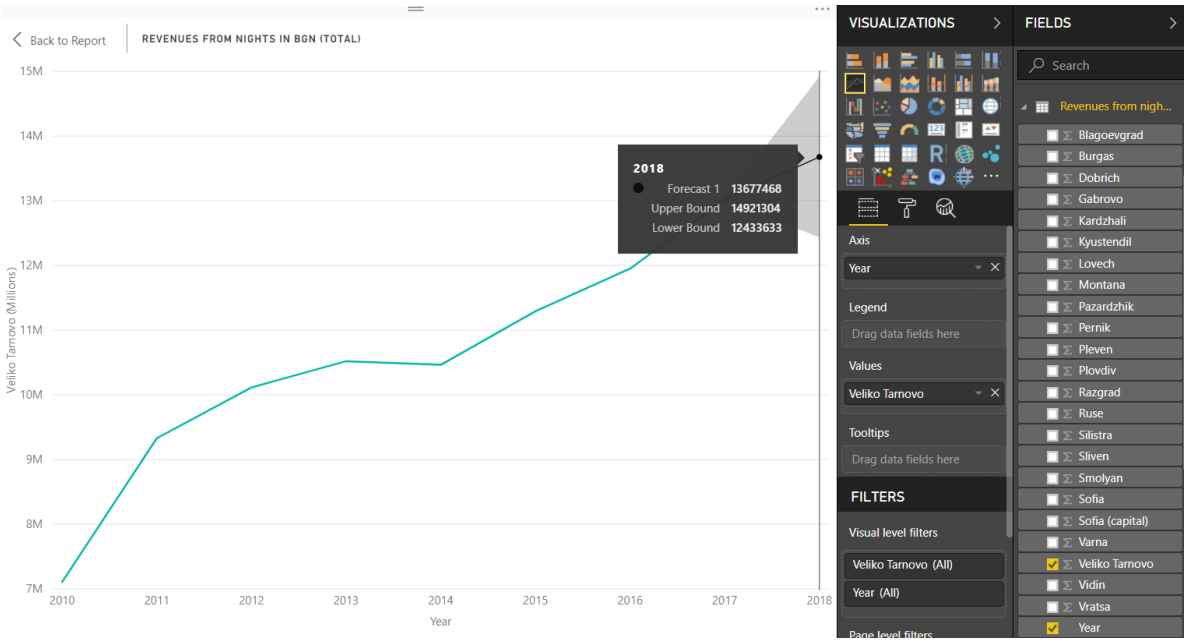


Figure 2. Overnight stays in Veliko Tarnovo District

Demographic sequences are examined and analyzed in [5, 8]. Comparing methods for demographic data classification is performed in [8]. Various methods such as decision trees, support vector machines, K-Nearest Neighbors (K-NN), neural networks are used for classification.

A product recommendation system is proposed in [12], which uses additional information, including demographic data about users, in order to improve the accuracy of the recommendations made.

In the present paper, the features of Power BI for implementing data analysis as clustering, correlation, analysis, descriptive analysis by tree decision are examined.

III. Advantages and Disadvantages of Applying Data Mining through Power BI Units

Microsoft Power BI is a flexible tools for applying Business Intelligence, which facilitates users to self-generate data analytics managed in the cloud for the collaborative work and sharing [3]. This tool is distinguished from a data warehouse because it allows:

– *User-defined, flexible data retrieval and analysis;*

This includes creating calculated columns; changing measures and dimensions on-the-fly, and without the need for extra use of other IT and ETL (extracting, transforming, and loading) tools.

Typically, data warehouses require data to be imported (via ETL tools) into the warehouse;

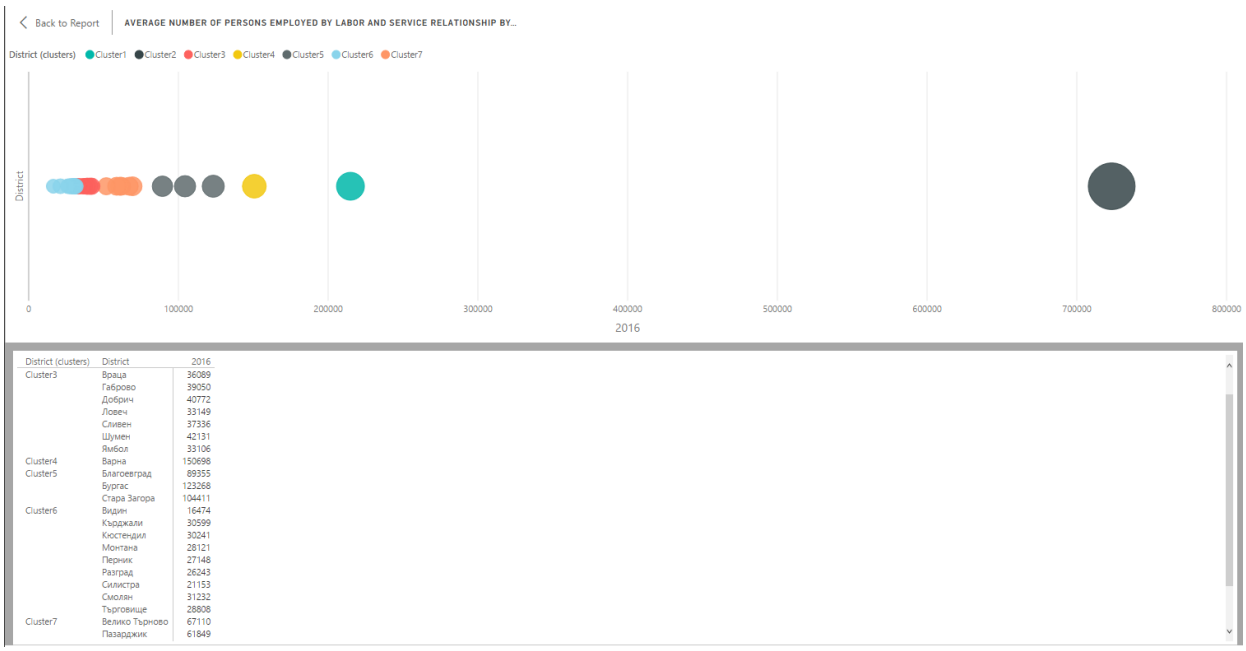
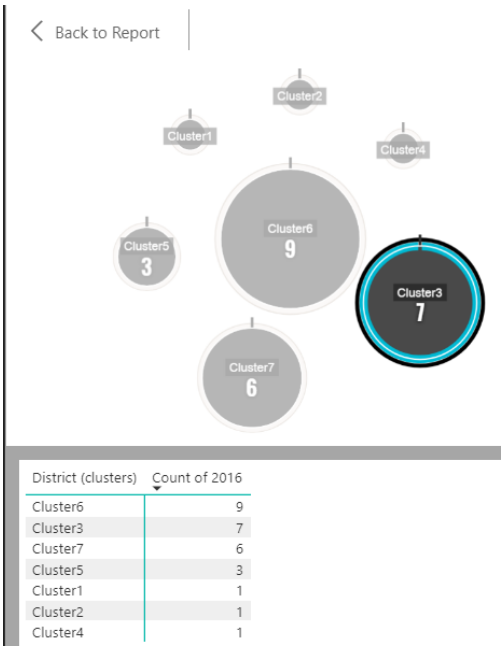


Figure 3. Clustering of the Regions According to the Number of Employers



District (clusters)	Count of 2016
Cluster6	9
Cluster3	7
Cluster7	6
Cluster5	3
Cluster1	1
Cluster2	1
Cluster4	1

Figure 4. The Clusters Visualized with Cluster Map

generating cubes with data; then analyze the data, thus making the BI process heavier, dependent on the IT technology used and difficult to repeat in short periods.

– *Working with a graphical interface;*

Online publishing in such a way that allowing data to be shared from a central location;

A. *Advantages of Power BI*

The benefits of Power BI can be summarized as follows:

(1) It provides the ability of analysts who can retrieve source data, create a dataset, transform or manipulate data, visualize data, and publish the resulting reports and dashboards;

(2) The basic Power BI technology is a database that supports table structures for data used by Power Pivot. By this way a compromise between performance and ease of use is achieved. On the other hand, data cubes require more complex structuring and query language, such as MDX (MultiDimensional eXpressions);

(3) The language DAX (Data Analysis Expressions) is based on a relatively simple statement used to create computed columns and measures;

(4) Q&A feature. This feature allows retrieving a response from the data by asking a question by performing natural language processing. Q&A is the most commonly mentioned benefit for self-service Business Intelligence;

(5) Actively maintained and updated. The available visuals are continuously updated by the community. Interactive visualizations of geographic maps are authorized by Bing Maps. Dynamics CRM integration also evolves;

(6) The free version reduces user limitations.

B. *Disadvantages of Power BI*

The following major limitations and disadvantages of Power BI can be identified:

(1) A given dataset may include multiple datatypes, but Power BI reports and dashboards can only generate data from one dataset; Similarly, Power BI can not combine imported data together with data available from real-time connections;

(2) Power BI can not work with files larger than 1 GB. The limitation refers to .pbix files, which are a type of archive files and compress data until they have to be used. However, the maximum allowed file size may limit Power BI to a subset of the data extracted from the data warehouse(s) of the organization concerned;

(3) For a dataset size, the limit is up to 1GB when using the free cloud service PowerBI.com. An example solution is to create multiple smaller datasets. For Pro version of PowerBI.com cloud service, the limit is up to 10GB;

(4) The granularity is not supported. Microsoft Power BI does not allow for creating scheduled reports, customized user views, notifications, and reports.

(5) In the presence of multiple different use cases and user profiles that require customization, Microsoft Power BI is not the best solution.

(6) Data cleaning. Microsoft Power BI has the ability to clean and transform imported data to achieve improvement of the data quality.

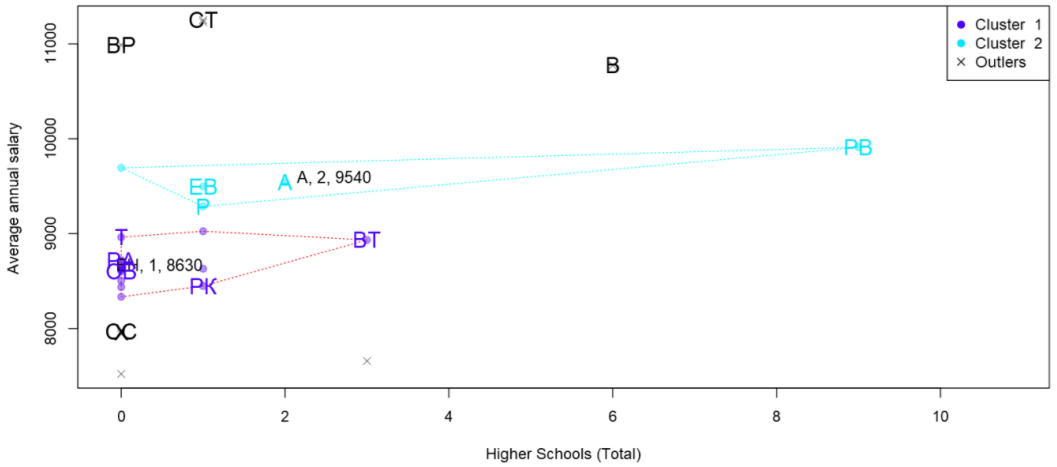


Figure 5. Clustering Based on the Average Annual Salary per District and the Number of Higher Education Institutions

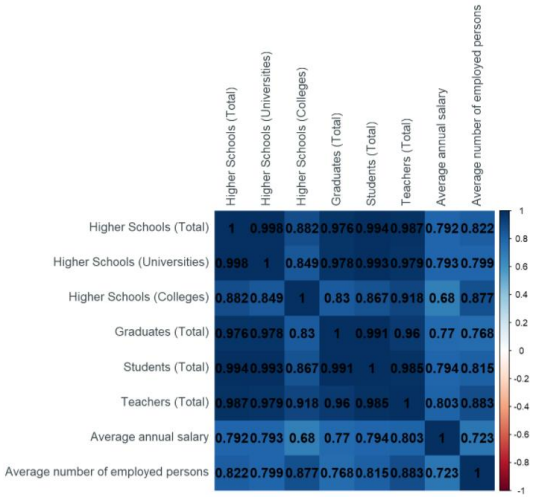


Figure 6. Complex Correlation Matrix

However, for more sophisticated transformations, it is necessary to use other tools such as MS SQL Server Integration Services or R script [4];

(7) This Microsoft tool is often used to supplement instead of replacing other reporting tools. In many cases, Power BI does not replace the corporate data warehouse. The tools for company data storage can be used for rarely changing reports when large-volume data processing is required. On the other hand, Power BI can be used for single and continuous, frequently changing analyzes of smaller datasets.

Based on the performed study of the main advantages and disadvantages of Power BI, we can conclude that the analysis of demographic data can be successfully done using this tool. Experimenting with specific data is described in the following section.

IV. Analyzing and Visualizing Demographic Data through Power BI

In addition to standard data representation in graphics and charts, Power BI allows the implementation of forecasting algorithms based on available archive data; hidden previously unknown dependencies in the data.

In this research, the analysis and visualization are applied on the demographic data obtained from the website of the National Statistical Institute in Bulgaria concerning

- the average number of employed persons;
- average salary;
- number of universities;
- number of university graduates by district, etc.

The visualizations used are:

- Cluster map;
- Correlation plot;
- Clustering with outliers;
- Decision Tree Chart.

A. *Forecasting*

Forecasting in Power BI is based on a method called exponential smoothing, which is a modification of the Holt-Winters method. The concrete example shown in Fig. 2 refers to overnight stays in the Veliko Tarnovo district at a 95% confidence interval.

B. *Clustering analysis*

Power BI Desktop provides the ability to automatically find clusters of data. Clustering allows groups of similar data to be discovered in the dataset.

Fig. 3 shows clustering of the regions according to the average number of persons employed by labor and service relationship by districts in 2016. As a result, seven clusters are obtained.

The visualization *Cluster Map* is designed to represent clusters of related data. An appropriate data summary can be selected – number of objects in each cluster (Fig. 4); sum; average; minimum, maximum value, etc.

The visualization *Clustering with Outliers* uses DBSCAN algorithm (*Density-Based Spatial Clustering of Applications with Noise*), which also finds the exceptions.

Fig. 5 depicts the clusters found, when the points being derived based on the average annual salary per district and the number of higher education institutions in the respective district.

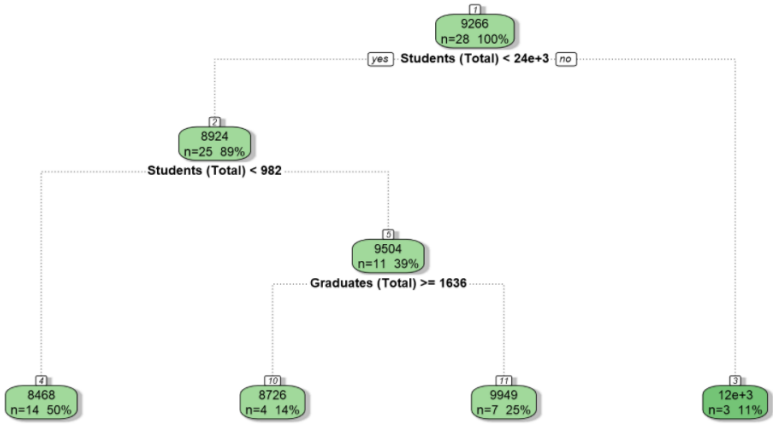


Figure 7. Correlation Tree between the Number of Students, the Total Number of Graduates and the Average Annual Salary per District

C. *Correlation analysis*

The visualization *Correlation Plot* is based on correlation analysis, a method that aims to find a relation between different attributes in a given dataset. The presence of a relation between two attributes means that when there is a change in one attribute, the other one is also changed. For this purpose, a correlation coefficient is calculated, i.e. the so-called Pearson coefficient [7], whose value is in the range –1 to +1.

The values close to:

–1 indicate a negative relationship, i.e. the high numerical values of an attribute refer to low numerical values of another attribute;

+1 indicate a positive relationship, i.e. the high numerical values of an attribute refer to high numerical values of another attribute;

0 – lack of correlation.

On Fig. 6, the correlation between the following attributes is examined: the number of higher schools; graduate number; number of students; number of lecturers; average annual salary; average number of persons employed by labor and service relationship by districts in 2016.

D. *Decision tree*

The visualization *Decision Tree Chart* implements a recursive splitting during the decision tree construction.

Fig. 7 shows the decision tree obtained by the following settings:

- the input variables are the total number of students and the total number of graduates in the respective fields;
- the target variable is the average annual salary per district.

Applying the presented visuals to demographic data combined with business data could provide descriptions and summaries of data; forecasts; dependencies that are helpful in supporting appropriate business decisions.

V. CONCLUSION

Advanced and easy-to-use tools are intensively developed and refined to perform complex data analysis. It's no accident that Gartner recognizes Microsoft as a leader in BI platforms for 11 consecutive years. The present paper explores the benefits of using Power BI to implement algorithms to mining demographic data and conveniently visualize the results obtained.

ACKNOWLEDGMENT

The work presented in this paper was partially supported by the research project “Analyses and research with cloud technologies for business”, “St. Cyril and St. Methodius” University of Veliko Tarnovo.

REFERENCES

[1] Preslav Dimitrov, Can the Bulgarian Tourism solve the problem with poverty in Bulgaria?, In book: The Fight with Poverty in Bulgaria: Macroeconomics, Ethical Aspects, Eurointegration, the Role of TourismChapter, Fridrich Ebert Stiftung: Beureau Bulgaria & Institute for Economics and International Relations, Sofia, 2014.

[2] Nontembeko Dudeni-Tlhone, Jenny P. Holloway, Sibusisiwe Khuluse, Renée Koen, Clustering of Housing and Household Patterns Using 2011 Population Census, 55th Annual Conference of the South African Statistical Association, 2013.

[3] Esat Erkec, How to create advanced analytics using Power BI and R scripts, SQLShack, 2018, available at: https://www.sqlshack.com/use-advanced-analytics-power-bi/

[4] Oliver Frost, Power BI: Transform Data with R, 2017, available at: https://www.consolidata.co.uk/explore/blog/power-bi-transform-data-with-r/

[5] Dmitry I. Ignatov, Ekaterina Mitrofanova, Anna Muratova, Danil Gizdatullin, Pattern Mining and Machine Learning for Demographic Sequences, International Conference on Knowledge Engineering and the Semantic Web, 2015, pp 225-239.

[6] Elza Jurun, Nada Ratković, and Ivana Ujević, A cluster analysis of Croatian counties as the base for an active demographic policy, Croatian Operational Research Review, vol. 8, 2017, pp 221–236.

[7] Daniil Maslyuk, Pearson correlation coefficient in DAX, 2017, available at: https://xxlbi.com/blog/pearson-correlation-coefficient-in-dax/

[8] Anna Muratova, Pavel Sushko, Thomas H. Espy, Black-Box Classification Techniques for Demographic Sequences: from Customised SVM to RNN, In Proceedings of the Fourth Workshop on Experimental Economics and Machine Learning, Dresden, Germany, 2017, pp. 31-40.

[9] Shafiza Mohd Shariff, Mark Sanderson,and Xiuzhen Zhang, Correlation Analysis of Reader’s Demographics and Tweet Credibility Perception, 38th European Conference on IR Research, 2016, pp 453-465.

[10] Oksana Shubat, Abilova Makhabat, Anna Bagirova, Anton Ivlev, The Use Of Cluster Analysis For Demographic Policy Development: Evidence From Russia, Proceedings 30th European Conference on Modelling and Simulation, 2016.

[11] Ondřej Šimpach – Marie Pechrová, Cluster Analysis of EU’S Regions According to Demographic Criteria, The 10th International Days of Statistics and Economics, Prague, September 8-10, 2016.

[12] R. Satish Srinivas, C. S. Anish Balaji and P. Saravanan, Online Product Recommendation using Relationships and Demographic Data on Social Networks, Indian Journal of Science and Technology, Vol. 9, no. 44, 2016.

[13] M. L. Stone, Big Data for Media, Reuters institute for the study of journalism, University of Oxford, Report, 2014.

[14] Alexandru Topîrceanu, Gabriela Grosseck, Decision tree learning used for the classification of student archetypes in online courses, Procedia Computer Science, vol. 112, 2017, pp.51–60.

Collective Behavior Simulation of Mobile Robots Using Lego NXT Sets

Marta Seebauer and Peter Udvardy
Óbuda University, Alba Regia Technical Faculty, Székesfehérvár, Hungary
seebauer.marta@amk.uni-obuda.hu, udvardy.peter@amk.uni-obuda.hu

***Abstract* — For response to the challenges of Industry 4.0 a new subject the "Simulation of Multi-agent Mobile Robots" was introduced into the curriculum of Mechatronic Engineering MSc course. It was an important aspect to make the study program for practice-oriented. Lego NXT sets were used for the simulation collective behavior of mobile robots. The paper summarizes the experiences of these practices.**

I. INTRODUCTION

The forth industrial revolution holds the name of Industry 4.0 and focuses on the smart technology and the automation of production by introducing intelligent robotic systems based on artificial intelligence.

The aim of robotics is to compensate the demand on labor market while improving the quality and also the efficiency of production. Japan was the predecessor of the introduction of industrial robots in the 1980s. This was the first success story of artificial intelligence applications after the crisis in 1970s.

In the Executive Summary World Robotics for 2018. year about the industrial robots the International Federation of Robotics announced that the robot sales increased by 30% to 381,335 units in year 2017. (Fig. 1) This is a new peak for the fifth year in a row. The main drivers of this exceptional growth in 2017 were the metal industry (+55%) and the electrical/electronics industry (+33%). [1]

TABLE I.
THE DEFINITIONS OF ARTIFICIAL INTELLIGENCE

	Human	Rational
Think	„The exciting new effort to make computers think …, machines with minds, in the full and literal things" [2] "The automation of activities that we associate with human thinking, activities such as decision-making, problem solving, learning…" [3]	„The study of mental faculties through the use of computational models." [4] "The study of the computations that make it possible to perceive, reason, and act." [5]
Act	"The art of creating machines that perform functions that require intelligence when performed by people" [6] "The study of how to make computers do things at which, at the moment, people are better." [7]	„A field of study that seeks to explain and emulate intelligent behavior in terms of computational processes." [8] "The branch of computer science that is concerned with the automation of intelligent behavior." [9]

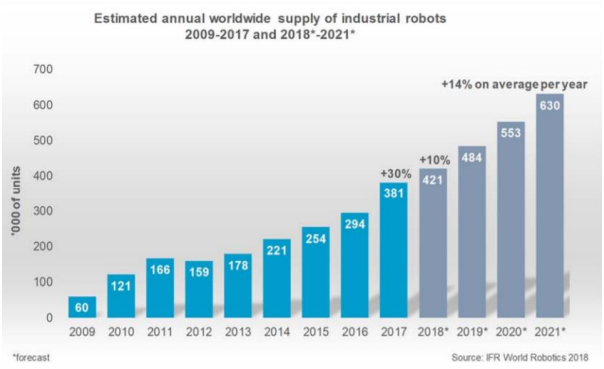


Figure 1. Estimated annual worldwide supply of industrial robots 2009-2017 and 2018-20121 [1]

The robot sales in the automotive industry increased by 22% and remained still the major customer of industrial robots with a share of 33% of the total supply in 2017. There are five major markets representing 73% of the total global sales volume in 2017: China, Japan, the Republic of Korea, the United States and Germany.

The enterprises of Middle-Transdanubian Region already plan to investigate into the smart manufacturing and need the relevant, and high quality human resources.

Óbuda University is prepared for the challenges of the industry and introduced a new subject with the topics of the simulation of multi-agent mobile robots into the curriculum of Mechatronic Engineering MSc course.

II. ARTIFICIAL INTELLIGENCE AND THE INTELLIGENT AGENTS

John McCarthy formulated the definition of artificial intelligence in Dartmouth College in 1956. Several other definitions are presented in the Table I. Before that this field of informatics held the name of cybernetics which was the science of control at that time. The agent technology was developed in 1990s.

A. Intelligent Agents

In artificial intelligence, an intelligent agent is an autonomous entity which observes the environment through sensors and acts upon it using actuators. [10] Some definitions of intelligent agents emphasize their autonomy, and so prefer the term autonomous intelligent agents. Still other authors [11] considered goal-directed behavior as the essence of intelligence and so prefer a term borrowed from economics, rational agent.

TABLE II.
ROBOT COMPLEXITY [12]

Computing Performance [MIPS]	Biological Example	Function	Capability
10	insect	Transportation, guarding, floor cleaning	Mobile operation
5 000	lizard	Lawn mowers, cleaning, garbage collection, warehousing, transportation	No adapting to changing environment
100 000	mouse	Service	Reinforced learning
5 000 000	monkey	Internal simulation for complex problem solving	Fast learning, object and parameters recognition
100 000 000	human	Abstraction and generalization, application of AI	Symbolic thinking, concept using

The structure of intelligent agent is presented in Fig.2. The humans are not in all cases rational. Unlike rationality the agents can imitate also human behaviors.

The agents can be also divided in depend of realization on software or hardware agents. The hardware agents are commonly known as robots. (Table I.)

The most commonly used industrial robots are fixed but the last time the customer increasingly prefer mobile robots. The Table II. shows the required computing performance for realization several capabilities.

B. The Agents Environment

The part of robot simulation is the programming of robots environment.

(1) Accessible – non accessible

The environment is accessible when agent sensing devices have access to the entire state of the environment.

(2) Deterministic – non deterministic

The environment is deterministic if the next state of the environment is clearly defined by its current state and the actions chosen by the agent.

(3) Episodic – non episodic

The environment is episodic if the agent's experience is split into episodes in which the agent's behavior does not depend on the actions of the previous episodes.

(4) Static – Dynamic

The environment is dynamic if it can change while the agent is thinking. If the environment does not change over time, but the agent's performance can do it, the environment is semi-dynamic.

(5) Discrete – Continoues

The environment is discrete if there is a clearly defined finite set of perceptions and actions.

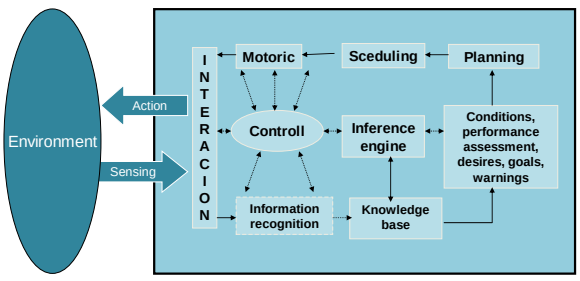


Figure 2. The Intelligent Agent

III. LEGO NXT ROBOTS

LEGO NXT is available since 2006. Two years later the NXT 2.0 was announced. In the original version three motors and four sensors were available (ultrasonic, sonic, light and collide sensors).

The 2.0 version had more parts and had a new colour sensor in the kit. The Brick had an USB interface and four robots could be created. This kit had a special version called 'Educational version' specially designed for educational purposes.

Fig.3. presents the typical structure of NXT robots.

IV. SIMULATED ROBOTS BEHAVIOR

Six different behavioral patterns were selected fro realization.

(1) Coward: this type of robot goes forward slowly and when faces with any kind of barrier it moves backward and changes its direction

(2) Curious: this robot moves with a moderate velocity and when meets another robot or a wall it stops and waits for a while and then moves into another direction

(3) Hungry/voracious: this robot moves around quickly and search for energy source. If the robot meets any barrier it doesn't stop but goes forward at a moderate speed.

(4) Aggressive: this individual goes through fast and in case of disturbance it doesn't change its direction and moves forward without any stop

(5) Hyperactive: this robot goes through very fast and avoid collision with rapid change of its direction

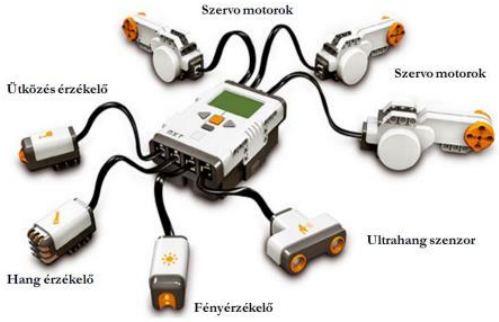


Figure 3. The LEGO NXT robot

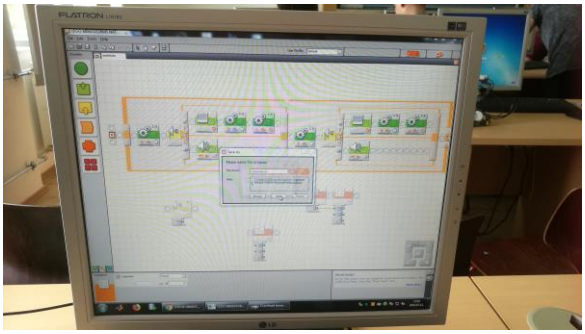


Figure 4. The software development environment of Mindstorms LEGO NXT robot

Fig.4. shows the software development environment of Mindstorms LEGO NXT robot. The different robots were programmed for different behavior pattern. The testing situation is presented on Fig.5.

V. LEGO MINDSTORMS EV3 ROBOTS

Lego Mindstorms EV3 was released as the third generation of LEGO robot series in 2013. In its name the EV refers to evolution and 3 means the third generation. The main advantage of the EV3 set is renewed the programmable brick in which the processor was changed from ARM7 to a stronger and faster ARM9 CPU. Compared to the former NXT and NXT2 versions the EV3 brick contains a USB connection and an SD card slot which can be used up to 32GB.

A. LEGO Mindstorms EV3 Sets

The brick's display also was changed in size. Wifi and Bluetooth connections are also available. The set contains a complete tutorial for building and programming 5 robots (EV3RSTORM, TRACK3R, SPIK3R, GRIPP3R, R3PTAR).

There are two kind of EV3 sets namely Home edition and Educational version. The EV3 home edition set contains 1 programmable brick, 2 large motors and 1 medium motors, 3 sensors (further sensors are available as extras), 1 remote controller which also can be used as a beacon, USB cable and hundreds of LEGO building components. The Educational version contains more sensors.

The EV3 can be fully programmed and controlled in its own software environment called LEGO Mindstorms EV3 home edition is based on LabView environment and the brick can be programmed by using blocks. The Educational version contains teacher resources, tutorials, lesson plans, classroom materials, etc. [15].

A benefit of EV3 robots is that they can be programmed and controlled on Android and IoS platform. The Robot commander app and the EV3 programmer app are available for both platforms. The apps connect to the brick via Bluetooth without using a computer. The programming app uses drag-and-drop programming interface. These application contains pre-programmed codes for the 5 robots (EV3RSTORM, TRACK3R, SPIK3R, GRIPP3R, R3PTAR) and own programmes also can be uploaded to the robots.

Through the advanced functionality, the EV3 robots are more appropriate for behavior patterns realization.

Figure 5. The moving LEGO NXT robot

B. Sensors

The Home edition contains 1 touch sensor, 1 infrared and 1 colour sensor. The Educational version contains 2 touch sensors, 1 colour sensor, 1 ultrasonic sensor and 1 gyroscope. The microphone from the NXT2 set is compatible with the EV3 set.

Using the touch sensor, the robot 'feels' when it bumps into its environment. The sensor detects when the button is pressed or released. The touch sensor can be used as a trigger when the aim of the use of this sensor is not the detection of its environment but the control of the robot's motors. The touch sensor also reacts to a double click function.

The infrared sensor can be used in two ways. The sensor can measure the distance and force the robot to react or can follow an IR source (generally the remote controller) in beacon mode. On the other hand the IR sensor can detect the remote controller's commands and force the robot to fulfil it.

The colour sensor can detect the colour of the surface or the intensity of light. 7 colours can be detected and each colour has its own numeral value (ie. white equals to 1). This sensor must be close to the surface up to 1-2 centimetres for perfect detection). If the sensor measures its environment's light intensity it can be done in two ways: the sensor emits red light and measure the reflection (reflection light intensity) or just measures the environment's light intensity without producing any outgoing light source (ambient light intensity).

The ultrasonic sensor measures the distance of the surrounding objects up to more than 2 metres. The sensor emits high frequency sound waves and measures the reflectance and time. If at least two robots use ultrasonic sensor the robot can only listen to these signals and use them for interaction.

The gyroscope detects changes in rotation and orientation. The rotation is measured in rotation per second and the maximum value is 440 degrees per second. The sample rate of the sensor is 1kHz. Angles can be measured and balancing robots can be built by using gyro sensor.

Large and medium motors are also can be used as sensors as the motors are equipped with a built-in rotation sensor. The sensor can measure the total number of rotations or the angle of rotation in degrees [16].

CONCLUSIONS

(1) The robotics and smart manufacturing are a part of interest of artificial intelligence.

(2) The multi-agent mobile robots are a perspective field of research in the age of Industry 4.0.

(3) The LEGO NXT robot sets were successful used for simulation of collective behavior of robots.

(4) The LEGO EV3 robot sets provides more possibilities for experiments.

REFERENCES

[1] International Federation of Robotics. "Executive Summary World Robotics 2018. Industrial Robots" https://ifr.org/

[2] J. Haugeland „Artificial Intelligence: The Very Idea". Cambridge: MIT Press. 1985

[3] R. Bellman "An Introduction to Artificial Intelligence: Can Computers Think?" Boyd & Fraser, 1978 -146 p.

[4] E. Charniak and D. V. McDermott "Introduction to Artificial Intelligence" Addison-Wesley Pub.Company, 1985 - 701 p.

[5] P. H. Winston "Artificial Intelligence" Addison-Wesley, 1984 – 527 p.

[6] R. Kurzweil "The Age of Intelligent Machines" Viking, 1992 - 565 p.

[7] E. Rich and K. Knight ""Artificial Intelligence" McGraw Hill, 1991.

[8] R. J. Schalkoff "Artificial Intelligence: An Engineering Approach" McGraw-Hill, 1990 - 646 p.

[9] G. F. Luger and W. A. Stubblefield "Artificial Intelligence: Structures and Strategies for Complex Problem Solving"Addison-Wesley, 1993 - 824 p.

[10] G. F. Luger "Künstliche Intelligenz. Strategien zur Lösung komplexer Probleme" ISBN 3-8273-7002-7 Pearson, 2001. pp. 45-47.

[11] S. Russell and P. Norvig "Artificial Intelligence: A Modern Approach" 3rd Edition, ISBN 978-0136042594 Pearson, 2009, pp. 971-1052.

[12] Szabó Richárd: Mobil robotok szimulációja, - ELTE Eötvös Kiadó, Budapest,2001 – ISBN 963-463-476-1

[13] I. Nagy "Behaviour Study of a Multi-Agent Mobile Robot System during Potential Field Building" Acta Polytechnica Hungarica Vol. 6, No. 4, 2009. – pp. 111-136.

[14] A. Pásztor, T. Kovács, and Z. Istenes, Swarm Intelligence Simulation with NXT Robots Using Piconet and Scatternet. 5th International Symposium on Applied Computational Intelligence and Informatics, SACI2009, Timisoara, Romania, IEEE Xplore Digital Library, Digital Object Identifier: 0.1109/SACI.2009.5136241

[15] LEGO MINDSTORMS EV3 robots https://www.lego.com/hu-hu/mindstorms/products/mindstorms-ev3-31313

[16] LEGO Engineering EV3 Sensors http://www.legoengineering.com/ev3-sensors/

Navigation possibilities for Autonomous Guided Vehicles

Péter, Udvardy
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
udvardy.peter@amk.uni-obuda.hu

***Abstract*— Autonomous Guided vehicles are commonly used in industrial automatization. Smart manufacturing contains many AGV solutions for every type of industrial activities and the navigation methods of these vehicles are cardinal points of their applicability. This paper introduces the common navigation techniques and their industrial appearences.**

Keywords— navigation, AGV, forklift

I. INTRODUCTION

Autonomous Guided Vehicles or Automated Guided Vehicles (AGVs) consist of two main parts namely the vehicle and the controlling system. Navigation is an important part of this system which means exact positioning and path tracking. The navigation method depends on the industrial environment type, inner and outer application need different approaches and techniques. Each technique has advantages and disadvantages and the optimal solution depends on the exact knowledge of parameters.

The main navigation methods are:

- Natural navigation
- RFID based localization
- Machine vision based navigation
- Inductive fixed guideline
- Magnetic tape guidance
- Light coding
- Laser navigation system

II. NAVIGATION TECHNICS

Natural navigation

The essence of natural navigation is that there is no need to install any driving paths or markers on the wall and the existing paths and boundaries can be removed. The AGV navigates freely and reacts in a very short time due to its sensors. Object detection happens in less than 60 milliseconds which is faster than human beings' reaction time.

The factory's area is mapped and stored in system memory and only some reference points are needed for precise navigation as the system navigates by laser sensors using the existing environment.
Natural navigation is originated from NDC Concept Team at Kollmorgen.

Fig. 1. Natural navigation[1]

Raw data is collected by laser sensors then a map of the plant or warehouse is created. The baseline map is editable and any part of the map can be removed or new parts can be added if necessary. During navigation the real environment and the baseline map is compared by the algorithm. Natural environment navigation is an optimal solution for controlled environments and for specific applications like deep lane storage systems.
One of the main advantages of this system is flexible, there is no need to install any signals or markers thus the implementation time is very short and the installation costs are relatively low. If an expansion is needed in plants or warehouses using this solution there will not be any restrictions for widening the system, the only necessary step is a new mapping. The natural navigation system continuously scans its environment and can intervene when necessary. This navigation type has been being available for 2-3 years.

RFID based localisation

Radio frequency identification (RFID) based location is commonly used in industrial environment. This technology uses wireless radio waves for tracking and is very flexible with an excellent accuracy. During navigation RFID technology generates a great amount of data thus valuable data filtering is necessary for effective navigation.
In industrial plants and warehouses the RFID transponders are implanted into the floor and are covered with resin. The RFID system uses electromagnetic field to identify and distinguish RFID tags which are embedded into different objects. RFID tags contain coded information which can be decoded by RFID readers. The two-way-communication readers are called interrogators as they send signals to the RFID tags and hear their responses. There are two types of tags: one is the passive RFID tag which has no extra power source and is readable from short distances and the other one is the active RFID tag which has a local power supply and can be used from greater distances (even 15 metres or more). Although the RFID tag is often not visible it can be used in both cases, too.
During navigation the fixed RFID tags have a special interrogation area which tell them apart from the other RFID tags. When an AVG forklift uses this local identification method these zones can be easily identified. The RFID tags and readers can use a wide frequency range from 120kHz to 10 GHz depending on the field of usage.
Using RFID navigation, antennas play an important role in positioning. The size and the position installation of the antenna determinate the read range thus the precision of navigation. Placing the tags horizontally or vertically also influences navigation [2].

Machine vision based navigation

Machine vision is a digital image based technology which is very suitable for process automation and automated or autonomous guidance. Machine vision includes hardware elements and sensors, software background and integrated systems, too. The machine vision process consists of the following elements: image capturing, image processing and any kind of use of the output results.
AGVs navigation is not the only interest area of machine vision techniques, it can be used in automated inspections, quality controlling, code reading, etc.
Image capture is the first step made by different types of cameras, lenses and sensors. Smart cameras and other sensors can be connected to the central unit via wired or wireless connectors. Sensors can be laser diode, etc.
In machine vision systems there are different sorts of captured images: 2D or 3D, visible light, multi- or hyperspectral images, x-ray or infrared images, etc. Visible light images can be monochromatic or real coloured. The camera/image resolution is an important issue in machine vision based systems when moving objects or objects by moving sensors must be captured without quality loss. The image type depends on the field of usage; industry needs different images than human navigation methods or self-storage systems.
Image processing type depends on the further objectives of machine vision based systems. There are many image processing methods and only the main types are enumerated: image segmentation, pattern or shape recognition, edge detection, pixel counting and deep learning process. Image processing gives output information for the AGV navigation.
The output of machine vision image process in AGVs systems is the exact position and steering/orientation of the AGV and the decision making possibility for the future.
There are some problems with machine vision based systems. The relatively long period of image capturing/acquisition and image processing limits the velocity of the AGV. The other problem is the limited capability of navigation accuracy which can be improved by using other sensor parallel.

Inductive fixed guideline

Inductive guidance or wire guidance of AGVs is a traditional navigation method and needs more preparatory work in contrast with other available guidance techniques as the factory's floor must be cut and the guidance wire have to be built in the floor. In this system there is a frequency generator as a current and frequency driver for the guide wire. The steering antenna is located on the AGV together with the sensors which are implemented into the moving AGVs and these sensors navigate on the continuous embedded path.
The advantage of this technology is a high reliability and high accuracy in positioning. This kind of navigation is not influenced by environmental conditions such as snow outside or dirt by the wheels' abbreviation, etc. A massive disadvantage of these systems is the inflexibility as the path cannot be changed easily when necessary. In industrial environment metal installations in the ground also influence navigation and may cause positioning mistakes. Additional sensors are recommended to increase the preciseness of this method (see sensor fusion). The installation of the wire system is not feasible financially.

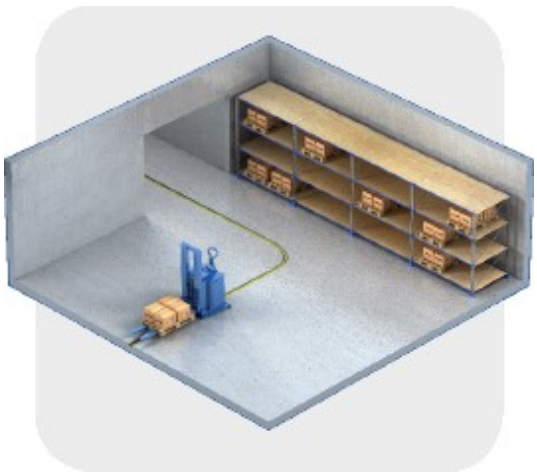


Fig. 2. Inductive navigation [1]

Magnetic tape guidance

Magnetic tape guidance technique in AGV navigation starts with the implementation of the guidance tape which is placed and fixed on the surface of the plant's ground. The width of the adhesive tape depends on the different technologies but at all events is between 20 and 50 millimetres. The tape can be laid in any kind of shapes as it is necessary for routing. Some manufacturers offer heavy load resistant durable magnetic tapes that can be built into the surface by carving.
The sensor, similar to the inductive guidance technique is implemented on the guided vehicle. The other similarity to the inductive method is that the guidance tape is continuous but in contrast with the mentioned system it is visible. The path can be changed easily as it is removed and a new tape is fixed on another route. This guidance system is mainly recommended for automatic guided carts (AGCs).
Magnet sensor like HG G-19600ZA detects the magnetic tape's magnetic field both in vertical and horizontal directions and continuously calculates the deviation from the path.

Light coding

Light coding technique can be regarded as the previous evolutional step of structured light technology. Structured light technique uses light source when cameras measure reflections. This method uses structured light source as the emitted visible light is modified and structured during radiation. The structured light is similar to a zebra or a black and white pattern. Because of the triggered light source, the measurement must be very strict and precise in timing to enrich acceptable results. Recently new techniques use blue light source instead of white light and generate 3D-point-cloud and reach higher resolution and more accurate measurements.

Light coding technique is the next step of structured light where laser source emits near infrared (NIR) light constantly and a filter transforms it into a dotted pattern. This dotted pattern is detected by an IR camera and the distance of the dots are estimated considering the distortion and other parameters of the dot cloud. The commonly known manifestation of this sensor appears in Microsoft Kinect [3].

Laser navigation system

The introduction of AGV system started in the mid 1970's at Tetra Pak manufacturing and warehouse facilities in Sweden. They wanted to improve their effectiveness and safety while moving huge paper rolls around. The speed limit of AGVs was 1m/s and it was guided by a wired floor network. In the 1980's they began testing with laser navigation and in 1991 laser navigation was implemented [1].
Laser navigation system is based on the triangulation method. Laser emits signals from the AGV and if at least three reflections come back the position and heading of the AGV can be calculated from the directions of the reflected laser beams [4].

Laser navigation systems sometimes have some disadvantages like slow response and low accuracy.

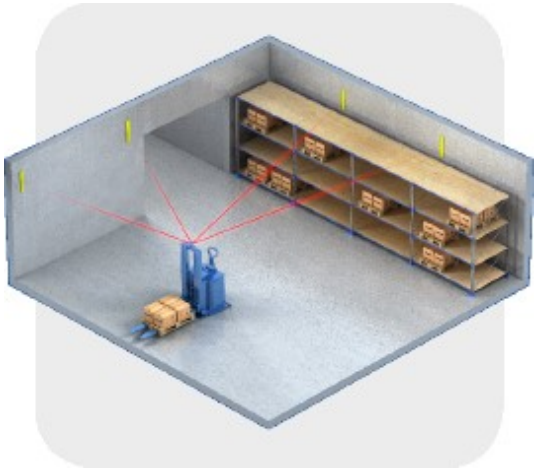


Fig. 3. Laser navigation

III. Main AGV producers and their solutions

Dematic uses four types of navigation system in their AGVs: laser navigation, magnetic tape guidance, natural navigation and wire (inductive fixed) navigation. All these navigation methods are available separately or in combination (see sensor fusion). Dematic offers software solution, too. The company uses lead acid and hydrogen fuel cell batteries and the batteries can be changed manually or automatically or can be charged without removing them from the AGV. Dematic has different types of AGV solution:

•compact AGVs – ultra compact AGVs for tight places, there are three types: Compact Tugger, Compact CB Fork, Straddle Fork

•standard AGVs – commercial standard solutions with forklift, chassis, etc.

•Hybrid AGVs – traditional forklift ruled by driver or works as an AGV

•Custom E'gv – special needs can be solved as asked by consumers (i.e. 180-degree cheese rotation during maturation)

•E'tow systems – rail in the floor guides the transport system, the in-floor chain conveyors are good for large transportation

Since 1960 Crown company has been dealing with forklifts and material handling equipment. They have 19 facilities and are present all around the world. They produce AGVs systems for warehouse automation, too.

For the navigation Crown has wire and rail guidance for minimal aisle width but AGV can become a man guided forklift on any occasion. They have solutions for low level order picking and for very narrow-aisle automation. There is an auto fence function available for TPS series. These machines are operated by lithium ion batteries with 48 or 72/80 volts depending on the series.

IKV Robot Nanchang company is located in China. Their work is based on German technique and their field of interest covers AGVs, industrial automation, intelligent warehousing and industrial controllers.

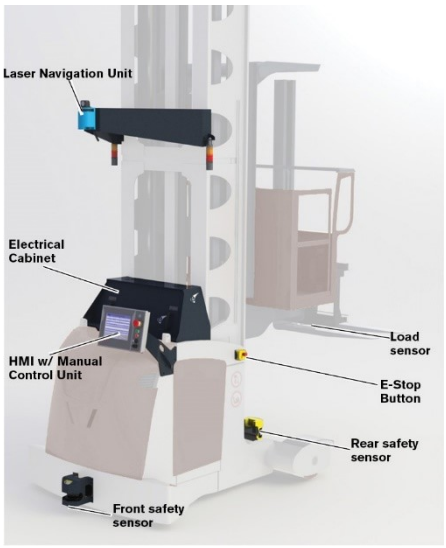


Fig. 4. Hybrid AGV [5]

The IKV Robot has the following AGVs:

•Bidirectional Roller AGV – navigates by magnetic tape guidance, photoelectric or laser guidance and it carries maximum 1000 kilograms

•Laser guided forklift – it has a 24V lead acid battery and can carry up to 1000 kilograms

•One way lurking AGV – this AGV navigates under the car or pallet and transports it. It navigates by magnetic tape-, light electrical or laser guidance and can carry up to 1000 kilograms.

•Backpack AGV – this AGV works with magnetic tape guidance system and has a 24V lead acid battery.

•Magnetic guidance forklift AGV – versatile system equipped with hydraulic lifting system and it can carry up to 2000 kilograms.

•Heavy Burden Carriers AGV – this AGV was planned for large coil and plate transport and for large assembly. This AGV navigates by magnetic tape guidance, photoelectric or laser guidance and it carries maximum 1000 kilograms.

Toyota corporation is one of the leading forklift producers of the world as they manufactured more than 500 000 forklifts just in the USA. These products are available for warehouses, manufacturing plants and retail storefront. In AGV solutions Toyota offers two systems:

•BT Staxio SAE160 AGV Autopilot Stacker – its load capacity is 2 tons and can lift up things up to 5 metres. It navigates with laser technology/natural navigation technique and has a manual mode, too.

•BT Optio OAE 120CB AGV Autopilot Order Picker – it can load 1.2 tons, lift up things to 4.1 metres and suitable for order picking purposes. ERP system and barcode scanner are integrated on-board.

•BT Reflex RAE250 AGV Autopilot Reach Truck – its load capacity is 2.5 tons and lift height is 10 metres. ERP system and barcode scanning are integrated on-board.

•BT Movit TAE500 – AGV Autopilot Tow Truck – this AGV can tow up to 5 tons, navigated by natural navigation technique and ERP system is integrated on-board.

•Tug-cart mouse AGV – this AGV can navigate in narrow aisles and is powered with 24V. It can carry about 1350 kilograms.

•L-cart AGV – this AGV is powered by 24V lead acid battery. It can carry up to 1550 kilograms and can be variously built up as necessary.

Atab is a Swedish company dealing with AGVs for 30 years. Not only the hardware but the whole solution is available for consumers. They produce forklifts, carriers and assembly AGVs. The product name of their solution is MAXAGV. Their products are categorised into three groups: FX means forklift, CX means carrier and AX means assembly. In this paper I am trying to present only the FX series.

•FX10 – there are three versions of FX10 where the difference is in lifting height and in maximum weight.

•FX15 – this forklift AGV is designed for pallets and stillages. There are also three versions of it considering the maximum lift height and the maximum weight. Battery operates this AGV for 16 hours.

•FX30 – this model is similar to FX15 but it has higher load and battery capacity.

•FX40 – this model can load pallets and racks. It has a wide massive chassis.

Rocla is a Finnish company dealing with intelligent material handling systems. The company has been dealing with automation intelligence since 1983 and since 2007 a fully modular serially manufactured AGV has been being manufactured here (AWT). This company manufactures warehouse trucks and forklifts and around 8000 AGVs have been being shipped to their partners since then. The Rocla AGVs use three types of navigations namely laser triangulation, magnetic spots and floor wire system. The AGV solutions are:

•Low lift AGV – this AGV is called ATX16 and has 2m/s driving speed, 1600 kilograms' load weight, 150 millimetres lifting height.

•Fork over lift AGV – this AGV is called ATX12, its lifting height is 1.7 metres. Automated navigation is mixed with manual picking of loads.

•Straddle AGV – its lifting height is 0.86 metres and the load weight is 2500 kilograms.

•Counterbalance AGV – this AGV is a good solution for gravity-flow racks with its 0.84 metres lifting height and 2500 kilograms' load weight.

EK Automation is a German company existing since the 1980's. This company delivered around 9000 AGVs for 1000 industrial plants. EK AGVs navigates with different methods such as natural-, optical-, laser guidance and hybrid systems are also available for special tasks. The AGVs' controller system is PLC and PC based.

Beewatec systems is a German company dealing with AGVs among others. Low cost modular systems and lean processes are the main factors of their AGVs.

•Bee spurmeisse straddle forklift transporter – this is a versatile AGV system with different load weight capacity from 1000 to 2000 kilograms.

•Bee mini – there are three subtypes in this AGV group namely Carry transporter, Pickup transporter and Underride transporter.

•Bee Truck transporter – this transporter is a compact tractor and pulls up to 4000 kilograms.

Savant company supplies AGVs and AGCs for industry, commercial and medical fields. Savant AGVs practice inertial navigation techniques and can carry maximum 20 tons. Fork AGVs can carry 2 tons. These AGVs use 'Virtual path' navigation.

•DF40 – this forklift AGV can carry around 1800 kilograms of weight and has a 600 AH battery with 8 hours working time capacity.

•DT-100 – this AGV can carry more than 20 tons and is powered by 1100 AH battery which is 2 tons.

•DC-60 – this AGV carries about 3 tons and use a 48V power system.

•DL-40 – it carries 2 tons and has a 630 AH battery.

Axter is a French company with more than 25 years' experience in mobile robotics. Their AGVs operate mainly with laser guided and natural navigation methods but sensor fusion solutions are also available depending on the need of consumers. There are three types of Axter AGVs:

•Intelligent AGC – these AGVs carry small pallets.

•Special carriers – extra needs resulted in special solutions. Size, weight or nature of products influence the shape of the AGVs.

•Standard vehicles – these AGVs include pallet trucks, stackers, tractors and reach tracks. Their maximum load weight can be up to 3.5 tons.

Sesto robotics is dealing with industrial automation and the headquarters are located in Singapore. They have AGVs and IMRs (Intelligent Mobile Robots). The company manufacture the following AGVs which use laser based navigation.

•S200 – V2 version of this AGV is a compact and can carry about 200 kilograms

•S300 – this AGV is a trolley transporter which can carry 300 kilograms

•S500 - biggest AGV of Sesto with the 500 kilograms' payload capacity.

Fig. 5. Sesto 59S500 [6]

Aerocom is an English company since 2000 and is dealing with AGVs and pneumatic tube systems in warehouses. Their AGVs are not off-the-shelf but can be adapted to the requirements. Their AGVs are the followings:

•Phoenix series – these AGVs include high-lift guided fork trucks, high shift stackers, etc.

•Caesar series – these AGVs include drive-under tractors, platform trucks with roller conveyor, heavy-duty transporters and outdoor vehicles which operate down to -20 degree Celsius.

•Mayesto series – this group consists of high-rack stackers

•Motormouse series – these AGVs are supplied with different load handling components

Jungheinrich is a German manufacturer which has been dealing with industrial and warehouse intralogistics since the 1960's. In their AGV systems the company uses laser based navigation. These AGVs are based on a serial manufactured machines which are equipped with special techniques. The automated solutions are the followings:

•ERC215a – this stacker truck AGV lift up height up to 4 metres and can carry 1.5 tons

•ERE225a - this stacker truck AGV lift up height up to 2.4 metres and can carry 2.5 tons

•EZS350a - this AGV can tow up to 5 tons and can be used in narrow aisles

•EKS215a – this AGV can lift loads up to 6 metres and the maximum payload is 1.5 tons

Linde is a material handling company with around 13000 people worldwide. The company was founded in 1904. Linde is a leading company in forklift truck manufacturing with 80 series and 20 model variants.

Linde AGVs determinate their location by natural navigation. The Linde Matic range are the followings:

•P-Matic tow tractor – this AGV can carry 5 tons and its battery operates with 24V 375Ah

•L-Matic pallet stacker – this AGV can lift up weight up to 1.9 metres and the payload capacity is 1200 kilograms.

•L-Matic AC counterbalanced pallet stacker – this AGV can lift up weight up to 1.9 metres and the payload capacity is 1200 kilograms.

•T-Matic pallet truck – this AGV can lift up 3 tons up to 1.2 metres.

•K-Matic narrow aisle truck - this AGV can lift up 1 ton up to 7.2 metres.

SEW was founded in 1931 in Germany. The company started with electric motors and gearmotors. The company is the market leader in drive automation industry and also dealing with AGV solutions.

•The Maxolution up to 500 kilograms transporting vehicle can carry up to 500 kilograms and it can navigate considering the local requirements: RFID, inductive fixed guideline, magnetic tape guidance or SLAM (simultaneous localisation and mapping).

•The Maxolution up to 1500 kilograms transporting vehicle can carry up to 1500 kilograms. With additional componenets it can accept pallets, boxes or frames and can navigate with the maximum tolerance of 10 milimetres.

Agilox is an Austrian company dealing with supply chain management and intralogistics. The company was founded in 2008. They manufacture intelligent guided vehicles. The Agilox system is a wireless LAN based service.

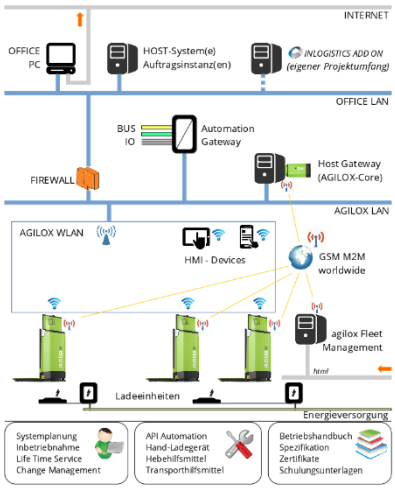


Fig. 6. Agilox navigation system [7]

IV. Conclusion

This paper was to show the autonomous guided vehicles, their navigation possibilities and techniques and the main AGV manufacturers and their solutions.

As a conclusion we can summarize that the navigation techniques are in continuous development, during navigation process less signals and markers or guides are used and the recognition of the environment by sensors come more and more into view.

Some special industrial applications or warehouse solutions require wired path installation to the floor but generally these techniques are not up-to-date, not flexible at all and cost a lot.

Sensor fusion can improve navigation accuracy on one hand and on the other hand it can be used for an environment where both inner and outside navigation are necessary.

Main AGV providers and their solutions are also introduced in this paper. There are many AGVs and navigation possibilities available on the market and most of the AVG manufacturers offer not only the machines and guidance systems but the software background as a complete solution, too.

References

[1] www.ndcsolutions.com

[2] www.jbtc.com

[3] L. Pérez, i. Rodríguez, N. Rodríguez, R. Usamentiaga, and D. F. García, Robot Guidance Using Machine Vision Techniques in Industrial Environments: A Comparative Review, Sensors, vol. 16, p. 335, Mar. 2016.

[4] N. Ida, Y. Tanaka, and Y. Satoh, Development of the Autonomous Guided Vehicle System Using Laser Navigation Method, in Robotics, Mechatronics and Manufacturing Systems (T. Takamori and K. Tsuchiya, eds.), pp. 637–641, Amsterdam: Elsevier, 1993.

[5] AGV car, AGV system, automatic guided carts, forklift AGV-IKV Robot – www.ikvrobot.com

[6] www.sestorobotics.com

[7] www.agilox.com