AIS 2020

15th International Symposium on Applied Informatics and Related Areas organized in the frame of Hungarian Science Festival 2020 by Óbuda University

PROCEEDINGS

November 12, 2020
Székesfehérvár, Hungary

ISBN 978-963-449-209-2

Automated Robotic Manufacturing Cell in Research and Education

Máté Péter Turcsányi
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
turcsimate@gmail.com

Levente Vadász
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
levente.vadasz997@gmail.com

Márta Seebauer
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
seebauer.marta@amk.uni-obuda.hu
ORCID: 0000-0001-8459-8975

Károly Széll
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
szell.karoly@amk.uni-obuda.hu
ORCID: 0000-0001-7499-5643

***Abstract*—As industry is applying more and more robotic solutions it becomes also necessary to improve the theoretical and practical education of automated cells on university level. The paper proposes a research and educational environment in which the students can safely get acquainted with the basics of modern automation. The main hardware components of the robotic cell are an ABB IRB 1400 6-axis industrial robotic arm, a KUKA KR-16 6-axis industrial robotic arm and an oval conveyor belt.**

Keywords—robotic assembly, educational robots, industry 4.0, digital education

I. Introduction

As digital education is increasingly gaining ground [1]–[3], state-of-the-art topics like robust engineering solutions [4]–[6], predictive maintenance [7]–[10], industry 4.0 [11], [12] or sensor technology [13]–[15] need new concepts for the structure of the university curriculum. Thus, a further challenge of this aspect is the modification of the laboratories to meet these new requirements.

The need for automation is becoming more and more important in engineering life. Even the simplest process requires extreme precision and attention. This paper proposes an idea of designing a system that can be programmed at different levels. As well as to create a solid foundation that corresponds to larger-scale further developments and ideas for research purposes.

Thanks to the developments of the Robotics Center at the Alba Regia Technical Faculty of Obuda University, the engineers studying at the institution can access industrial solutions during their studies. A universal automated cell designed in the robot laboratory can be used by any student at the university to expand his knowledge, test his ideas and inventions. The automated cell includes an ABB and a KUKA industrial robot arm, a large and a small industrial assembly line, laser scanners and different communication protocols. The elements of the cell work together to create a system that requires mechatronic knowledge to operate, thus providing a workspace for electrical, computer and mechanical engineers and undergraduate mechatronic engineers.

II. Problem Description

During the design, the following aspects were considered:

- Possibilities provided by the available tools
- Acquisition of necessary tools
- Implementation of communication protocols
- Integration into various university courses

The opportunities provided by the available tools were quite limited. It was mostly data blockage that caused this hurdle. The robotic arms were the greatest challenge in this aspect, as they use DeviceNet as communication protocol, thus their controller is not accessible. In terms of the conveyor belt the PLC had to be replaced and the great number of input and output signal had to be handled.

III. Structure of the Environment

The automated cell (see Fig. 1) consists of 2 industrial robot arms, a conveyor belt controlled by PLC, a small conveyor belt and safety laser scanners. We chose Siemens and Balluff products for the conveyor belt by merging TIA Portal and IO-Link. Here we had to consider the conditions for the creation of an instructor system that is more versatile in terms of programmability.

Fig. 1: Experimental setup (Picture of the system)

A. KUKA industrial robot arm

Technical data of the KUKA robot arm (see Fig. 2):

- Maximum reach 1612 mm
- Maximum payload 20 kg
- Rated payload 16 kg
- Pose repeatability (ISO 9283) ± 0.04 mm
- Controller KR C2

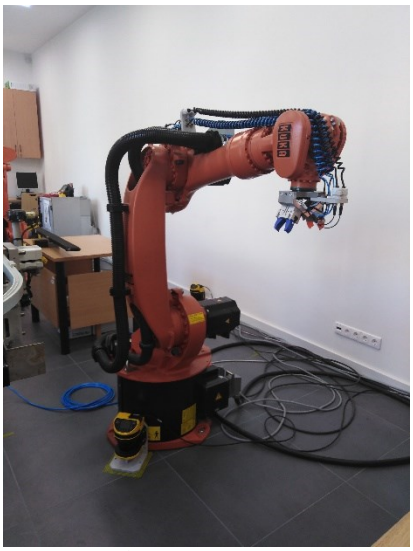


Fig. 2: KUKA KR-16

B. ABB Industrial Robot Arm

Technical data of the ABB robot arm (see Fig. 3):

- Maximum payload: 5 kg
- Maximum reach: 1.44 m
- Integrated signal supply 12 signals on upper arm
- Integrated air supply Max. 8 bar on upper arm
- Position repeatability 0.05 mm (average result from ISO test)
- Max. TCP velocity 2.1 m/s
- Continuous rotation of axis 6

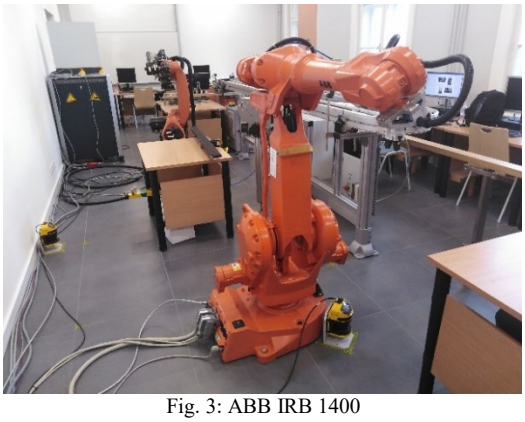


Fig. 3: ABB IRB 1400

C. Conveyor Belts

The greater conveyor belt is is an oval track (see Fig. 4). It currently has four stations, two operator-side and two robot-side sections. The stations are equipped with different sensors according to their side. The robot side is equipped with several sensors, providing data and information about the workpiece's actual position. In case a sensor fails, we get feedback, which makes the system more stable. The conveyor can be controlled manually as a separate unit with its own manual unit or in automatic mode via the PLC. The belt can also run backwards.

Fig. 4: Oval conveyor belt

The smaller conveyor (see Fig. 5) is powered by an electric servo drive. Depending on the output signal of the control unit, it is activated or inactivated. Its job is to transport a workpiece between the two robots independently of the oval assembly line.

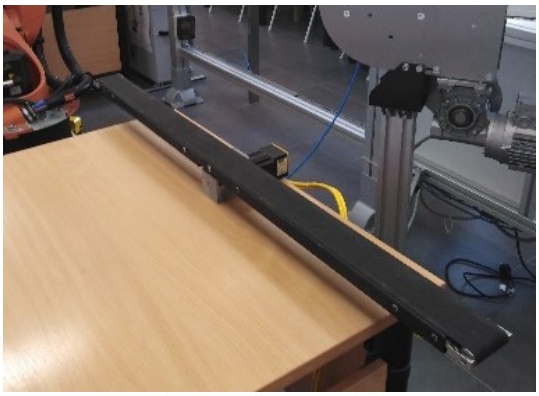


Fig. 5: Small conveyor belt

The entire belt is electronically and electropneumatically controlled (see Fig. 6). The feedback on the status of the actuators are solved via the REED relays.

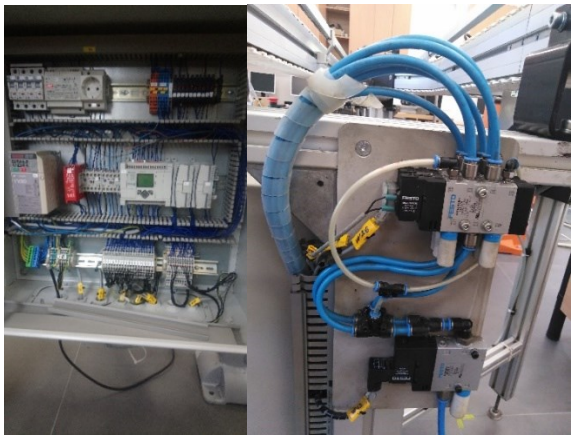


Fig. 6: Electrical cabinet of the conveyor and valve island

It is important that the work tray is always returned to the same points where the robots are set. Stopping the belt each time is not an option in this case. If several trays are running on the belt at the same time, switching the frequency converter on and off can also reduce the lifetime of the motor. Therefore, stopping the tray on the track is solved via pneumatic actuators (see Fig. 7). The tray has a unique design, a negative pattern has been etched from its sides so that when it reaches the given point, the sensors detect it, the actuator equipped with a cylinder stops the tray and then a side clamping cylinder fixes its position. This situation persists until there is a response from the robot and the sensors on that side that the workpiece has been removed. If the condition is met, the actuator releases the tray.

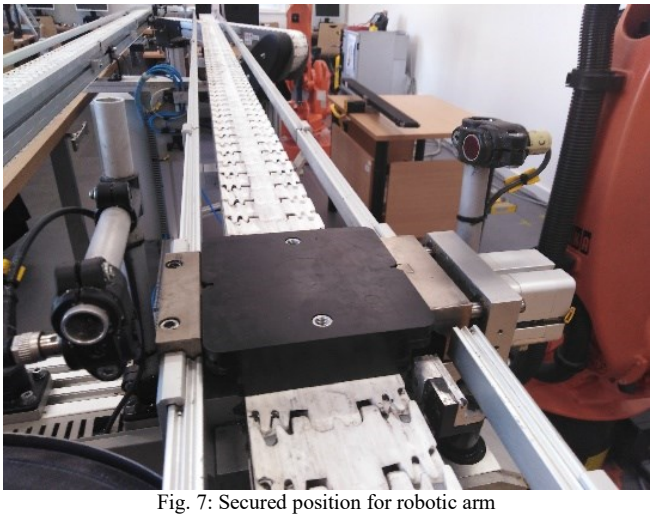


Fig. 7: Secured position for robotic arm

On the operator side, it is possible to remove the workpiece and inspect it. The tray is stopped via a simple pneumatic actuator (see Fig. 8). This is also important from a quality management point of view.

Fig. 8: Position for operator

D. Safety Scope

The safety circuit consists of emergency stop pushbuttons located in the cell and on the robots' teach pendant, and laser scanners. These are handled by a PILZ safety PLC (see Fig. 9). If the conditions are met, i.e. no emergency stop button is active, the devices are not in fault state, the laser scanners are active and no signal is given, the safety circuit is set up. This signal is retrieved by the control unit and transmitted to all devices. This also provides a secondary security. When this circuit is interrupted, the system must stop.

Fig. 9: Electrical cabinet of the safety PLC

E. Network Communication

The network communication of the robotic cell (see Fig. 10) is based on the following protocols:

- PROFINET
- DeviceNet
- Ethernet/IP

PROFINET is an Industrial Communication Protocol created by Siemens. The term PROFINET stands for: PROcess FIeld NET. Which means it is Ethernet based, using RJ45 ports and cables to connect different types of devices equipped with the PROFINET interface. The dataflow of this communication is 100 megabits per second, and the response time is less than 1 millisecond, which includes good results for high speed applications. To connect several devices, a simple switch can be applied. In the case of PROFINET we differentiate three types of addresses:

• Device Name: The name of the device that is connecting to the PROFINET

• IP address: The communication surface

• MAC address: The physical address of the device

The critical ones in this setup are the Device Name and the IP address (mostly a simple IPv4 address with a submask is enough). If a device is malfunctioning it can be replaced easily, that is the reason why the MAC address is not that important in this setup.

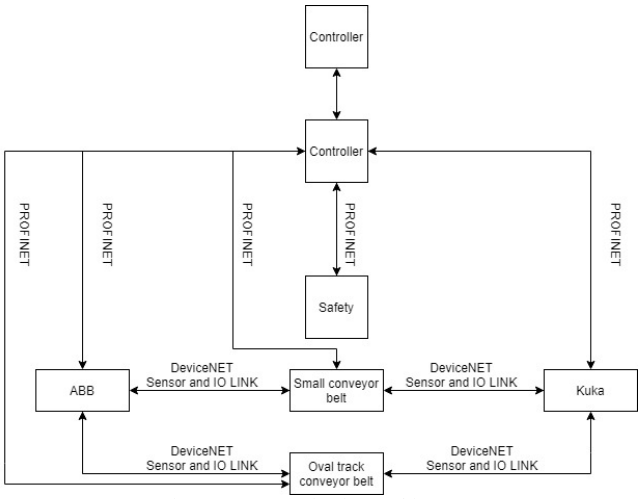


Fig. 10: Communication architecture

IO-Link is a relatively new technology. It transmits data and signals to the control unit using point-to-point communication. Should any problem occur, whether a drive failure or a sensor contact fault, the IO-Link Master transmits this fault to the control unit, making troubleshooting faster and easier. It performs a diagnostic test while the system is operating so that there is no data conflict between the alarms in addition to other digital signals. This diagnostic can be set manually to see what specific information we want to extract. Different tolerance fields and details can be set. Larger damages and losses can also be prevented as a result.

The cell has been designed to provide control of the elements by various means. Currently, the system is controlled by a PLC that feeds data to the main controller, (which can be another PLC or) in this case a computer, by using Node-RED or/and ROS.

Controlled by a computer, all elements of the cell are connected to a local network using PROFINET, with a dedicated PC as the main controller. The main program runs in Node-RED, which handles all incoming and outgoing signals from the PLC. It is also easier to simulate individual events in a computer environment, so it can be used to test different systems in the cell. Thanks to its user-friendly interface, we can process large amounts of information in a short time.

The Robot Operating System (ROS) provides accurate information about the status and position of robots. ROS is an open source platform for building robotic software. The platform provides a wide range of development support, including a 3D simulator that can simulate the behavior of the robot and the environment in detail. Such a simulator can also be used to simplify testing, as it does not require the actual robotic hardware and the creation of a real environment, the basic operation can also be checked in a simulated world.

IV. Operating the Environment

When the cell is powered on, it can be started automatically by pressing a button. The cell stops working immediately if there is no communication connection with one of its elements or the signal of the security circuit is lost. The security circuit checks the status of the equipment, any program errors, and the status of the laser scanners.

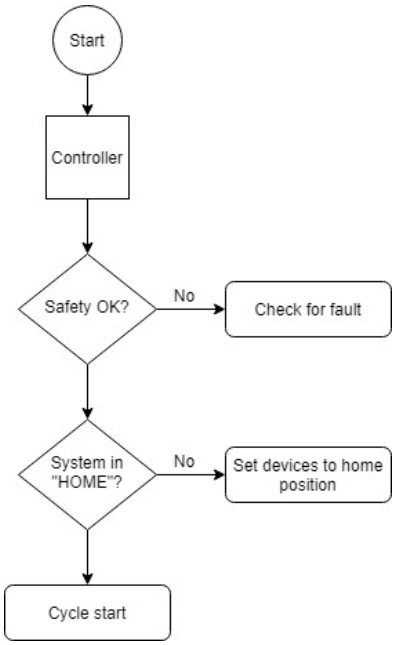


Fig. 11: Process architecture

There are 4 workstations on the large assembly line. At the first workstation, the KUKA robot arm places the workpiece on the belt on a work tray secured by a pneumatic side clamp. When the robotic arm leaves the work area, the conveyor starts and transports the work tray to 2 operator stations, where the controller can stop the workpiece using pneumatic cylinders and sensors if certain conditions are met. When the work tray reaches the fourth and final station, the pneumatic side clamp fixes the tray and the ABB robot arm takes the workpiece off and places it on the small conveyor belt, which returns it to the first robot and the process can start all over again.

It is a standard industrial practice, for example, to take samples for quality control. The number of cycles can be calculated using the sensors. In the controlling program it can be set for example to every 10th element to stop at the first operator station where the workpiece must be checked and continue the process in case of "OK" feedback, otherwise return the tray empty without the workpiece.

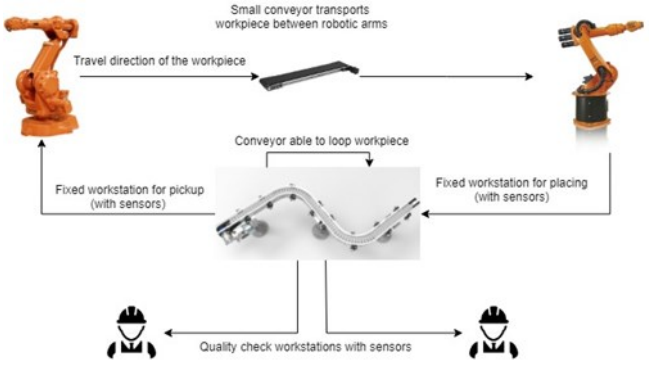


Fig. 12: Workflow

V. Conclusion

The paper introduced a robotic cell for university studies and research purposes. The purpose of the cell is to simulate a complex industrial environment where the methods and tools used in today's practice can be found. This, together with the security system and network communication solutions,

creates an environment in which the problems and solutions of almost every branch of engineering can be found.

ACKNOWLEDGMENT

The authors thankfully acknowledge the financial support of this work by the Howmet Aerospace Foundation, the Hungarian State and the European Union under the EFOP-3.6.1-16-2016-00010 and 2019-1-3-1-KK-2019-00007 projects.

The research has been supported by the NTP-HHTDK-19-022 grant. The authors declare no conflict of interest. The authors would like to express their gratitude to those who have provided their help in the various experiments.

REFERENCES

[1] M. Pogatsnik, "Measuring Problem Solving Skills of Informatics and Engineering Students," in *IEEE Joint 19th International Symposium on Computational Intelligence and Informatics and 7th International Conference on Recent Achievements in Mechatronics, Automation, Computer Sciences and Robotics, CINTI-MACRo 2019 - Proceedings*, Nov. 2019, pp. 93–98, doi: 10.1109/CINTI-MACRo49179.2019.9105277.

[2] M. Pogátsnik, "Életpálya-építés serdülő- és ifjúkorban - kutatás közben," *Egyéni Kül. Szerepe Tanulásban És Pályaválasztásban*, pp. 149–162, 2015.

[3] M. Pogatsnik and R. B. Kendrovics, "Communication and Reading Comprehension among Informatics and Engineering Students," in *SAMI 2020 - IEEE 18th World Symposium on Applied Machine Intelligence and Informatics, Proceedings*, Jan. 2020, pp. 235–240, doi: 10.1109/SAMI48414.2020.9108764.

[4] G. Györök and B. Beszédes, "Fault tolerant power supply systems," in *11th International Symposium on Applied Informatics and Related Areas (AIS 2016)*, 2018, pp. 68–73.

[5] G. Györök and B. Beszédes, "Highly reliable data logging in embedded systems," in *2018 IEEE 16th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, Feb. 2018, pp. 49–54, doi: 10.1109/SAMI.2018.8323985.

[6] G. Györök and B. Beszédes, "Adaptive Optocoupler Degradation Compensation in Isolated Feedback Loops," *2018 IEEE 12th Int. Symp. Appl. Comput. Intell. Inform. SACI*, pp. 167–172, 2018.

[7] G. Manhertz and Á. Antal, "The effect of air-fuel equivalence ratio change on the vibration components of an internal-combustion engine," *RECENT Innov. Mechatron. 2 1-2*, pp. 1–6, 2015, doi: 10.17667/riim.2015.1-2/13.

[8] G. Manhertz and Á. Bereczky, "Development of a vibration expert system to analyze and predict malfunctions in internal-combustion engine," in *Proceedings of ARES'14 : Workshop on Application of Robotics for Enhanced Security*, 2014, pp. 24–29.

[9] G. Manhertz, D. Modok, and Á. Bereczky, "Evaluation of short-time fourier-transformation spectrograms derived from the vibration measurement of internal-combustion engines," in *2016 IEEE International Power Electronics and Motion Control Conference (PEMC)*, Sep. 2016, pp. 812–817, doi: 10.1109/EPEPEMC.2016.7752098.

[10] G. Manhertz, G. Gardonyi, and G. Por, "Managing measured vibration data for malfunction detection of an assembled mechanical coupling," *Int. J. Adv. Manuf. Technol.*, vol. 75, no. 5, pp. 693–703, Nov. 2014, doi: 10.1007/s00170-014-6138-3.

[11] É. Hajnal, "Big Data Overview and Connected Research at Óbuda University Alba Regia Technical Faculty," in *AIS 2018 - 13th International Symposium on Applied Informatics and Related Area*, 2018, pp. 1–4.

[12] A. Selmeci and T. Orosz, "Usage of SOA and BPM changes the roles and the way of thinking in development," in *2012 IEEE 10th Jubilee International Symposium on Intelligent Systems and Informatics*, Sep. 2012, pp. 265–271, doi: 10.1109/SISY.2012.6339526.

[13] Z. Pentek, T. Hiller, and A. Czmerk, "Algorithmic Enhancement of Automotive MEMS Gyroscopes with Consumer-Type Redundancy," *IEEE Sens. J.*, pp. 1–1, Aug. 2020, doi: 10.1109/jsen.2020.3017094.

[14] Z. Pentek, T. Hiller, T. Liewald, B. Kuhlmann, and A. Czmerk, "IMU-based mounting parameter estimation on construction vehicles," in *2017 DGON Inertial Sensors and Systems, ISS 2017 - Proceedings*, Dec. 2017, vol. 2017-December, pp. 1–14, doi: 10.1109/InertialSensors.2017.8171504.

[15] T. Hiller, Z. Pentek, J. T. Liewald, A. Buhmann, and H. Roth, "Origins and Mechanisms of Bias Instability Noise in a Three-Axis Mode-Matched MEMS Gyroscope," *J. Microelectromechanical Syst.*, vol. 28, no. 4, pp. 586–596, Aug. 2019, doi: 10.1109/JMEMS.2019.2921607.

The impact of digital education from a teacher's perspective

Judit Módné Takács
Alba Regia Technical Faculty
Obuda University
Szekesfehervar, Hungary
modne.t.judit@amk.uni-obuda.hu

Monika Pogatsnik
Alba Regia Technical Faculty
Obuda University
Szekesfehervar, Hungary
pogatsnik.monika@amk.uni-obuda.hu

***Abstract*— In our study, we present a survey we conducted on the experiences of the transition from traditional education to digital teaching from the perspective of teachers.**

We examined the availability of the necessary tools for digital teaching (computer, web camera, microphone) and internet access for teachers and students. Several issues have been assessed: the time required for daily preparation and teaching, the need for uniform program use within a given educational institution. The experiences on the pros and cons of digital education have been gathered. We mapped the motivational tools used to maintain attention in online education. We collected teachers' suggestions for missing teaching materials.

Keywords—digital education, teaching, motivation, maintain attention.

I. INTRODUCTION

As a result of the epidemic situation, there is an explosive methodological renewal in the Hungarian education. The transition to distance learning has necessarily been rapid, with most educational institutions successfully tackling the challenge. In order to implement the sudden transition effectively, the official bodies in Hungary made several recommendations and collected useful materials for schools and teachers, yet extremely diverse methodological solutions were used in the months following the introduction of the digital curriculum in the spring term of 2020, exploiting the potential of distance education at different levels and qualities [1]. The traditional Prussian frontal classroom teaching, where the teacher is the sole custodian of knowledge and the learner is responsible for sitting in a disciplined position and memorize the lesson - has suddenly entered the 21st century information society, where the emphasis is on knowledge creation technology, information acquisition, processing and sharing [2].

The potential for innovation in distance education is the most important impetus for the renewal of education and training systems than ever before [3]. According to a trend summary analysis by Sharpe [4], during the epidemic period, there was a significant increase in innovative activity in dealing with problems. Teleworking did not mean that work performance deteriorated, creative solutions and new communication techniques were invented, and they even helped to strengthen social relations.

Digital education is one of the most influential factor in both teaching and learning, depending on the digital tools available, the level of digital competence of the teacher and the learner. The efficiency of educational work depends on whether the scheduled, content-based and time-limited curriculum is handed over and mastered. Can the teacher, as a "transmitter", use the available tools, apply them in effective, efficient teaching, and can the learner / student as a "receiver" master the digital content [5].

The teaching community has shown unprecedented cooperation [6], [7]. Over just one weekend, thousands of Facebook groups were formed. The sharing of knowledge and curriculum has evolved to a tremendous extent in a few days. The specialization of the groups took place in a short time, a very close collaboration has developed to maintain education, and as a result, the transition to digital distance learning has been possible in a few days. Based on the economic situation of the country's regions, the necessary tools for the implementation of digital education were not available everywhere. Financial opportunities and daily livelihood difficulties limited the receptive base [5].

In our research, we assessed teachers in different educational institutions about the solutions they use, what tools they have access to, how they try to maintain education, and motivate students.

II. METHOD

A. Procedure and instruments

A survey was organized among the partner schools of the University of Óbuda regarding the feasibility of online education within the EFOP-3.4.4-16-2017-00019- STEM "Strategic Developments of the University of Óbuda" project. The survey tool was an anonymous, online questionnaire with a pre-recorded set of questions. Regarding the types of questions, there were open-ended questions, where the respondent could also express their own opinion, and closed-ended questions, among which there were mainly selective question types with several choices.

The aim of the research is how well the teachers of the schools participating in the survey felt organizational, technically, personally for this task, and what kind of experience they gained and what attitude they had towards digital education.

B. Participants

The 214 participating teachers from different educational institutions, also from various regions within the country (mostly from Fejér county and Borsod-Abaúj-Zemplén county) took part in the survey. In terms of distribution, the majority of responses, 61%, came from secondary education institutions.

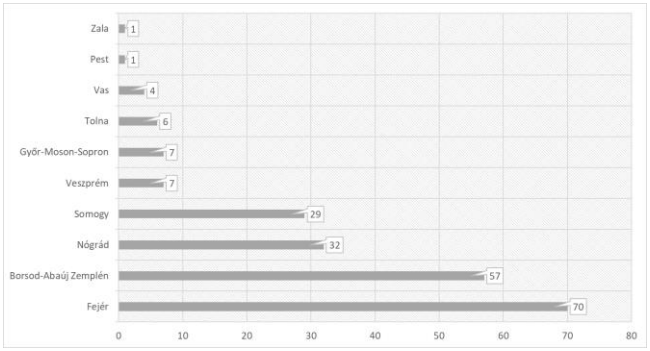


Fig. 1. Sample of responses by counties

The distribution of participants in the questionnaire by profession showed that nearly 45% of the respondents teach in the human field, 35% in STEM (Science, Technology, Engineering, Math) and 20% in primary school as a general teacher.

It follows that the research was mainly assisted by educators who are not directly related to the field of informatics, which made their work much more difficult in the transition to online education, as they only have the IT knowledge necessary for everyday life.

They had to learn to use digital tools and applications in a short time, such as video editor, image editor, online test manager, more accurate knowledge of various cloud services, and more familiar with the platform supporting more education in the initial period.

III. RESULTS

A. *Experience with the availability of IT supplies for online education*

The transition to online education has created an adequately new life situation. Educators also had to adapt to a new way of teaching. However, this also required the strengthening of different competencies.

Examining the availability of essential IT tools for online education at the start of the new form of education, 95% of respondents had the necessary digital tools that could be suitable for implementing online education (laptop, personal computer, smartphone, tablet). 36% of the respondents were provided with the necessary IT equipment by the educational institution, while the rest had to carry out this transition from their resources.

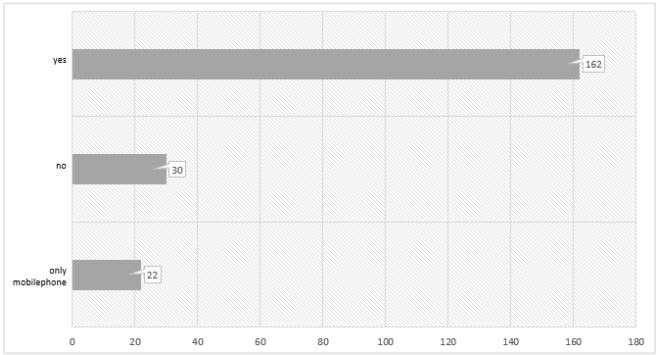


Fig. 2. Sample response with the availability of the appropriate IT tool

In traditional education, the use of digital tools, especially by humanities teachers, cannot be said to be too typical, while in this new life situation it has become essential and has mostly required the use of digital competencies.

Another interesting fact is that a fully online education, which requires the use of audio-visual devices, ¾ respondents had only the means to transmit/display visual and audio data, as the rest could only fill these gaps by using their smartphones.

B. *Teachers'workload experiences over time*

Traditional education takes place at a fixed time, while online education has been a full-time activity in many cases. This situation made it necessary for everyone involved (teacher, parent, student) to have a new kind of time management.

Traditional, fixed-time, 45-minute lessons have been replaced by more freely organized, more flexible education, but this has strongly required teacher-parent collaboration, especially in the younger age group, and increased teacher-student collaboration in the older age group.

Mutual responsibility came into focus, as ordinarily, education has been mostly the responsibility of the teacher, with the parent only providing a supportive, helpful function. In the new situation, the teacher and the parent shared the teaching practice.

In many cases, this task was problematic to achieve with the parents' modified work schedule and way of working.

In terms of their workload, the new form of education also meant additional tasks for teachers. 68% of the respondents worked more than 5 hours, while 22% of these teachers more than 10 hours a day.

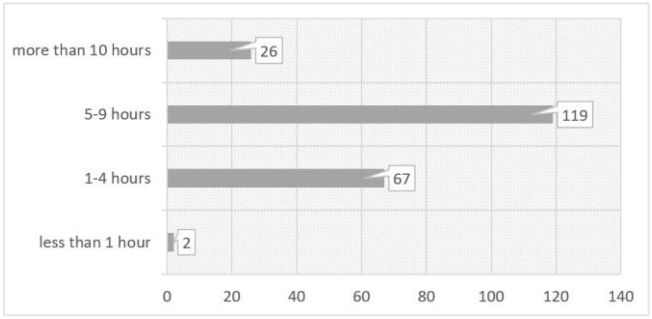


Fig. 3. Time spent per day preparing and presenting digital lessons

Experience has shown that in the traditional work schedule, the family life of teachers is separated from education, while in this situation, education and family life are present at the same time.

As a result, other burdens also appeared in the lives of teachers, especially in families with small children, where they also had to take part in the education of their children.

C. *Examining the existence and possible lack of online educational materials*

The curricula, textbooks and lesson plans required for traditional classroom education were already available to experienced educators who had been teaching for several years. However, in the online world, new methods and new types of educational materials would have been needed, which were limited.

Existing online learning materials either did not or only partially fit into the required framework curricula, training system or were not available in the required language.

Teachers formed online groups on different platforms, which tried to fill these gaps and help each other by supporting each other. Nevertheless, the research found that 93% of respondents would find it useful to create and publish an online knowledge base.

The majority of survey participants, more than 4/5, highlighted the lack of instructional videos available in their native language. Teachers would find it extremely useful in their work to have a collection of these instructional videos gathered, organized on a variety of topics, accessible to all.

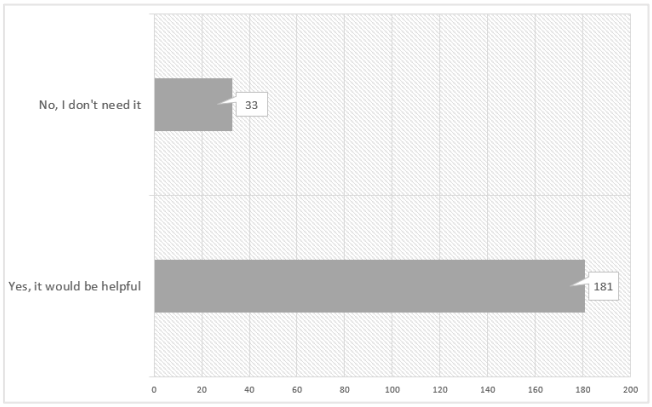


Fig. 4. The usefulness of an online course supporting digital education

Also, nearly 80% of participants would find online training in support of digital education useful. Even educators teaching in the field of informatics want to strengthen their competencies for online education, both professionally and methodologically.

D. *Experience the benefits of digital education.*

At the end of the last school year, a completely unusual situation arose due to the Covid-19 virus. In essence, schools, parents, children had to switch from traditional education to online education overnight.

Below we present the positive experiences of the educators involved in what areas of this transition were affected in practice.

One of the substantial benefits of digital education was that educators were able to learn and try new methods. This new interface required a kind of change of attitude the individuals involved went through huge digital developments and were able to try out innovative ideas.

On the one hand, online education forced the instructors to use digital tools, get to know the online teaching materials and try out and use new programs and applications for them. On the other hand, methodological diversity characterizes this period.

In traditional education, students are present in person, so discipline is strongly present, teachers can easily follow metacommunication cues. The presence of students in online education depended on the availability of the IT tools available to them and their proficiency in the digital world. Discipline was less common in education, and classroom aggression had no place in this form. This side of teacher work also became much smoother, as there was no need to discipline, it resulted in less conflict.

According to some respondents, more modest students were more easily influenced in this medium because they were not necessarily heard by others, thus putting less pressure on them.

Educators reported about weaker students to perform several times better in online education than in school, where they tend to "pass the opportunity" to active, better-able students. Everyone can work at their own pace. Children will become independent, they will be more informed, and they will utilize and expand their IT knowledge. Metacommunication cues also remained traceable with appropriate program selection and use.

Several reported that using online, chat lessons made it feel like they had moved in with their families. They were able to gain insight into students 'learning environments. Educators were able to experience how and in what way parents support and help their children. With younger children, parents can spend more time with their children, pay more attention to them and develop their digital competencies, while encouraging older children to work independently, be creative, seek individual solutions while their digital competencies develop much faster.

Some have also reported motivation-enhancing effects as a result of a more playful method of curriculum delivery closer to students. At the same time, it is easier to spot inactive and active performance students, and you can give individual feedback more than once.

The great advantage of online teaching is that it gets used to regularity and accuracy, as it requires concise wording, and educators have highlighted that the new online space teaches stakeholders about cultured digital communication.

E. *Negative experiences of digital education*

In the following, we present the negative experiences of educators involved in online education and the areas affected by this transition in practice.

There are sharper opinions and differences about the disadvantages. These differences are likely to be influenced by problems arising from regions with different economic characteristics, or a humanities teacher may face different problem types than a real education teacher in digital education. Besides, negative feelings and experiences can be strengthened by age specifics.

There was a great deal of agreement among teachers regarding the problems with the examinations. The distribution and control of dissertations caused quite a headache in the form of online education. The instructors could not be sure about the students, the proportion of their answers, or the independence of their knowledge. In addition

to correcting, storing, and preparing online tests is much more complicated and time-consuming than using a printed, available worksheet produced in previous years. Evaluation is problematic, looking from a few perspectives - as almost only online tests are available, and it does not differentiate enough. There is no way and enough time to listen to individual reports.

Another problem area is the lack of community and its positive effects. It was impossible or difficult to contact more students at all. One reason for this may have been students who did not have the IT tools needed for digital learning or had to share these tools within the family, so they had limited access to it. It was more effortful to motivate them to work together, and it became necessary for a person within the family to monitor their work and the student's presence in class. It was problematic to establish the necessary contact with disadvantaged children and to aid the learning process at a distance.

Overall, all parties lacked a personal presence, with many educators also listing the lack of physical presence of students, the importance of eye contact, and the lack of body language as natural causes of traditional education as reasons for personal motivation.

F. Motivational tools used to maintain attention during online education

The last area we wanted to explore is what motivational tools the pedagogical society uses in this online space to keep children's attention and for successful teacher-student collaboration.

Opinions on the success of maintaining motivation were divided, with some respondents reporting that they had not been able to find the right tools and felt left alone without help in an impossible life situation. That is probably why many people would take part in training in support of digital education. Not only because of the possible lack of digital competencies but also because of the methodological shortcomings that would help teachers in this situation.

Educators who had positive experiences with student motivation reported the following.

Educators can easily keep children's attention with playful tasks, joyful learning experiences, positive feedback, humour, and empathy. Many have highlighted the importance of personal, possibly individual, conversations through various online channels. The importance of parental support in this unconventional process is significant.

Overall, punishment, bad grades, and inconsistency have led to results in almost no one in this process, as creating a stress-relieving atmosphere, encouragement, and positive motivation are the only weapons for teachers in this situation as well.

IV. SUMMARY AND FURTHER RESEARCH

The research examined several issues related to the introduction of emergency online education. Only 35% of the teachers involved in the research were provided by their school with an IT tool to implement their education, which is a very low rate. Of the 214 participants, 22 could communicate with students in real time via their mobile phones, and no other image capture device was available for them.

There is a lack of motivational tools in digital education, and even IT teachers expressed their need for further training because they have a low knowledge of methods.

Further elements of our research examine the operation of digital education from other perspectives, and the results of surveys conducted among parents and students will be published later.

A further area of research would examine the impact of the semester spent in digital education on the normal education system. What are the elements that are incorporated into classroom education based on the experience of digital education? Based on feedback from students, parents, and co-workers, which methods were considered good, what is worth to keep, and what they would change.

REFERENCES

[1] Czirfusz, D., Misley, H., Horváth, L. (2020): A digitális munkarend tapasztalatai a magyar közoktatásban. Opus et Educatio, Vol 7, No 3. http://opuseteducatio.hu/index.php/opusHU/article/view/394/685

[2] Castells, M. (2005): A hálózati társadalom kialakulása. Az információ kora I. Gazdaság, társadalom, kultúra. Budapest: Gondolat - Infónia.

[3] Benedek, A. (2020): Távoktatás másként!!! – A digitális kor pedagógiai kihívásaihoz. Opus et Educatio, Vol 7, No 3. http://opuseteducatio.hu/index.php/opusHU/article/view/387

[4] Sharpe, R. (2020): Leading and learning through uncertainty. Universitiy World News, 2020.07.11. https://www.universityworldnews.com/post.php?story=20200708171028609

[5] Proháczik, Á. (2020): A tantermi és az on-line oktatás (tanítás és tanulás) összehasonlító elemzése. Opus et Educatio, Vol 7, No 3. http://opuseteducatio.hu/index.php/opusHU/article/view/390

[6] Széll Károly, Manhertz Gábor (2016): Development of a web-based training module in robotics In: Korondi, Péter (szerk.) 2016 International Symposium on Small-scale Intelligent Manufacturing Systems (SIMS) Piscataway (NJ), Amerikai Egyesült Államok : IEEE, (2016) pp. 1-6. , 6 p.

[7] András Czmerk, Károly Széll, Péter Korondi (2013): Development of a servo-pneumatic system in distant learning In: Korondi, Péter (szerk.) Proceedings of CERiS'13 - Workshop on Cognitive and Eto-Robotics in iSpace, Budapest, Magyarország : BUTE Department of Mechatronics, Optics and Mechanical Engineering Informatics, (2013) pp. 50-54. , 5 p.

.

MIG MAG Robot Welding in Research and Education

István Elek Kiss
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
istvanelekkiss@gmail.com

Bence Gábor Gyöngyösi
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
gyongyoska12@gmail.com

Márta Seebauer
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
seebauer.marta@amk.uni-obuda.hu
ORCID: 0000-0001-8459-8975

Károly Széll
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
szell.karoly@amk.uni-obuda.hu
ORCID: 0000-0001-7499-5643

***Abstract*— As the number of welding experts as well as the cost of automation is decreasing, the application of robotics is gaining significance in this field. The aim of the paper is to introduce the automation of the production of a particular part as a use case for university studies and experimental setup for further research. The part was previously hand-welded, but the increasing annual order number made automation necessary. The paper introduces a new welding jig utilizing a robotic arm.**

Keywords—MIG MAG welding, robot welding, welding design, welding programming, digital education

I. INTRODUCTION

As digital education is increasingly gaining ground [1]–[3], state-of-the-art topics like robust engineering solutions [4]–[6], predictive maintenance [7]–[10] or industry 4.0 [11], [12] need new concepts for the structure of the university curriculum. Thus, a further challenge of this aspect is the modification of the laboratories to meet these new requirements.

Welding plays a special role in robotics. Beyond positioning the correct path planning has an extremely large effect on the quality of the seam. The design of the welding jig, choice of the appropriate sensor technology [13]–[15] and the programming of the robot arm is a challenge for education as there are a lot of safety requirements for which a common robotics laboratory is not properly prepared.

The second section of the paper describes the basics of the problem of a welding robot like fixing the workpiece, coordinate systems of the robot arm or welding. The focus of the third section is the design of the welding jig mentioning those topics which might be important in education. The fourth section introduces a simulation environment for robot programming and the updates needed after transferring the program to the robot controller.

II. PROBLEM DESCRIPTION

During the design, the following design steps were done:

- positioning of the workpiece
- design of the welding jig
- programming of the robot

Determining the position of the workpiece is fundamental. When positioning, we must pay attention to fixing the degrees of freedom of the workpiece, which means that it cannot be moved in any direction. Rigid bodies have six degrees of freedom, as shown in Fig. 1. These freedoms are the displacements along the x, y, z axes and the rotation about these axes.

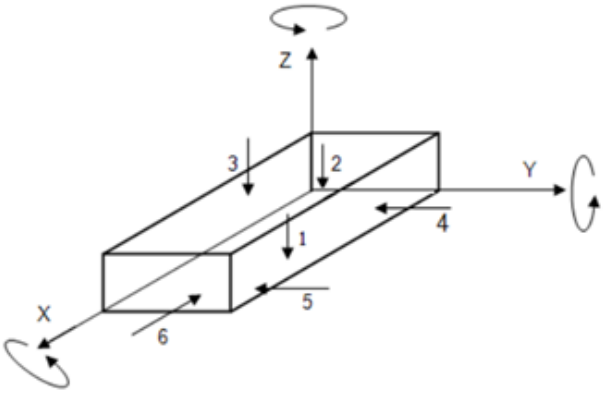


Fig. 1: 6 degrees of freedom [16]

In Fig. 1 we also see 6 points, which symbolize at which points the surfaces of the body are supported and thus all degrees of freedom can be taken away. This way it can be achieved that the body cannot move. In our case, this is a very important factor so that the position of the workpiece is fix in the device in such a way that it does not move in either direction during welding. [16]

After determining the position of the workpiece, the next step was to design the fixation. When mounting, it is important that the part does not move during welding and does not suffer from deformation or surface damage. In addition, it must even be considered that no large surface damage remains in the finished piece in the end, do not have imprints or scratches larger than accepted. When designing the clamp, the surfaces suitable for the clamp must be selected. This depends on the shape of the workpiece, the rigidity, the location of the positioning elements and supports and, in our case, it is also a crucial aspect that the welding robot can approach the paths of the seam.

Solid knowledge of the different coordinate systems used in robot programming (see Fig. 2) is also essential for university studies.

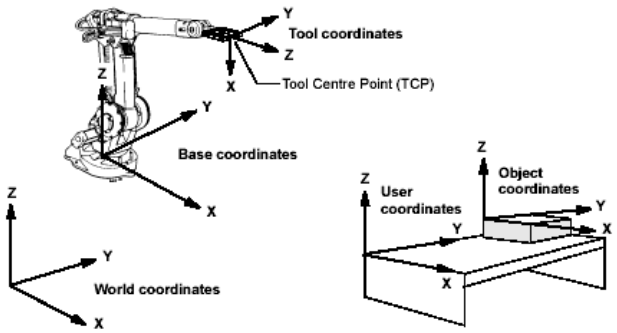


Fig. 2: Coordinate systems [17]

III. Welding Jig Design

The workpiece (see Fig. 3) consists of two parts, a bent "U" profile and a plate. Both parts are made on a laser cutting machine and one is then bent on a bending machine. In the case of welding, the point is that the plate is welded to this "U" profile at a certain height. The first design step is the development of the welding jig. Numerous requirements must be considered:

- main dimensions
- tolerances
- type of welding seam
- location of welding seam

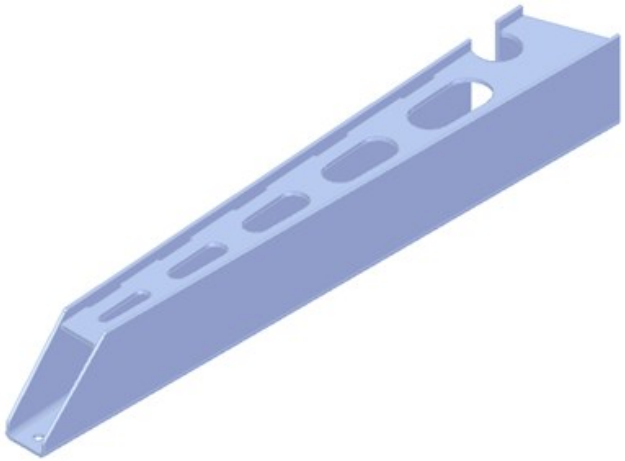


Fig. 3: Workpiece

The whole device is based on a S355MC 10mm thick steel plate, which is a commonly used commercially available material. The seats and bumpers are made of C45 material. The closed section in the middle of the device is a 100x80x4 closed section made of S235JR material. The other two smaller parts of the moving support were also made of S235JR material in a size of 60x20x2. Clamps are used to fix the parts. In order to make both a right-hand and a left-hand piece of the part, which means that the two are mirror parts of each other, a jig has been designed on which both parts can be made at the same time. It was also necessary to have a moving support on which, on the one hand, the plate part was to be welded to the profile and, on the other hand, to hold against the clamps used to fix the part on its side. The table is 1000 mm high and was assembled from 80x80x4 closed sections. The palette is welded to this table (see Fig. 4).

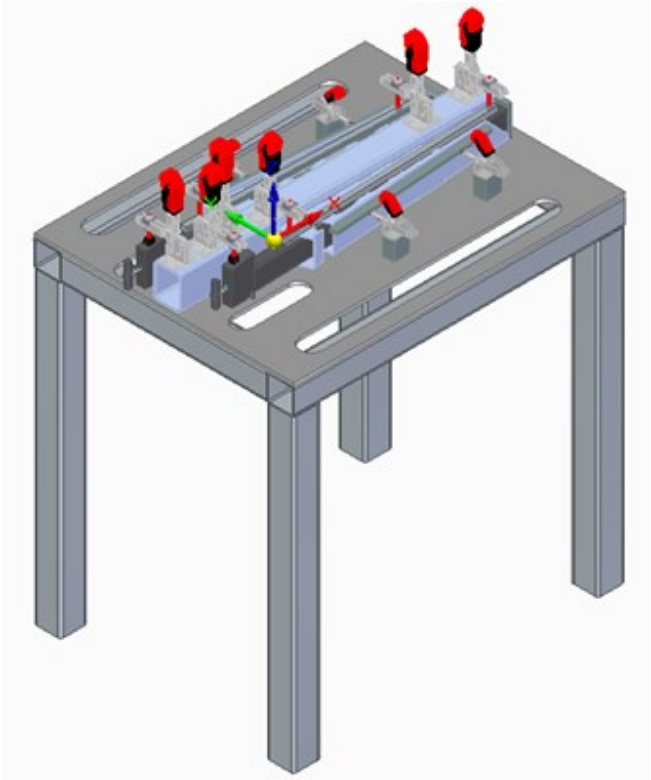


Fig. 4: Welding jig

The first version of the welding jig can be seen in Fig. 5. A common use case of welding jig design can be demonstrated on this part where the moving support is positioned parallel to the central closed section.

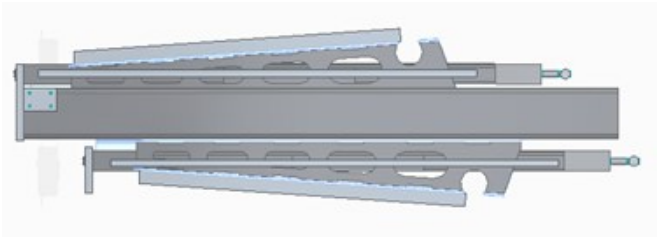


Fig. 5: First version of the welding jig

All the parts are fixed properly considering the first requirement concerning the welding jig, nevertheless the welding robot cannot reach all the necessary positions on the workpiece, therefore it would not be possible to weld. That is why it was necessary to change its position as shown in Fig. 6. For this reason, the front bumper was also milled so that the moving support could collide completely on it. Thus, the robot can make the necessary movements also near the closed section.

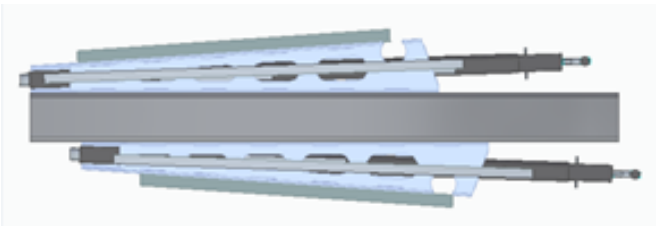


Fig. 6: Second version of the welding jig

IV. Testing in Simulation Environment

3D simulation program called Roboguide was used as simulation environment. In this program we can program the robot and test the finished robot programs in a safe simulation environment.

The program was used for two goals. The first goal was to make sure that the device is suitable in the sense that the robot will have access to all the parts that need to be welded and there will be no collision with the device anywhere. The most problematic position of the welding jig from this point of view is the farther side of the workpiece. Fig. 7 shows that this is accomplished and there is no collision.

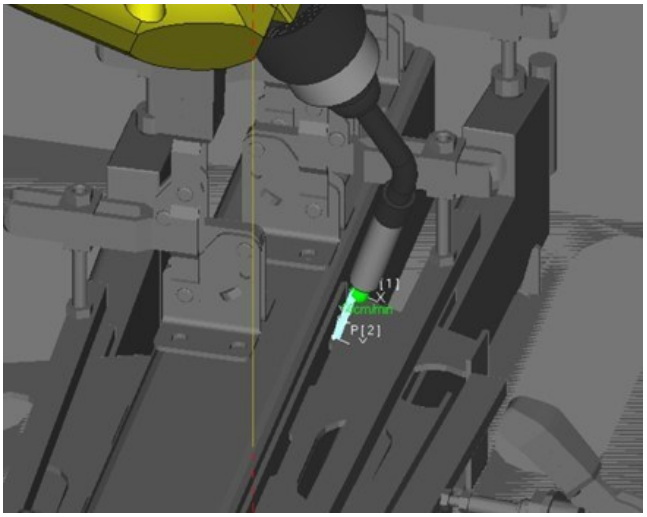


Fig. 7: Simulation environment

Fig. 8 shows an error code. This error code was encountered several times during testing. It means the robot supposed to move to a point it could not reach. That was why there was still a bit to adjust on the device or the position of the device. When placing the device, it was designed in the robot cell, care must be taken to ensure that the robot arm reaches every point to be welded.

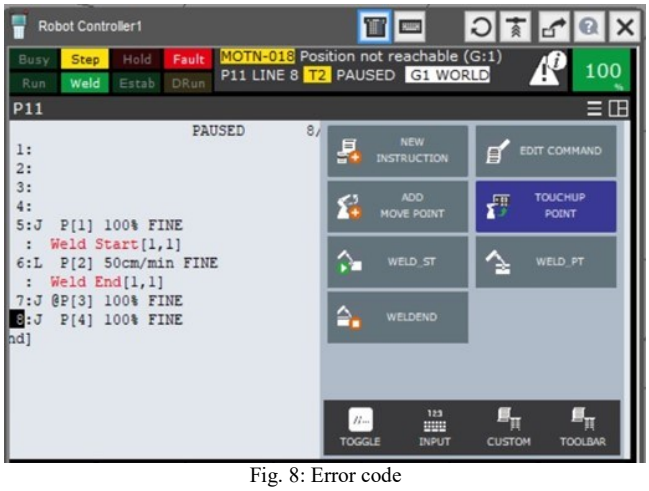


Fig. 8: Error code

The second goal was to program the robot in the simulation environment. The purpose was to implement the welding sequence in a safe environment. The possible collisions or bad seams are not crucial in a virtual environment. After this step, the program can be transferred to the real robot controller, where only the relevant position registers must be updated for the real environment of the robot cell. Then the program must be run in test mode and it must be checked that everything runs properly and there are no collisions.

For the programming, the teach in method was applied from point to point and recorded each point for the robot, making sure not to collide with the elements of the device. Since the program was ready in advance, it was easy to update the program. Nevertheless, there were questionable cases where the robot movements were repeatedly run between two points to see if it is possible to fit in the given tight spaces.

Programming was started by setting the basic data. The robot controller is built so that the welding parameters, such as current, voltage, wire feed speed etc., can be set and they are saved in a so-called JOB which can be called during programming to set the predefined parameters.

V. Conclusion

The paper introduced a use case for robotic welding design with educational and research purposes. The design process can be presented form the first steps till the working robot cell starting from a computer laboratory and finishing in the workshop. The proposed environment facilitates the education of several engineering field. The paper introduces a curriculum for the following topics:

- fixing 6 DOF of a workpiece
- coordinate systems in robotics
- welding jig design
- robot programming in simulation environment
- application of the simulation results in the real robotic environment

Acknowledgment

The authors thankfully acknowledge the financial support of this work by the Howmet Aerospace Foundation, the Hungarian State and the European Union under the EFOP-3.6.1-16-2016-00010 and 2019-1-3-1-KK-2019-00007 projects.

The research has been supported by the NTP-HHTDK-19-022 grant. The authors declare no conflict of interest. The authors would like to express their gratitude to those who have provided their help in the various experiments.

References

[1] M. Pogatsnik, “Measuring Problem Solving Skills of Informatics and Engineering Students,” in *IEEE Joint 19th International Symposium on Computational Intelligence and Informatics and 7th International Conference on Recent Achievements in Mechatronics, Automation, Computer Sciences and Robotics, CINTI-MACRo 2019 - Proceedings*, Nov. 2019, pp. 93–98, doi: 10.1109/CINTI-MACRo49179.2019.9105277.

[2] M. Pogátsnik, “Életpálya-építés serdülő- és ifjúkorban - kutatás közben,” *Egyéni Kül. Szerepe Tanulásban És Pályaválasztásban*, pp. 149–162, 2015.

[3] M. Pogatsnik and R. B. Kendrovics, “Communication and Reading Comprehension among Informatics and Engineering Students,” in *SAMI 2020 - IEEE 18th World Symposium on Applied Machine Intelligence and Informatics, Proceedings*, Jan. 2020, pp. 235–240, doi: 10.1109/SAMI48414.2020.9108764.

[4] G. Györök and B. Beszédes, “Fault tolerant power supply systems,” in *11th International Symposium on Applied Informatics and Related Areas (AIS 2016)*, 2018, pp. 68–73.
[5] G. Györök and B. Beszédes, “Highly reliable data logging in embedded systems,” in *2018 IEEE 16th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, Feb. 2018, pp. 49–54, doi: 10.1109/SAMI.2018.8323985.
[6] G. Györök and B. Beszédes, “Adaptive Optocoupler Degradation Compensation in Isolated Feedback Loops,” *2018 IEEE 12th Int. Symp. Appl. Comput. Intell. Inform. SACI*, pp. 167–172, 2018.
[7] G. Manhertz and Á. Antal, “The effect of air-fuel equivalence ratio change on the vibration components of an internal-combustion engine,” *RECENT Innov. Mechatron. 2 1-2*, pp. 1–6, 2015, doi: 10.17667/riim.2015.1-2/13.
[8] G. Manhertz and Á. Bereczky, “Development of a vibration expert system to analyze and predict malfunctions in internal-combustion engine,” in *Proceedings of ARES’14 : Workshop on Application of Robotics for Enhanced Security*, 2014, pp. 24–29.
[9] G. Manhertz, D. Modok, and Á. Bereczky, “Evaluation of short-time fourier-transformation spectrograms derived from the vibration measurement of internal-combustion engines,” in *2016 IEEE International Power Electronics and Motion Control Conference (PEMC)*, Sep. 2016, pp. 812–817, doi: 10.1109/EPEPEMC.2016.7752098.
[10] G. Manhertz, G. Gardonyi, and G. Por, “Managing measured vibration data for malfunction detection of an assembled mechanical coupling,” *Int. J. Adv. Manuf. Technol.*, vol. 75, no. 5, pp. 693–703, Nov. 2014, doi: 10.1007/s00170-014-6138-3.
[11] É. Hajnal, “Big Data Overview and Connected Research at Óbuda University Alba Regia Technical Faculty,” in *AIS 2018 - 13th International Symposium on Applied Informatics and Related Area*, 2018, pp. 1–4.
[12] A. Selmeci and T. Orosz, “Usage of SOA and BPM changes the roles and the way of thinking in development,” in *2012 IEEE 10th Jubilee International Symposium on Intelligent Systems and Informatics*, Sep. 2012, pp. 265–271, doi: 10.1109/SISY.2012.6339526.
[13] Z. Pentek, T. Hiller, and A. Czmerk, “Algorithmic Enhancement of Automotive MEMS Gyroscopes with Consumer-Type Redundancy,” *IEEE Sens. J.*, pp. 1–1, Aug. 2020, doi: 10.1109/jsen.2020.3017094.
[14] Z. Pentek, T. Hiller, T. Liewald, B. Kuhlmann, and A. Czmerk, “IMU-based mounting parameter estimation on construction vehicles,” in *2017 DGON Inertial Sensors and Systems, ISS 2017 - Proceedings*, Dec. 2017, vol. 2017-December, pp. 1–14, doi: 10.1109/InertialSensors.2017.8171504.
[15] T. Hiller, Z. Pentek, J. T. Liewald, A. Buhmann, and H. Roth, “Origins and Mechanisms of Bias Instability Noise in a Three-Axis Mode-Matched MEMS Gyroscope,” *J. Microelectromechanical Syst.*, vol. 28, no. 4, pp. 586–596, Aug. 2019, doi: 10.1109/JMEMS.2019.2921607.
[16] A. Szabó and I. Kozma, *Gyártóeszközök tervezése és gyártása*. 2011.
[17] *Coordinate systems, http://www.ipacv.ro/proiecte/robotstudio/textbooks/file/robot_motion.htm, Accessed: 2020-11-08*. .

Detecting the effects of emotions and higher dimensional facial vectorization on facial recognition in a smart mirror system

Andras Toth
Obuda University
Alba Regia Technical Faculty
sba.andras.toth@gmail.com

Patrik Toth
Obuda University
Alba Regia Technical Faculty
patrikthetoth@gmail.com

Szabolcs Meszaros
Obuda University
Alba Regia Technical Faculty
villam983@gmail.com

Jozsef Halasz
Obuda Universit,
Alba Regia Technical Faculty
halasz.jozsef@amk.uni-obuda.hu

Abstract—

As new technologies are introduced; their full potential might not be apparent at first. As facial recognition is used more and more in several different devices and services, one good idea might as well separate one product from the rest. As it has been demonstrated in a previous paper, introducing facial recognition into a smart mirror is not only feasible, it can also be practical. There are DIY solutions that provide the functionality mentioned above, the present concept strives to offer something more.

The aim of this paper was to investigate the possibility of detecting the effect of facial emotions via the same 128-dimensional facial recognition system and compared it one that utilizes 512-dimensional vectors to represent the human face.

The project utilized a face recognition pipeline used Euclidian distances to classify the users. These users were artificially created, with neutral, angry and happy emotions were applied to their faces. All together more than 30000 distances were measured, these were the basis for this paper. General linear model was used to analyze these distances.

The results showed that the solution with 512-dimensional vectors revealed significantly higher distances between different users. Within the same users, the emotional content was able to increase distances, and this effect was more prominent with 512-dimensional vectors compared to 128-dimensional ones.

In conclusion, our result indicate that the 512-dimensional solution had higher sensitivity and the effect of emotional content on facial detection must be considered in later studies.

Keywords— emotion, Euclidean distance, face recognition, neural network, smart mirror

I. Introduction

Smart devices are part of our everyday life now [1]. Most companies try to adapt their products to fit their customers as best as possible [2]. The team's vision is about a smart mirror, that utilizes face recognition to show meaningful information to its users while they perform menial tasks [3].

The team's smart mirror project consist of three major components, one of these were the face recognition pipeline. Over the course of the project some results pointed out, that the current state of the pipeline is not sufficient, thus new technologies were adopted [4][5].

Because of this change in the underlying technologies, we were able to achieve greater accuracy for our face recognition system. Using this newly gained accuracy a new goal was set, to detect the effects of different emotions on the face recognition pipeline accuracy, and using this data prepare our system to differentiate emotions on our users' face.

The aim of this paper was to investigate the possibility of detecting the effect of facial emotions via the same 128-dimensional facial recognition system and compared it one that utilizes 512-dimensional vectors to represent the human face.

II. Methods

A. Creating the face repository

For the creation of the face repository, a software called FaceGen Modeller (demo version) [6] has been chosen. The 4 basic face features (Figure 1) for this paper have been created with the software's randomizer option. In this software it is possible to adjust the faces so called Action Units [7]. These action units are responsible to different movements in the face itself, in this way it is possible to create facial expression like happiness or anger. For every measuring point with given facial expression and intensity with the needed action units [7] an XML document has been created for the purpose of reusability, these were used for the creation of the faces presented in this dataset. For each model 2 different set of pictures were created with different intensity of anger or happiness applied. (Figure 2) All users were created with the male preset of FaceGen Modeller to avoid any potential biases in the face recognition pipeline.

These sets were created for each emotion: neutral with the intensity between 0 and 10 %, low with the intensity between 20 and 30 % and high, with the intensity value between 50 and 60 % with 1 percent increments for all sets. These were then captured with the help of a program called ShareX [8]. For the proper file name format, that contained the user's name, the applied emotion, and its intensity the program Bulk Rename Utility [9] was used.

B. Facial Recognition pipeline

The pipeline consists of three layers, namely: Detection, Representation and Classification. The detection layer finds a face in the input image, and after cropping the image to reduce complexity, it is fed into the representation layer, which is tasked with vectorizing the cropped image, and as such, provides us with an n-dimension vector that is specific to the face in the input image. After this, we store these vectors, or embeddings and once another face is encountered, we can compare them, which ultimately is the responsibility of the classification layer (Figure 3).

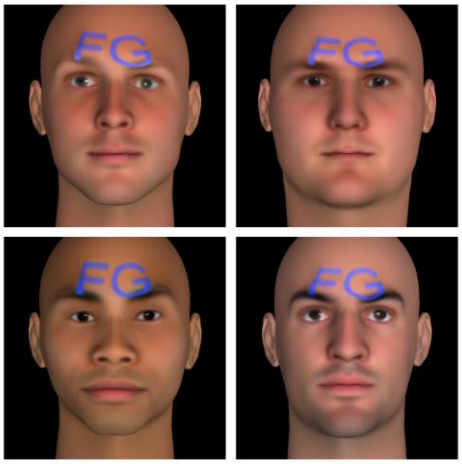


Figure 1. Four representative users in their neutral state
These were created using FaceGen Modeller [4].

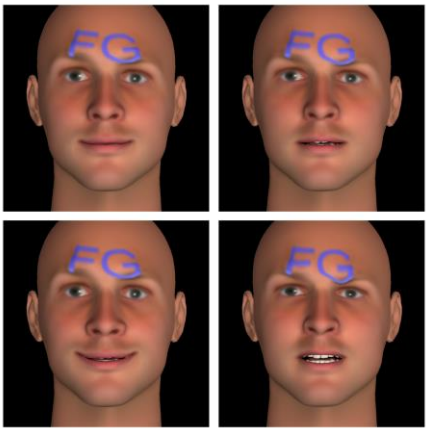


Figure 2. Representative emotions from a specified user
Top left low intensity happiness, lower left high intensity happiness
Top right low intensity anger, lower right high intensity anger
These were created using FaceGen Modeller [4].

C. *Measuring accuracy*

Once we have these embeddings, we can easily compare two faces by way of calculating the Euclidian distance of the two vectors, thus giving us a scalar distance measurement that we can use to determine the accuracy of the recognition pipeline. For example, 0 means the two faces are identical, while a distance of 1.6 means they are nothing alike. In practice, we see a threshold of recognition at around 0.6, as in, if the distance is larger than 0.6, the two users are not the same.

D. *Evaluating the gathered data*

More than 30 thousand data points were gathered during the tests, these contained the data set user's data and evaluating user's data such as username applied emotion, the emotions intensity, and the Euclidean distance within the 2 users. This type of distance was used in the FaceNet paper [10] to represent how closely two face represents the same person. The researchers in that paper used 1.1 as a segmentation threshold. Distances below 1.1 between two faces were considered to belong to the same user.
The project used Python 3 with Seaborn, Numpy and Pandas for data evaluation.

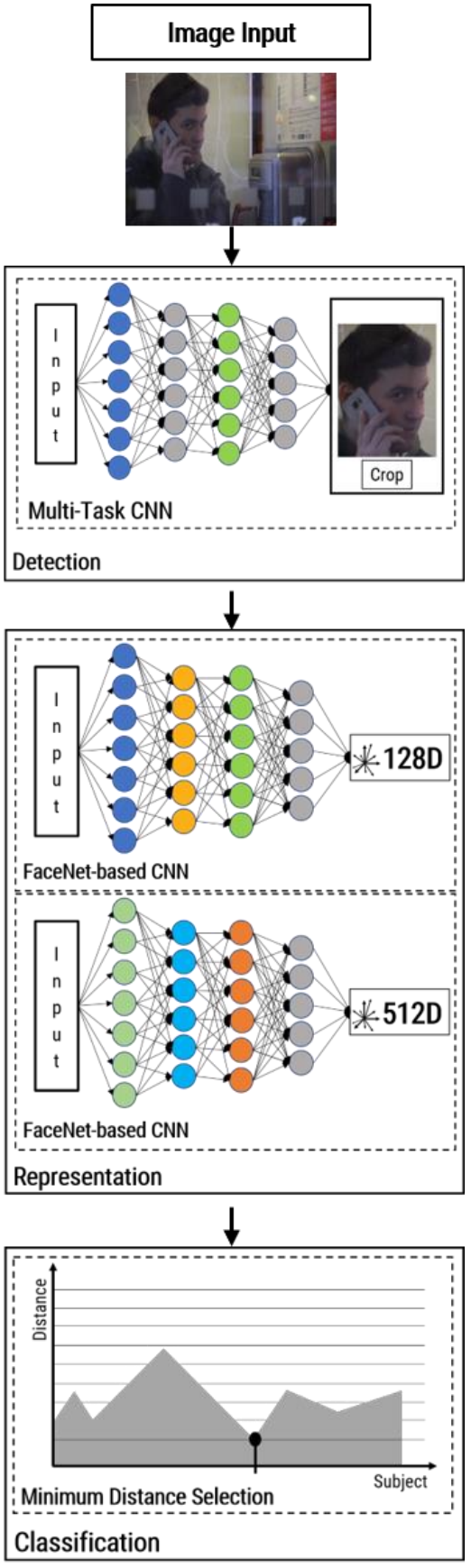


Figure 3. The assembly of the three staged face recognition pipeline.
The first stage's task is to detect and crop human faces.
The second stage converts the cropped images to a vector.
The third stage measures distances between two vectors.

E. Smart mirror system software architecture

The given results were originally gathered to evaluate two different face recognition and identification systems. The chosen system was used as the basis for our software architectures face recognition system. This was used to access user related data from the central database to show relevant information to our users. These were transferred with simple HTTP Request and Responses, using a RESTful [11] solution. The given software architectures design can be seen on Figure 4.

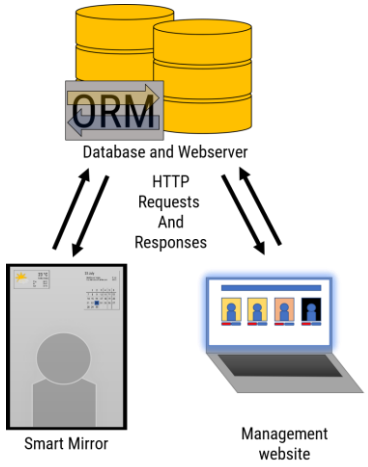


Figure 4. Software architecture of the team's smart mirror system.
The smart mirror client communicates with a webserver using HTTP request and responses.
The central webserver uses ORM technology to store user data.

III. RESULTS

Effect	SS	Degr. of Freedom	MS	F	p
	Univariate Tests of Significance for Distance Sigma-restricted parameterization Effective hypothesis decomposition				
Intercept	10828,53	1	10828,53	485749,1	0,000000
{1}Emotion	12,01	1	12,01	538,7	0,000000
{2}Intensity	22,82	1	22,82	1023,5	0,000000
{3}Dimension	167,37	1	167,37	7508,1	0,000000
{4}Within_User	3793,46	1	3793,46	170168,2	0,000000
Emotion*Intensity	4,00	1	4,00	179,3	0,000000
Emotion*Dimension	0,00	1	0,00	0,0	1,000000
Intensity*Dimension	0,88	1	0,88	39,6	0,000000
Emotion*Within_User	10,20	1	10,20	457,6	0,000000
Intensity*Within_User	18,78	1	18,78	842,5	0,000000
Dimension*Within_User	40,04	1	40,04	1796,0	0,000000
Emotion*Intensity*Dimension	0,32	1	0,32	14,2	0,000166
Emotion*Intensity*Within_User	2,46	1	2,46	110,5	0,000000
Emotion*Dimension*Within_User	0,35	1	0,35	15,6	0,000079
Intensity*Dimension*Within_User	0,75	1	0,75	33,5	0,000000
1*2*3*4	0,25	1	0,25	11,3	0,000776
Error	690,17	30960	0,02		

Table 1. Distance statistics.
General Linear Model (STATISTICA 7.0) was used in the examined 4-factorial design. $p < 0.05$ represents statistically significant difference between groups. The results of the interactions are also presented.

STATISTICA 7.0 was used to establish General Linear Model with 4 factors. Factor 1 was Emotion (Anger versus Happiness), Factor 2 was Intensity (Low versus High), Factor 3 was Dimension (128 versus 512) and finally, Factor 4 was Within User (Within User versus Between User). Interactions between factors were also analyzed. $p < 0.05$ were considered statistically significant (Table 1).

The initial results showed quantifiable difference with the new, 512-dimension representational layer compared to the former 128-dimensional one (Figures 5 and 6).

As one can see in Figures 5 and 6, the 512-dimensional system produces a bigger gap between those distance values that were measured with pictures belonging to the same user, as opposed to those value that were the result of comparing two vectors that belonged to different users.

All figures include a separator line at the distance value of 1.1, as recommended as a threshold value for separating users from each other [10]. On Figure 7, it is clearly visible that the 512-dimension system produces higher distances, this effect is more pronounced in those situations when the compared

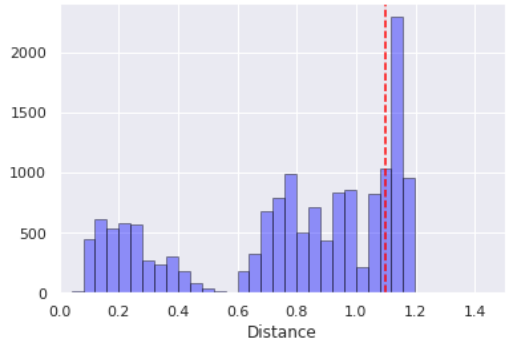


Figure 5. Distances measured with 128D representational layer.
Histogram of measured distances, bins represent the number of observations of given distance.

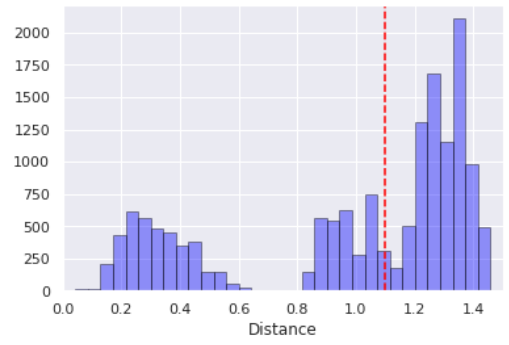


Figure 6. Distances measured with 512D representational layer.
Histogram of measured distances, bins represent the number of observations of given distance.

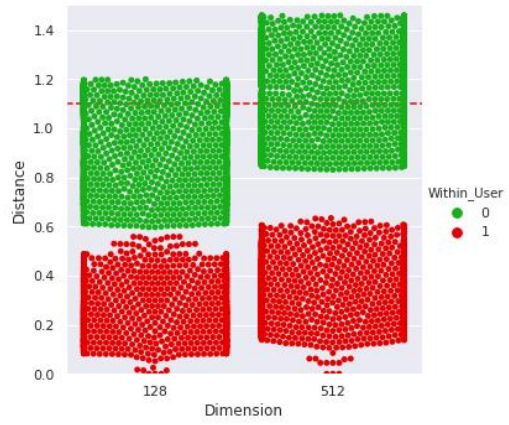


Figure 7. Distances with given Dimension
Dots represent measured distances
Dot coloring indicates that the two vectors belong to the same user or not. (Red dots: same user; Green dots: different user)

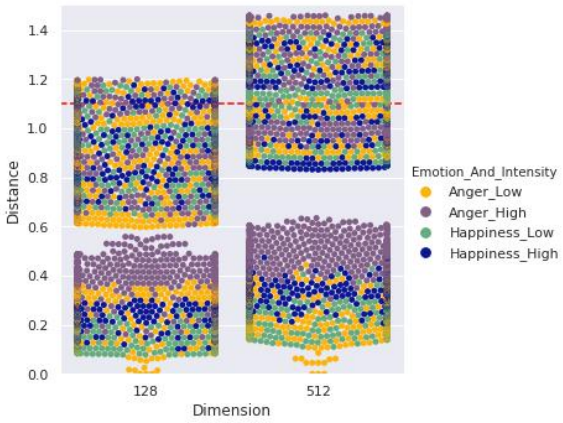


Figure 8. Emotions and their intensities.
Dots represent measured distances.
Dot coloring represent the emotion and intensity combination.
Layering only is expressed in measurements belonging to the same users (dots with smaller distances).

vectors belonged to different users. In measurements, when the two compared vectors belonged to the same user the lowest possible distance values were generated by those vectors that represented a user with low intensity anger, in turn the maximum possible values were generated by faces that represented high intensity anger. Between these datapoints were the ones generated by vectors that represented faces with happiness. In general vectors containing low intensity happiness had lower distances than those with high intensity happiness. These effects can be seen on Figure 8.

IV. Discussion

The main findings of present study were the following:

A. Dimension findings

We believe the additional dimensions in the representational layer helped our system to represent the users with more granularity. With help of this new representational layer the system is more likely capable of evading user misclassification.

B. Within and between users findings

The two measured systems performed well enough, given that all the within user measurements were much lower than the 1.1 value provided by the FaceNet paper[10].
The 512-dimensional system provided a much bigger gap between the pools of measurements, thus lowering the possibility of misclassification.

C. Emotion findings

As one can see, the measured distances for the emotion anger produced higher distances between the low intensity and high intensity datapoints. This effect is the result of more active Action Units, [7] i.e. more facial features are active at the same time. Generally, the human face uses more muscle movement to represent anger, these muscle groups are called Action Units, these were introduced in the system Facial Action Coding System, by Carl-Herman Hjortsjö [12].

With the gathered data it was possible to deduce which emotion had bigger effect on the measured distances, and how much the measured distance and its intensity correlate.

D. Intensity findings

As the provided measurements shows, the data points with higher intensities provided higher distances between the two data points.

V. Conclusions

The project was successfully finished, it resulted in a working system with a central database and webserver, a client program, and an administrator web interface.

In conclusion, our result indicate that the 512-dimensional solution had higher sensitivity and the effect of emotional content on facial detection must be considered in later studies.

It is self-evident, that those measurements that were performed between two different users resulted in higher distances, in this case emotion and their intensities did not resulted in a marked layering. Within the same users, different emotions resulted in a significant increase of distances, e.g. in high intensity anger, the distances doubled, compared to the neutral faces in 512-dimensional measurements.

The presented study proved, that the emotional qualities have great impact on the measured distances. In addition, it has been proven, that the 512-dimensional system handles these emotional qualities with greater finesse.

Acknowledgment

The research has been supported by the NTP-HHTDK-19-022 grant. The authors declare no conflict of interest. The authors would like to express their gratitude to those who have provided their help in the various experiments.

References

[1] M. R. Alam, M. B. I. Reaz, and M. A. M. Ali, “A review of smart homes - Past, present, and future,” *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, vol. 42, no. 6, pp. 1190–1203, 2012, doi: 10.1109/TSMCC.2012.2189204.

[2] G. Brussenskiy, C. Chiarella, and V. Nagda, “Smart Mirror Sponsored by Central Florida Inpatient Medicine Group 10 Senior Design II-Project Documentation 4-28-2014,” 2014.

[3] P. S. Tóth, A. T. Tóth, and S. Mészáros, “Concept and implementation of a smart mirror,” in *14 th International Symposium on Applied Informatics and Related Areas organized in the frame of Hungarian Science Festival 2019 by Óbuda University*, 2019.

[4] K. Zhang, Z. Zhang, Z. Li, S. Member, Y. Qiao, and S. Member, “(MTCNN) Multi-task Cascaded Convolutional Networks,” *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, 2016, doi: 10.1109/LSP.2016.2603342.

[5] A. Paszke *et al.*, “PyTorch: An Imperative Style, High-Performance Deep Learning Library,” in *Advances in Neural Information Processing Systems 32 (NIPS 2019)*, 2019.

[6] Singular Inversions Inc., “FaceGen Modeller.” Singular Inversions Inc., 2020, [Online]. Available:

www.FaceGen.com.

[7] E. B. Roesch, L. Tamarit, L. Reveret, D. Grandjean, D. Sander, and K. R. Scherer, "FACSGen: A Tool to Synthesize Emotional Facial Expressions Through Systematic Manipulation of Facial Action Units," *J. Nonverbal Behav.*, vol. 35, no. 1, pp. 1–16, 2011, doi: 10.1007/s10919-010-0095-9.

[8] ShareX Team, "ShareX." 2020, [Online]. Available: https://getsharex.com/.

[9] J. Willsher, "Bulk Rename Utility." Acebrook Pty Ltd, 2019, [Online]. Available: https://www.bulkrenameutility.co.uk/.

[10] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 07-12-June, pp. 815–823, 2015, doi: 10.1109/CVPR.2015.7298682.

[11] R. T. Fielding, "Architectural Styles and the Design of Network-based Software Architectures," 2000.

[12] C.-H. Hjortsjö, *Man's face and mimic language*. Studentlitteratur, 1970.

Tangible Results of the IRSEL Project

M. Verőné Wojtaszek
Institute of Geoinformatics
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
wojtaszek.malgorzata@amk.uni-obuda.hu

V. Balázsik
Institute of Geoinformatics
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
balazsik.valeria@amk.uni-obuda.hu

L. Földváry
Institute of Geoinformatics
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
foldvary.lorant@amk.uni-obuda.hu

A. Kovács
Institute of Geoinformatics
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
kovacs.anna@amk.uni-obuda.hu

B. Márkus
Institute of Geoinformatics
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
markusbela@gmail.com

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

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

***Abstract*— The objective of the IRSEL project was to develop an innovative learning platform, a LMS in the field of Remote Sensing Applications for Asian countries, China and Thailand. Beyond establishing the LMS at each Asian university, a Knowledge Pool of high-level e-Learning teaching materials for a wider scientific and engineering community has been developed, furthermore workshops, trainings for the teaching staff and a summer school for selected students has been organized. The LMS installed at 4 Asian universities serves the practical applicability of increasingly higher resolution Remote Sensing data coming from more and more sources for wide range of disciplines (including different tasks of Environmental protection, Agriculture, Forestry and fishery, Physical sciences, Engineering and engineering trades, Transport services, Security services). All intellectual products of the project are available via internet for the whole Remote Sensing society at large, while the full feature is hosted at the Thai and Chinese partners and will be involved in their higher education activity.**

Keywords— Remote Sensing applications, e-learning, LMS

I. INTRODUCTION

The EU space strategy according to [1] says "The Commission's aim is to optimise the benefits that space brings to society and the wider EU economy. Achieving this means boosting demand among public and private users, facilitating access to and use of space data, and stimulating the development and use of innovative downstream applications." The demand is in line with the practice observed in the field of Remote Sensing (RS), which has wide range of potential applications unused [2].

The Erasmus+ Capacity Building for Higher Education Key Action 2 project titled Innovation on Remote Sensing Education and Learning (abbreviated as IRSEL) has been conducted for the period 15 October 2017 to 14 October 2020. The wider objective of the IRSEL project was to develop an innovative learning platform, a Learning Management System (LMS) for Asian countries, China and Thailand, which already have relevant activity in the field of RS. Though the content of the implemented LMS does not correspond to any level of tertiary level education, it is basically developed for meeting from BSc to MSc level of RS related disciplines.

More specifically, the following activities were planned and completed: 1) establishing a LMS at each of participating Asian universities, 2) creating a Knowledge Pool of high level e-Learning teaching materials for a wider scientific and engineering community, and 3) organizing workshops, trainings and a summer school for the students, teaching and the administration staff.

Specific Project Objectives of the IRSEL project has been:

1) To ensure researchers and academic staff in RS and geospatial sciences for studies integrated into world-wide sustainability academia (research) community

2) To strengthen and integrate RS and geospatial science into Multi-Inter-Trans-Cross-Disciplinary sustainability studies and research of Socio-Ecological Systems

3) To enhance the role of Asian institutions in socio-ecological systems studies and researches for the benefit of Asian region

4) To promote internationalization on the relevant knowledge areas

5) To enhance international cooperation between EU and Asian universities and research institutes

The LMS hosts 20 newly developed modules on Remote Sensing in the curricula of participating universities, improving the quality of higher education, delivering a background for studying the practical use of the Remote Sensing techniques. This would by time enhance the practical use of Remote Sensing on a wide range of applications serving the labour market and society. The competences and skills in the participating Higher Education Institutions (HEIs) will be developed using these learning modules. The developed LMS at 4 Asian universities will serve the practical applicability of Remote Sensing data for wide range of disciplines (including different tasks of Environmental protection, Agriculture, Forestry and fishery, Physical sciences, Engineering and engineering trades, Transport services, Security services). The

aim of the LMS is to foster the uptake of Remote Sensing applications to boost the benefits that Earth Observation (EO) brings to society and the wider economy. Society will benefit from the improved contacts between research and application for sustainable development.

Now the IRSEL project has reached its end. Instead of focusing on the project activities in general, this paper focuses only on those tangible results, which can openly be used by the international RS society. For up-to-date information on the IRSEL project, you may visit the official website at http://irsel.eu/, c.f. the QR code in Figure 1.

The project has been implemented in an international cooperation of 8 higher education institutions additionally to the 6 associated partners. The list of the involved institutions is summarized in Table I.

TABLE I. THE CONSORTIUM

Partners		
HEI	***abbr.***	***Country***
Obuda University	OU	Hungary
University of Natural Resources and Life Sciences	BOKU	Austria
Jagiellonian University	JU	Poland
University of Twente	ITC	the Netherlands
Fujian Normal University	FNU	China
Yunnan Normal University	YNNU	China
Asian Institute of Technology	AIT	Thailand
Khon Kaen University	KKU	Thailand
Associated partners		
HEI	***abbr.***	***Country***
Government Office of the Capital City Budapest, Department of Geodesy, Remote Sensing and Land Offices[1]	BFKH-FTFF	Hungary
Research Centre for Astronomy and Earth Sciences, Hungarian Academy of Science	MTA RCAE	Hungary
Yunnan Hanzhe Technology Co. Ltd.	-	China
Southwest Forestry University	SWFU	China
Mahasarakham University	MSU	Thailand
Thammasat University	TU	Thailand

II. KNOWLEDGE CENTRES AND KNOWLEDGE POOL

A key outcome of the IRSEL project was the development of 4 Knowledge Centres (KC) hosting LMS at each Asian partner Universities. Beyond the hardware and software demands, also some equipment for data acquisition is included in the KCs.

The LMS is host on the well-known open-source learning platform, the Moodle. The core material of the hosted knowledge pool is the 20 learning modules, additionally, freely accessible research and teaching materials, and scientific publications developed within the frame of this project, c.f. [3]-[10] are included.

[1] Now it is merged to Lechner Knowledge Centre.

Fig. 1. QR code of the irsel.eu webpage

By the end of the project, the content of the KCs is identical at each Asian partners. By time, Asian partners should modify, extend and update the content, including additional knowledge pool. It can be extended with the translation of the modules to local (Chinese or Thai) languages, additional Remote Sensing software and Remote Sensing data.

Generally, the knowledge pool hosted on the LMS will be involved at different levels of the education of the Asian partners, including distance and blended learning forms as well. For supporting the use of the LMS for self-study, a learning guide is provided. Similarly, for boosting the use of LMS for teaching, a teaching guide is also delivered.

A major part of the equipment purchase budget of the IRSEL project was dedicated to establishing the hardware and software demand of the knowledge pool. Beyond the hardware, software and data demands, also tools for data acquisition were invested, as no Remote Sensing application can be performed without acquiring data. Without the sake of completeness, the equipment park for data acquisition contains laser scanner, multispectral UAV, camera, spectroradiometer, plant canopy analyser.

Fig. 2. QR code of the registration webpage

Though the learning materials has been developed primarily for the benefit of the Chinese and Thai partner universities, according to the open access policy of the project, learning materials are also made available for the international RS society. After a registration at https://lms.irsel.eu/ website, you can get an access to download the learning modules in readable form. The downloadable content can be used offline, while the full feature, applying innovative tools in the learning material is available only on the Moodle platform of Asian partners.

III. Learning Material Development

As the e-Learning material content of the Learning Management System is the intellectual output of the IRSEL project, which can be accessed by the RS community, in this chapter we provide a review of the 20 modules.

Prior to the learning material development, the Needs Analysis was done. As a result, we've created a Needs Analysis Report. It is a document describing detailed needs and expectations from Asian Remote Sensing Society. The report contains results of needs analysis questionnaires, summarize the recommendations according to local demands, and list the targeted stakeholders and respondents from China and Thailand. The report provides necessary information for module development for the IRSEL project.

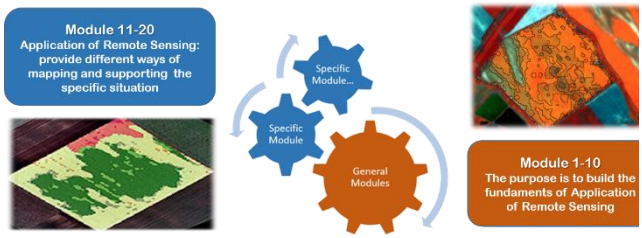


Fig. 3. Structure of the Learning Materials

The learning materials can be divided into two parts: the first 10 modules provide the fundamental knowledge on Remote Sensing at appr. BSc level, while the last 10 modules deals with actual Remote Sensing applications at appr. MSc level (Fig. 2). The actual course specifications are detailed below:

M1 Physical Principles of Remote Sensing: providing students with physical principles of Remote Sensing, the tool to obtain information on the earth from deci-meter level to km level locally and globally, as well as basic RS image processing techniques and skills. The main focus remains on theories and laws on nature of light, its interactions with the atmosphere and earth surface. Additionally, this module focuses on introducing spectrometer, their types, applications and approach to conduct laboratory experiments.

M2 Data Acquisition, Sensors and Platforms (passive sensing): The module provides an overview of different sensors and platforms for the acquisition of passive EO data with emphasis on satellites. EO data, and subsequently the sensors, can be characterized based on the different types of resolutions (spatial, spectral, temporal, radiometric). The main platforms - such as unmanned aerial vehicles (UAV), airplane, and satellite - are presented. Further, the data acquisition process and the different types of sensors (frame cameras, scanners) will be analysed.

M3 Data Acquisition, Sensors and Platforms (active sensing): This module aims at introducing basic and applications of active remote sensing. It covers the basics of imaging system and polarmetric of radar. The differences in type of sensor and availability of spaceborne and airborne sensor. The characteristics of scattering and reflection of microwave energy in various type of surface. An Introduction to Lidar system concepts. The exercises and case studies allow students to explore a range of practical techniques.

M4 Airborne Photogrammetry Remote Sensing Simulation: Understand and master the definition and classification of aerial photogrammetry and Remote Sensing; understand the mission, history and current status of aerial photogrammetry and Remote Sensing; master and thoroughly understand the basic theoretical basis for aerial photogrammetry. Through the study of this course, students are required to understand the whole process of photogrammetry operations, and they can analyse and interpret the aerial photogrammetry process from theory and practice and provide a theoretical basis for solving practical problems.

M5 Digital Image Processing I.: This module aims at introducing basic and advanced techniques of digital image processing. It covers the fundamental concepts required to understand and apply commonly used and more advanced algorithms for pre-processing of remotely sensed data, image manipulation, characterization, segmentation and feature extraction in direct space.

M6 Digital Image Processing II. / Image Classification and Interpretation: This module aims at introducing basic and advanced techniques of digital image processing. It covers the fundamental concepts required to understand and apply commonly used and more advanced algorithms for classification of remotely sensed data. It focuses on an image classification knowledge, techniques and skills for getting information from imagery and ability to solve complex tasks based on Remote Sensing. Emphasis is placed on gaining a practical understanding of the principles behind each technique and a consideration of their appropriateness in different applications. (For an example see Fig. 4)

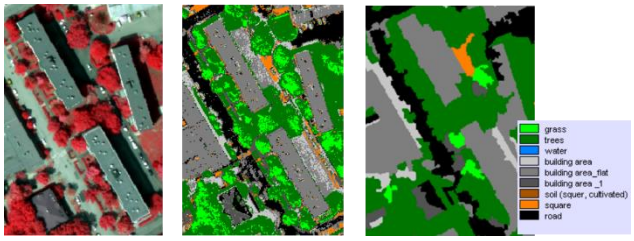


Fig. 4. Pixel and object approach for land cover classification (example)

M7 Available Software Applied in Remote Sensing: This module aims at introducing software tools available for applications in Remote Sensing. RS software landscape will be introduced and their practical methodologies of using this software for real-world problem solving will be explored. Some selected fully-fledged software packages including proprietary and free and open source will be used for image processing, analysis and visualization processes or for information extraction.

M8 Land Change Detection: This module aims at introducing concepts of land change and how basic and advanced techniques of digital change detection can be applied to detect and monitor changes on the land surface. It builds on fundamental concepts related to principles of Remote Sensing and basics of image processing and extends knowledge and skills of students into a temporal dimension. Topics are illustrated with examples and case studies supported by remotely sensed and ancillary data.

M9 Terrain modelling and analysis: introducing concepts of digital elevation models, and how basic and advanced techniques can be applied to analyse and visualize terrain and surface. It builds on fundamental concepts related to principles of photogrammetry, Remote Sensing, and image processing, and extends knowledge and skills of students into a third dimension. (For an example see Fig. 5)

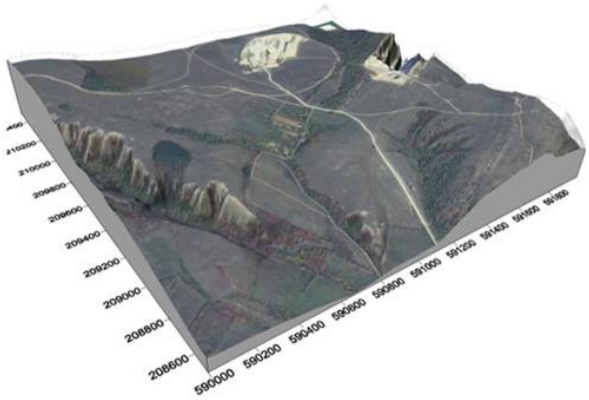


Fig. 5. An orthophoto projected onto a LIDAR-borne terrain model

M10 Remote Sensing and Geographical Information Systems (GIS): aims at the basic of geographic information system (GIS) and data type for integrated with Remote Sensing data. The integration between GIS and Remote Sensing data for data acquisition, analysed, presentation. Technology trends in Remote Sensing and GIS in the future.

M11 Application of Remote Sensing in Agriculture: This module aims at providing the principles, potentials and challenges of Earth observation technologies in the context of crop type mapping, change detection and, more generally, crop condition monitoring. It combines theoretical and practical sessions. The scope of the practical (computer-based laboratory exercises) is to consolidate the theory, to learn & train practical skills and to apply problem-solving methodologies to real-life examples. The main focus lies on the added values of multi-spectral data to retrieve crop bio-physical indicators and the use of multi-temporal data to study the changes and evolution of these indicators over time.

M12 Vegetation mapping and monitoring: providing students with knowledge of satellite Remote Sensing processing techniques and analysis for vegetation mapping and monitoring. Derived products of vegetation index, water stress etc. are generated to track the length of growing season, vegetation health and anomalies.

M13 Application of Remote Sensing (optical satellite data) in Forestry: This module presents different application possibilities of Remote Sensing in forestry. The main focus remains on tree species classification on different levels of detail. Additional contents are biomass / growing stock estimation and change detection applications for forest monitoring. Starting with the presentation of existing and free available data products the module provides some basics about different (active and passive) Remote Sensing data and analysis techniques.

M14 Monitoring the environment by using of RS: This module aims at introducing RS methods of mapping, monitoring and modelling of the resources for management of the environment and solving environmental problems. It will help learners to understand the methods of analysing environmental problems using RS tools for identification and application of possible solutions to support decision making.

M15 Application of Remotes Sensing in Water Management: In this module the theoretical background of the different water budget components is discussed. For each of the components the role of Earth observation is highlighted. To assess the water balance component, you need data. Data can be obtained from the ground and/or from satellite observations and/or already available satellite products. This module explains how to gather and combine data from these two sources, and where applicable, it refers to procedures already included in linked courses.

M16 Oceans/See and Coastal Monitoring: The main objective of the module is to give an overview about Earth observation solutions related to ocean and coastal monitoring. After a generic overview, which covers all application fields, some aspects will be described in detail, including water quality and sea surface monitoring and coastal processes monitoring.

M17 Remote Sensing in Archaeology: This module aims at providing students with knowledge of airborne and terrestrial Remote Sensing in archaeology in order to perform archaeological site detection and identification. It covers the fundamental concepts for data pre-processing and analysis of the multi various airborne and terrestrial Remote Sensing to derive an integrated archaeological mapping and interpretation of detected structure and features. (For an example see Fig. 6)

Fig. 6. Example for using orthophoto of an archeologic application in Székesfehérvár

M18 Application of (optical) Remote Sensing in Urban Environment: providing the principles, potentials and challenges of Earth Observation technologies in the context of land cover mapping, monitoring and change detection within urban environment. The scope of the practical (computer-based laboratory exercises) is to consolidate the theory, to learn and train practical skills and to apply problem-solving methodologies to real-life examples. The main focus lies on the added values of multi-spectral data to retrieve vegetation bio-physical indicators and the use of multi-temporal data to study the changes and evolution of these indicators over time. (For an example see Fig. 7)

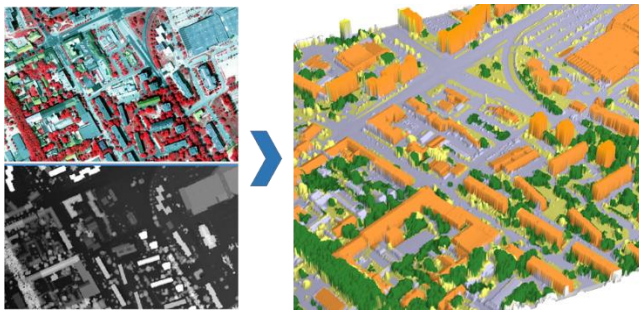


Fig. 7. Example for land cover mapping by OBIA

M19 Disaster Monitoring: This module aims at introducing the benefit of satellite images to monitoring disaster as flood, drought, tsunami, Hurricanes, earthquake, pollution. The data selection for analysed and detected each kind of disaster. The exercises and case studies allow students to explore a range of practical techniques.

M20 Weather and climate monitoring with Remote Sensing: introducing basics of satellite meteorology and climatology, principles of data processing and analyses. The main focus remains on particular spectral channels applications and the interpretation of processed weather and climate information. The module content will give also a broad overview of satellite weather and climate data records and databases as well as data integration towards assessing and monitoring climate change.

IV. Summary

The IRSEL project has aimed to develop an innovative learning platform, a Learning Management System for China and Thailand in the field of Remote Sensing applications. More specifically, a LMS at each of participating Asian universities has been established, a Knowledge Pool of high level e-Learning teaching materials for a wider scientific and engineering community has been created, finally, workshops, trainings and a summer school for the teaching and the administration staff has been organized.

The project has involved equipping people with the skills, experience and motivation to plan and develop complex educational materials. The project breaks down the traditional educational scheme what contribute to sustainability of results. It promotes (a) multi-method approaches: learning materials (modules), tutorials, webinars, meetings, etc., (b) critical, analytical and integrative thinking, (c) problem solving and research capability, (d) creative and innovative, (e) global and locally relevant information.

It has focused on the use of partnerships to build networks (knowledge platforms) and relationships, also has improved the communication between different sectors of discipline. The outcomes foreseen go in line with needs of educational and research institutions as well. In addition to the tangible results there are many areas of exploitation, such as cooperation in academic sector to offer new e-Learning approaches, also cooperation with geospatial industry to offer new innovative business products, furthermore the benefit of the society due to the improved contacts between research and application for sustainable development.

Acknowledgment

The paper is supported by the Erasmus+ Capacity Building in Higher Education, Key Action 2 with project ID 586037-EPP-1-2017-1-HU-EPPKA2-CBHE-JP.

Legal Notice

References

[1] European Commission: “Communication From The Commission To The European Parliament”, The Council, The European Economic And Social Committee And The Committee Of The Regions, Space Strategy For Europe, Com(2016) 705 Final, Brussels, 26.10.2016, p. 13.

[2] Ferrari, A., Cachia, R. & Punie, Y. 2009. Innovation and creativity in education and training in the EU member states: fostering creative learning and supporting innovative teaching. Luxembourg: European Communities. European Commission Joint Research Centre, JR 52374.

[3] Shifaw, E., Sha, J.M., Li, X.M., Shang, J.L., Bao, Z.C.: "Remote sensing and GIS-based analysis of urban dynamics and modelling of its drivers, the case of Pingtan, China", Environment Development and Sustainability 22(10), 2018, DOI: 10.1007/s10668-018-0283-z

[4] Shifaw, E., Sha, J.M., Li, X.M., Bao, Z.C., Ji, J.W., Chen, B.C.: "Spatiotemporal analysis of vegetation cover (1984–2017) and modelling of its change drivers, the case of Pingtan Island, China", Modeling Earth Systems and Environment 4(2), 2018, p. 1-19, DOI: 10.1007/s40808-018-0473-6

[5] Bao, Z.C., Sondago, T.H., Shifaw, E., Li, X.M., Sha, J.M.: "Monitoring of beach litter by automatic interpretation of unmanned aerial vehicle images using the segmentation threshold method", Marine Pollution Bulletin 137, 2018, p. 388-398, DOI: 10.1016/j.marpolbul.2018.08.009

[6] Mansberger, R., Bauer, T., Immitzer, M., Vuolo, F.: "Bildungsprojekte im Fachbereich Photogrammetrie, Fernerkundung und GIS: Herausforderungen, Chancen und Grenzen", Dreiländertagung der DGPF, der OVG und der SGPF in Wien, Österreich – Publikationen der DGPF, Band 28, 2019, p. 256-269

[7] Shifaw, E., Sha, J.M., Li, X.M., Bao, Z.C., Zhou, Z.L.: "An insight into land-cover changes and their impacts on ecosystem services before and after the implementation of a comprehensive experimental zone plan in Pingtan island, China", Land Use Policy 82, 2109, p. 631-642

[8] Lin, X., Su, Y.C., Shang, J.L., Sha, J.M., Li, X.M., Sun, Y.Y., Ji, J.W., Jin, B.: "Geographically Weighted Regression Effects on Soil Zinc Content Hyperspectral Modeling by Applying the Fractional-Order Differential", Remote Sensing 11(6), 2019, p. 636, DOI: 10.3390/rs11060636

[9] Shifaw, E., Sha, J.M., Li, X.M., Bao, Z.C., Bahir, A.L., Belete, M., Ji, J.W., Su, Y.C., Addis, A.K.: "Farmland dynamics in Pingtan, China: understanding its transition, landscape structure and driving factors", Environmental Earth Sciences 78(17), 2019, DOI: 10.1007/s12665-019-8537-0

[10] Verőné Wojtaszek, M., Balázsik, V., Földváry, L., Márkus, B.: “IRSEL: Innovation on Remote Sensing Education and Learning”, Proceedings of the H-SPACE 2020 Conference, Budapest, 26-27 February, 2020, in press.

Status of the DSinGIS Project

V. Balázsik
Institute of Geoinformatics
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
balazsik.valeria@amk.uni-obuda.hu

L. Földváry
Institute of Geoinformatics
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
foldvary.lorant@amk.uni-obuda.hu

A. Kovács
Institute of Geoinformatics
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
kovács.anna@amk.uni-obuda.hu

B. Márkus
Institute of Geoinformatics
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
markusbela@gmail.com

A. Pődör
Institute of Geoinformatics
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
podor.anna@amk.uni-obuda.hu

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

M. Verőné Wojtaszek
Institute of Geoinformatics
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
wojtaszek.malgorzata@amk.uni-obuda.hu

Gy. Busics
Institute of Geoinformatics
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
busics.gyorgy@amk.uni-obuda.hu

G. Nagy
Institute of Geoinformatics
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
nagy.gabor@amk.uni-obuda.hu

***Abstract*— The objectives envisaged with the DSinGIS project is to establish a missing PhD level of the Uzbek educational system. The project has already established an accredited Doctoral School in the field of GISc, developed its programme, defined the requirements, advanced supporting teaching and learning materials in English and/or Uzbek languages. Also, the education capacity of 5 leading Uzbek universities has been developed by creating a Knowledge Centre at each Uzbek partner universities containing an e-learning platform with a jointly developed knowledge pool. However, there are certain delayed activities, which resulted in the extension of the project period by 6 months. These are the development of a Joint Research Centre, organizing a Summer School and the GI2020 scientific conference, and also the visiting research in the EU for 7 more Uzbek PhD students. The present paper provides an overview of the status of the activities.**
Keywords: GeoInformation Science, PhD programme, e-learning

I. INTRODUCTION

GeoInformation Science (GISc) is a relatively young science, however, has its roots thousands of years. It integrates three traditional geosciences (firstly, geodesy as the science of precise spatial data acquisition; secondly, geography as the science of studying human and physical aspects; finally, cartography as the science of making maps). The integration of these sciences is based on the rapidly evolving computer science. The methods of GISc are widely applied in other sciences, essential in decision making for sustainable development. GISc provides the theoretical foundation of handling geo-related (ie. spatially referenced) digital spatial data acquired primarily by satellite-borne methods. As a result, GISc delivers an essential tool for interpreting, visualizing and analysing measurements of Earth Observation satellite missions, such as Remote Sensing. It makes use of the methods of geospatial analysis and modelling, information systems design, geocomputation and geovisualization.

With the coordination of the Óbuda University, an Erasmus+ Capacity Building in Higher Education, Key Action 2 project is conducted named Doctoral Studies in GeoInformation Sciences (abbreviated as DSinGIS). The wider aim of the project is to support Uzbekistan in sustainable development by GISc. The objectives envisaged with the project is to establish a missing puzzle from the Uzbek educational system after the MSc level has been completed and before the DSc is targeted. The project established an accredited Doctoral School in the field of GISc, developed its programme, defined the requirements, advanced supporting teaching and learning materials in English or Uzbek languages, all developed in accordance to international standards and in accordance to the Uzbek education system.

As a support for the new Doctoral programme, a network of activities is conducted to improve the educational and research capacity of the Uzbek society. Among these activities, an international network of the 5 leading Uzbek universities is established. Also, their education capacity is developed by creating a Knowledge Centre at each Uzbek universities containing an e-learning platform with a jointly developed knowledge pool. The knowledge pool is also supporting research activity of future PhD students. Furthermore, a Joint Research Centre, 5 research labs are developed to improve the research capacity of PhD programmes. Finally, annual GI conferences are organized to provide a platform for presenting research results.

There are several challenges in Uzbekistan, where GISc may efficiently support solutions, for such issues as climate change, land degradation, heavy use of agrochemicals, diversion of huge amounts of irrigation water from the two main rivers of the region, water scarcity, the chronic lack of water treatment, e.g. Aral Sea, or the growing threat to air quality.

The DSinGIS project shows synergies with other EU-funded project like “Furthering the Quality of Doctoral Education at Higher Education Institutions in Uzbekistan” [1], “Environmental Protection in Central Asia - Disaster Risk Management with Spatial Methods” [2] and “Sustainable development in rural areas of Uzbekistan” [3] projects.

The DSinGIS project was originally intended to be completed by 14 October 2020, but due to the COVID-19 pandemic, some important tasks must have been postponed, accordingly, the project run was requested to be extended by half a year. The aim of the present paper is to overview the results so far, and also to outline the remaining tasks. The paper also intends to update the status and planned timeline of the project activities presented at H-SPACE conference in February 2020 [4]. For up-to-date information on the DSinGIS project, visit the website at http://www.dsingis.eu/home/, which can be reached by the QR code on Figure 1 as well. Also, actual tasks and challenges of project implementation can be reached at [5].

Fig. 1. QR code of the dsingis.eu webpage

TABLE I. THE CONSORTIUM

Partners		
HEI	***abbr.***	***Country***
Obuda University	OU	Hungary
Paris Lodron University of Salzburg	PLUS	Austria
Royal Institute of Technology	KTH	Sweden
Leibniz Institute of Agricultural Development in Transition Economies	IAMO	Germany
Tashkent Institute of Irrigation and Agricultural Mechanization Engineers	TIIAME	Uzbekistan
National University of Uzbekistan named after Mirzo Ulug'bek	NUU	Uzbekistan
Karakalpak State University named after Berdakh	KSU	Uzbekistan
Samarkand State Architectural and Civil Engineering Institute	SamSACEI	Uzbekistan
Tashkent Institute of Architecture and Civil Engineering	TIAC	Uzbekistan
Associated partners		
HEI	***abbr.***	***Country***
State Committee of Republic of Uzbekistan on Land Resource	GKZGDK	Uzbekistan
Ministry of Higher and Secondary Specialized Education	MHSSE	Uzbekistan
Supreme Attestation Commission under the Cabinet of Ministers	SAC	Uzbekistan

II. THE CONSORTIUM

The project is implemented in an international cooperation of 9 higher education institutes (HEI) additionally to the 3 ministerial institutions contributing as associated partners. The majority of the consortium is Uzbek HEIs consisting of all relevant universities in the field of Geoinformatics. The list of the involved institutions is summarized in Table I.

III. ESTABLISHMENT OF THE DOCTORAL SCHOOL

The DSinGIS project focuses on the development of a PhD level educational system. The demand on establishing a Doctoral School in the field of Geoinformation Science is in line with the aims of the Ministry of Higher and Secondary Specialized Education of Uzbekistan (MHSSE). The project specified a PhD programme and methods to deliver its courses.

As a major achievement of the project is that the establishment of the Doctoral School has already been registered in 2019 by the Supreme Attestation Commission under the Cabinet of Ministers (SAC) as a state-recognized PhD programme.

The PhD programme consists of 18 courses divided into 6 compulsory courses and 12 advanced courses in 3 specializations. Among the 18 courses, 8 are prepared in English and 10 in Uzbek language. The rationale behind the share of languages is a consequence of the project partners are expected to contribute equally to the course development, each HEIs developing 2 courses. The structure of the programme and the content of the courses are as follows.

I. Compulsory courses prepared in English

1) Spatial representations and Spatial Data Infrastructures (SDI). The course provides a comprehensive overview on the state-of-the art of SDI, the underlying principles, as well as technological and non-technological components of SDIs.

2) Spatial statistics. The course aims at advancing knowledge on spatial data analysis and spatial statistics. It focuses on methods that are relevant in fields related to sustainable resource use and development of rural areas, such as land use change, climate change, soil degradation, and spatial analysis of well-being.

3) Global Navigation Satellite Systems (GNSS). This course provides the students with an in-depth knowledge about global navigation satellite systems, in particular positioning methods and algorithms as used in the fields of geoinformation science. The course focuses on high accuracy positioning methods, long term static observation methods for deformation monitoring and reference networks, and on atmospheric effects on GNSS signals.

4) Visually interfacing with spatial information. This course aims at introducing the complex field of visually interfacing with spatial information. Techniques and tools as well as concepts and standards to find, filter and visualize spatial data are presented. Technical skills and human-computer interaction competencies are built up.

5) Research methodology and scientific communication. This course introduces to students general research methods as well as practical research process, with focus on critical and creative thinking, addressing also scientific writing and communication in different forms and different media. Furthermore, social impact of scientific research, commercialization of research results through innovation is concerned.

6) Advanced remote sensing and digital image processing: This course aims at advancing remote sensing and digital image processing knowledge, techniques and skills for getting information from imagery and ability to solve complex tasks based on remote sensing. Emphasis is placed on gaining a practical understanding of the principles behind each technique and a consideration of their appropriateness in different applications.

II. Courses for specializations (there are 3 specializations, each consists of 4 courses, mostly in Uzbek language; when it is in English, it is noted).

II/1. Geodesy:

7) Geodetic Reference Systems. This course aims to deepen the theoretical knowledge and practical skills for the development and management of research projects.

8) Advanced theory of errors. This course consists of studying the theoretical foundations of multivariate statistical analysis in relation to the processing and analysis of geodetic measurements.

9) Satellite gravimetry (in English). The course aims at advancing on physical geodesy knowledge from observational aspects, focusing on obtaining positioning and physical information from satellite-borne observations.

10) 3D laser scanning and mapping by UAV. This course focuses on application of 3D laser scanners and unmanned aerial vehicles in analysing data and creating digital maps or update existing maps.

II/2. Geoinformatics:

11) Geo-databases and distributed architectures. This course is on developing techniques and skills for designing and building a geospatial database, as well as managing such distributed geodatabases, and working with multi-user spatial data base.

12) Advanced thematic mapping. The course supports candidates in cartography, thematic mapping, cartosemiotics, contemporary issues of spatial data representation, use of automation and tools in geovizualisation.

13) Advanced spatial analyses. This course aims to provide knowledge and skills necessary to investigate the spatial patterns, advanced analytical and practical skills to identify and apply the correct analytical tools for problem solving, and to appropriately interpret the analysis results.

14) Integration of remote sensing and GIS. The main aim of this course is exploring the synergies of integrated remote sensing systems and GIS.

II/3. GIS applications:

15) Spatial decision support in land management. The course is aimed to get an idea of the current regulatory and legal acts that regulate the subject of green law, and the application of this knowledge in practical activities, with emphasis on the current legislation of Uzbekistan.

16) Land use economics. This course is aimed at promoting the knowledge of doctoral students in the field of land use and its economics.

17) Spatial simulation of environment. This course is a critical introduction to spatial simulation of ecosystems, embedding of the PhD student in modern research practices, introducing a young scientist into up-to-date context and language of the simulation domain, including proper software background.

18) Sustainable resource management (in English). This course introduces key concepts related to natural resource management for food security and sustainable development. The course summarizes major trends in changes in resource management globally, across scales, and by geographic zone and country, considering also impacts of global climate change on water and land use, and their implications for sustainable resource management.

IV. Knowledge Pool

Beyond the courses of the PhD programme, a knowledge pool is built providing a theoretical background for the PhD students in the field of GISc and related disciplines. The knowledge pool consists of a glossary of geospatial terms, proceedings of annual scientific conferences organized within the frame of this project, relevant digital libraries and scientific journals, 8 modules developed within the Tempus GE-UZ project [6], 18 modules developed within the Erasmus+ DSinGIS project, online library/database such as Scopus, EBSCO, Sciencedirect and Proquest, products of PhD students in GISc (e.g. DSinGIS Grantholders' technical reports, papers/articles, presentations and posters), MSc, PhD thesis related to GISc, GIS related scientific journals and articles, and also GISCA&GI-Forum conferences Proceedings [7]. The structure of the Knowledge Pool is presented on Fig. 2.

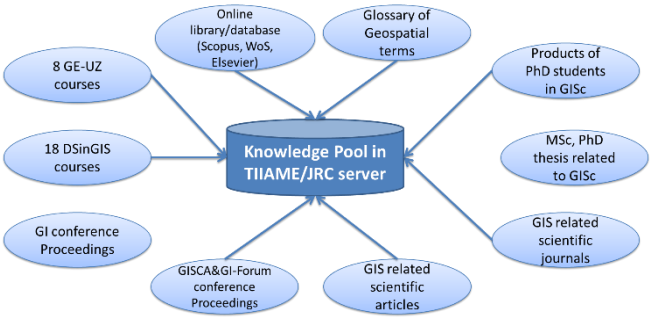


Fig. 2. Structure of the Knowledge Pool

The glossary of geospatial terms is containing more than 1000 terms in Uzbek, helping interdisciplinary communications.

GISc conferences are organized annually (by the date of the present paper, 2 of the planned 3 have already been completed successfully), the submitted papers are published in a conference proceedings and selected papers in local journals.

According to the DSinGIS work plan in total 15 selected doctoral candidates will receive grant to study in one of EU partner institutions (60 days/person) in 3 cohorts. The candidates have been carefully selected by the Admission and Examination Committee. The doctoral candidates will be actively involved in project evaluation, dissemination and

exploitation (see section VI. later). They should prepare travel reports focused on their research activities during their study visits in EU partner universities.

V. Knowledge Centres and Joint Research Centre

Advanced IT network is installed at each UZ partner institution, which handles a Moodle Learning Management System. The centres are equipped by GIS and Remote Sensing specific devices. The centres have an IT platform for acquiring and sharing knowledge, videoconference system will be installed for frequent communication between doctoral schools, and it helps collaborative, blended learning. Each centre is equipped with a server, a video conferencing system set, and 3 workstations with advanced GIS & Remote Sensing software. An interactive projector set supports professional presentations of doctoral students.

A sustainable Joint Research Centre (JRC) is also planned for the benefit of doctoral candidates in geoinformatics, not exclusively DSinGIS partners, but covering the interdisciplinary applications in the country. The JRC is established at TIIAME, but the related equipment purchase must have been delayed due to the COVID-19 related preventive actions in social distancing and closing several institutions. JRC will be equipped with advanced hardware and geospatial software. The technical / scientific staff will be offered by TIIAME for free. JRC will result directly improved quality of doctoral researches and increased collaboration within partners and wider GI community. The long run benefits of JRC: (1) UZ institutions are able to increase cooperation with EU research institutes, (2) Industry in geospatial sector able to offer innovative business products, (3) Society benefit from innovations, e.g. smart city concept to enhance quality, performance and interactivity of services, to reduce costs and resource consumption and to improve contact between citizens and government.

VI. Competence Development

There is a lack of experiences in the new educational methodologies and international tendencies in the Uzbek GI community. The aim of the project is to improve the knowledge, skills and competences of related Uzbek staff regarding management, administration, supervision and mentoring in doctoral studies. The competence development activity focuses on three target groups: 1) management/administration staff, 2) teaching staff and supervisors and 3) doctoral candidates.

There are trainings and workshops are organized in order to improve the competences of the different target groups of the Uzbek HEIs. Also, annual international GISc conferences are organized, which is a huge opportunity for the Uzbek Geoinformatic society to participate and present up-to-date researches in the field of GISc and related sciences. The main parameters of such events are summarized in Table II.

A major success of the competence development activities is the annually organized scientific conferences. On one hand, the number of the published papers is beyond expectation. But what is more important is that annually an increase of the level of the submitted papers can be observed. Also, the number of researches performed in international cooperation could be observed. Originally, 3 GI conferences were planned. In 2018 and 2019 it was implemented according to the plans. In 2020, the GI2020 conference was agreed to be organized jointly with the GISCA2020 international conference, several publications from the DSinGIS community has been submitted. The GISCA2020 conference was finally held in 1–3 June 2020 in an online form. The online participation provided a kind of challenge for the young Uzbek GI community, and their participation turn to be below the expectations. In order to provide an appropriate forum for the young Uzbek colleagues, a face to face conference has been decided to be held as soon as the international pandemic situation enables organization of such conferences. By September of 2020, the proposed date of the 3rd GI conference is January or February 2021.

TABLE II. Events for competence development

event	target group	topics	date and location
Training for management and administration	administrative staff, senior managers	management, administration, quality management, quality enhancement, internationalization issues	22-26 May 2018 Hungary
Workshop on learning support methodologies	Teaching staff, scientific advisors	learning support methodologies, eLearning tools and competencies in mentoring	15-18 October 2018 Uzbekistan
1st GI Conference	Teaching staff, research community, PhD students and PhD candidates	presenting and discussing research findings, interests of young and senior, Uzbek and international GI scientists	19-20 October 2018 Uzbekistan
Workshop on interdisciplinary doctoral courses	Teaching staff, scientific advisors	experiences on interdisciplinary GISc issues of doctoral courses	June-July 2019 Austria
Training on supervision and research methodologies	Teaching staff, scientific advisors	supervision and research methodologies in GISc	15-18 October 2019 Uzbekistan
2nd GI Conference	Teaching staff, research community, PhD students and PhD candidates	presenting and discussing research findings, interests of young and senior, Uzbek and international GI scientists	22-23 October 2019 Uzbekistan
Geoinformatics Summer School	PhD students	practical and methodological skills for advanced use of spatial analysis methodologies and techniques of GISc	postponed Uzbekistan
GISCA Conference [7]	Teaching staff, research community, PhD students and PhD candidates	presenting and discussing research findings, interests of young and senior, Uzbek and international GI scientists	1-3 June 2020 online
3rd GI Conference	Teaching staff, research community, PhD students and PhD candidates	presenting and discussing research findings, interests of young and senior, Uzbek and international GI scientists	postponed Uzbekistan

Also, a relevant contribution to competence development is that the project provides scholarship for the young PhD student generation for a 2-month visiting research at an EU partner university to get international experiences. The list of the awarded researchers and their parameters of visit is listed in Table III.

TABLE III. SCHOLARSHIP AWARDEES

name	HEI	host	date
Yakhshimurad Khudaybergenov	KSU	IAMO	October - November 2018
Mamanbek Reimov	TIIAME	OU	January – March 2019
Zokhid Mamatkulov	TIIAME	OU	January – March 2019
Otabek Avezbaev	TIAC	PLUS	March – May 2019
Kuwatbay Bekanov	KSU	IAMO	April – May 2019
Sitora Sodikova	SamSACEI	PLUS	May – June 2019
Ilhom Abdurahmanov	TIIAME	OU	February – March 2020
Medetbay Uteuliev	KSU	OU	February – March 2020
Azizjon Ruziev	NUU	OU	postponed
Abdujalil Muminov	NUU	IAMO	postponed
Vohidjon Niyazov	SamSACEI	OU	postponed
Sherzod Uzokov	SamSACEI	PLUS	postponed
Akbarjon Khamraliev	TIIAME	OU	postponed
Barno Khalilova	TIIAME	PLUS	postponed
Ilkhomjon Abdullaev	NUU	OU	postponed

The candidates are selected carefully by an Admission and Examination Committee. In addition to UZ supervisors EU teachers are assigned to mentoring learning, supporting researches. As it can be seen in Table III, 7 candidates among the 15 could have not fulfil their travel due to the COVID-19 pandemic.

VII. SUMMARY

With the coordination of the OU, an Erasmus+ Capacity Building in Higher Education, Key Action 2 project named Doctoral Studies in GeoInformation Sciences is conducted. The wider aim of the project is to support Uzbekistan in sustainable development by GISc. The objectives envisaged with the project is to establish a missing puzzle from the Uzbek educational system after the MSc level has been completed and before the DSc is targeted. The project established an accredited Doctoral School in the field of GISc, developed its programme, defined the requirements, advanced supporting teaching and learning materials in English or Uzbek languages, all developed in accordance to international standards and in accordance to the Uzbek education system.

As a support for the new Doctoral programme, a network of activities is conducted to improve the educational and research capacity of the Uzbek society. Among these activities, an international network of the 5 leading Uzbek universities is established. Also, their education capacity is developed by creating a Knowledge Centre at each Uzbek universities containing an e-learning platform with a jointly developed knowledge pool. The knowledge pool is also developed supporting research activity of future PhD students. Furthermore, a Joint Research Centre, a research lab is to be developed to improve the research capacity of PhD programmes. Finally, annual GI conferences are organized to provide a platform for presenting research results.

As a consequence of the COVID-19 pandemic, certain activities could have not been completed according to the schedule, which resulted in the extension of the project period by 6 months. These activities are the development of a Joint Research Centre, the organization of the Summer School and of the face-to-face GI2020 scientific conference, and also the visiting research in the EU of 7 Uzbek PhD students could have not been completed.

ACKNOWLEDGMENT

The paper is supported by the Erasmus+ Capacity Building in Higher Education, Key Action 2 with project ID 585718-EPP-1-2017-1-HU-EPPKA2-CBHE-JP.

LEGAL NOTICE

REFERENCES

[1] Erasmus+ CBHE Project “UZDOC: Furthering the Quality of Doctoral Education at Higher Education Institutions in Uzbekistan”, http://www.uzdoc.eu/

[2] Erasmus+ CBHE Project “Environmental Protection In Central Asia (EPCA): Disaster Risk Management With Spatial Methods”, http://eu-epca.eu/

[3] EU funded Project “Sustainable development in rural areas of Uzbekistan”, http://www.uzruraldev.eu/en

[4] Földváry, L., Balázsik, V., Márkus, B., Pődör, A., Verőné Wojtaszek, M., Abdurahmanov, I., Reimov, M.: "Doctoral School in Geospatial Science in Uzbekistan", Proceedings of the H-SPACE 2020 Conference, Budapest, 26-27 February, 2020, in press.

[5] Markus B. et al.: “DSinGIS project Handbook”, Szekesfehervar, 2017, p. 86.

[6] Tempus CBHE Project “GE-UZ: Geoinformatics: enabling sustainable development in Uzbekistan”, http://www.dsingis.eu/wp-content/uploads/2020/01/GE-UZ-brochure_w_link.pdf

[7] GIS in Central Asia Conference – GISCA 2020 on topic “Applied Geoinformatics for Sustainable Development”, https://gisca20.wordpress.com/

Virtual Reality Laboratory in the Help of Research and Education

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

Marta Seebauer
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
marta.seebauer@amk.uni-obuda.hu

Dániel Marton
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary

Károly Széll
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungarykaroly.szell@amk.uni-obuda.hu

Bálint Tőke
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary

Mátyás Rick
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary

***Abstract*—In 2020 a new Virtual Reality Laboratory was developed and equipped at Óbuda University Alba Regia Technical Faculty. The new laboratory consists of high performance PCs, VR Head Mounted Devices (HMD), complex depth cameras, projectors for the education and research in the field of virtual, augmented, and mixed reality. This laboratory helps the newly accredited Mobile Application and Game Development Specialization of the IT Engineer MSc course, the Mechanical Engineer BSc. and Mechatronic Engineer MSc. courses. Furthermore, it makes possible some educational projects, and research projects that are listed in this article.**

Keywords—Virtual Reality, Augmented Reality, education, Spatial vision

I. INTRODUCTION

Virtual reality, augmented reality and mixed reality have become part of our everyday lives. Historically publications on virtual reality have appeared in the scientific world since the 1960s. Then, at an increasing pace, the enhancement of related hardware tools, software tools and development environments took place. The developed systems can be of great interest in almost all segments of industry and society, highlighting healthcare and education as already successful applications[1]–[11].

Figure 1 shows the prevalence of VR and related concepts in scientific vocabulary. The two basic terms associated with it are Virtual and Augmented Reality. Virtual Reality: "A high-end user interface that involves real-time simulation and interaction through multiple sensorial channels (vision, sound, touch, smell, taste)" or "A computer-generated, immersive, multi-sensory information program which tracks a user in real time"

Augmented VR: the idea of taking what is real and adding to it in some way so that user obtains more information from their environment.

The number of publications containing the term VR in the title, based on Google Scholar search, grew rapidly between 2000 and 2010 and today the number of publications exceeded the 1,000,000. The number of publications containing augmented reality in the title

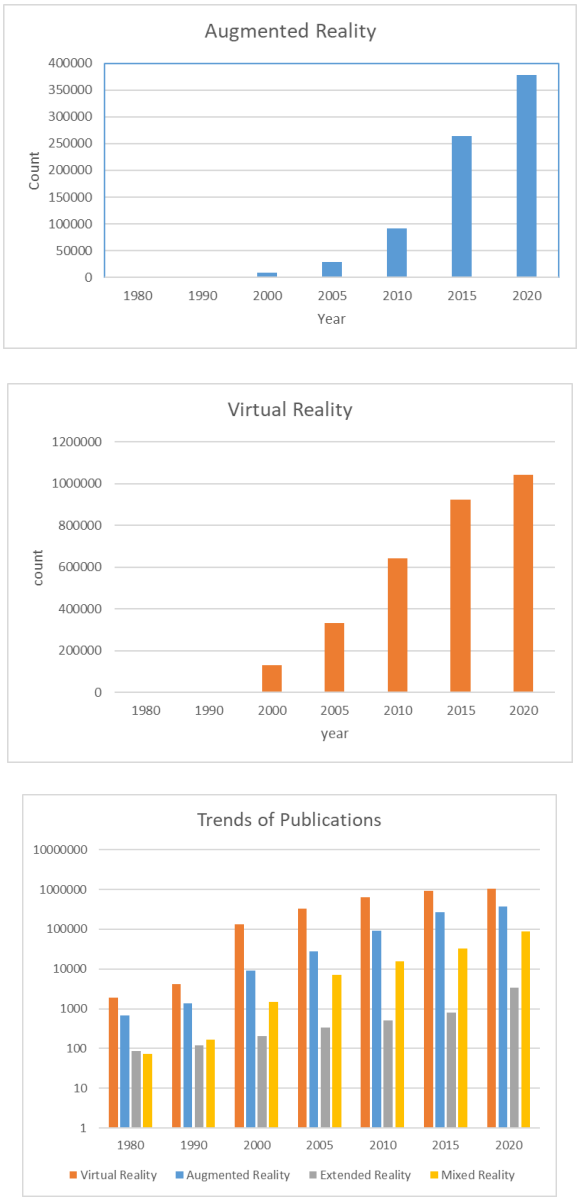


Figure 1 Number of VR related publications

increased rapidly after 2010 and till now reached hundreds of thousands. It is easy to see that this amount of publication unfortunately does not allow to follow the topic in every detail. Complementary concepts, mixed and augmented reality are also popular. The background of the large number of publications is that however the basic concepts and principles were laid down in the previous century, but the implementation struggles with many difficulties and childhood diseases, both in terms of hardware and software. However, the areas of application are very promising in the light of psychological studies. With the connection of COVID-19 pandemics, the impact of VR in distant education and in business communication has increased [6][11][1].It can be stated that there are currently no suitable tools and software that allow continuous (full-time) learning or work in a VR environment. We can expect a change in this regard in the future, as the tools will become easier and more convenient due to the ever-evolving technology. The psychological study of virtual reality is also intensive. This includes testing immersion, examining the psychological impact of activities in VR systems, and analyzing the health effects of VR devices to reduce adverse effects. VR is an explosively growing market segment around the world and its role in IT training became more central. In this article, we will present the construction process and results of the VR laboratory established at the Alba Regia Technical Faculty of Óbuda University, and evaluate the usability of the tools in education, as well as formulate research projects related to VR.

II. Composition of a VR Laboratory

In the spring of 2020, a new virtual reality laboratory was installed at the Alba Regia Technical Faculty of Óbuda University [12][13][5][7].

Two newly purchased PCs which are suitable for 3D design and computer graphics and related VR MR devices have been acquired. Laboratory equipment is listed in Table 1.

Table 1. Virtual Reality and Game Development Devices

Devices	Attributes
PC	PCT Blade Base Intel Intel Core i7-8700 3400MHz 8MB LGA1151 8GB DDR4 memory module 256GB 2,5" SSD SATA 2TB 3,5" HDD DVD inner writer Windows Pro 10 32/64
VR device	Oculus Quest VR Headset 128GB+Handheld Controller
MR device	MS HoloLens2 (under acquisition)
Depth sensors	Kinect for Xbox One Kinect for Xbox 360
Game platform	Xbox One

These devices complemented the existing PCs from previous purchases, Xbox One devices + Kinect for Xbox One sensors and Kinect for Xbox 360 depth sensors. Currently, the VR lab has a capacity of 10 people. In the future, the PCs are planned to equipped with suitable graphics cards, as well as to purchase additional sensors (Kinect for Azure, Orbbec).

An important aspect of the laboratory equipment was that it could be conveniently used for project-based education as well as research projects.

The installed software configuration enables computer graphics, 3D modeling, 3D sensing, and their use in software development and testing. We managed to create a test environment in which the software execution process of VR devices by a user can be objectively evaluated. The image seen in the device can be projected and recorded along with the interactions of the user.

Figure 2 Oculus Quest and HoloLens2

III. Methods

The following methods were used during the design of the system.

Depth sensor

The MS Kinect SDK was installed and applied to query the Kinect Depth Sensor.

Database

The data containing the movements were saved in a MongoDB database [10]. An MLP neural network was used to analyze the movements.

Implementing virtual reality

The Unity game engine provides easy integration with virtual reality devices, including Oculus devices, in this case by installing an Oculus SDK that can be easily integrated into the project.

Assets and VR development

The Oculus Quest development team has developed an integration package for the Unity game engine that includes a stable interface that allows the engine to

communicate with the virtual reality device and provide ready-made solutions to take full advantage of the device's features. This package is called "Oculus Integration" which is available through the Unity Asset Store. During the development of the source code, using the pre-made model in this package, the "prefrab" model, which provide the ability to move the player, called "OVRPlayerController". This model was created, which includes the usability of the HMD (Head Mounted Devices), provides different controller models and the player's real-time virtual spatial tracking.
The image of the Oculus VR headset was projected onto the projector with the SideQuest 3rd party software[14].

IV. PROPOSALS

The newly established research site can serve as a location for educational and research projects.

A. Educational projects

From the 2019/20 academic year, we managed to start the specialization called Mobile Application and Game Development in the master's degree in IT engineering with the John von Neumann Faculty of Informatics of Óbuda University. The Software Tools of Game Development and Multiplatform Game Development subjects taught in the specialization are based on this lab. The laboratory is suitable for the implementation of the course entitled Project Work, furthermore it can be used in the robotics engineering education by the implementation of robot simulation tasks. The topic of many dissertations is VR and AR, which can be realized by this laboratory.

A. Research projects

Within the framework of the intelligent data analysis research group of Alba Regia Technical Faculty of Óbuda University, we formulated projects that fit the research of the working group. Our main objectives are applications of 3D data visualization and data analysis.

Our primary goal is to visualize tabular data in an augmented and mixed reality environment. Our goal is to visually link the data extracted from the database or received from different data sources with the corresponding data source object to assist in the exploratory data analysis. Our second important project is the quantitative analysis of human behavior in the VR environment. During the investigation the movements of the user during the VR software execution are measured and then analyzed. The system developed in this way will be suitable for examining many basic research issues in quantitative psychology, sociology and pedagogy. In addition, it will be used for applied industrial research and for monitoring and supervising practical educational and industrial design or optimisation projects.

V. RESULTS

In the previously developed system, real-time recording and storage of human movements in a MongoDB database and high-precision classification of some basic forms of movement with an MLP neural network were accomplished[10]. Three games have been developed to make tests in the developed VR system. To test spatial orientation and navigation, we developed a software that generates a spatial maze in which the user aim is spatial orientation and finding the exit of the maze. Our second test application is a motion coordination application that is a VR implementation of a Darts game. The third application is suitable for examining a complex issue.

Figure 3 3D shape connection game with lowpoly design. The user gets different complexity 3D prism generated by the software according to the defined complexity level. The user has to rotate it to the appropriate orientation and fit into the keyhole.

Figure 4: Darts game with realistic design The user can play a classic darts game alone or with a different level computer player.

It is a 3D object matching game, in which the 3D prism generated by the game must be rotated into the appropriate orientation by the player to fit in a given scheme. The software makes possible to examine the hand movements implemented in the VR environment as well as the user's spatial vision. Measurements of the three test programs have begun. The following is a detailed re-measurement of the three test applications as well as an evaluation of the measurement results. In the future, it is worth further developing the tool system by solving the following tasks. By using several depth cameras at the same time, it is possible to examine a larger area or a group consist of more

than six people. By calibrating the depth cameras, it is possible to characterize the movements of the examined person in the real metric system and with spatial coordinates, as well as to measure them with the movements in the virtual space units. The next goal is to create test applications in which the subject can perform tasks in cyberspace without using a handheld controller.

Figure 5 VR Labyrinth test game in cyberpunk style. The user has to direct a small virtual machine to find the exit from the two layered 3D labyrinth.

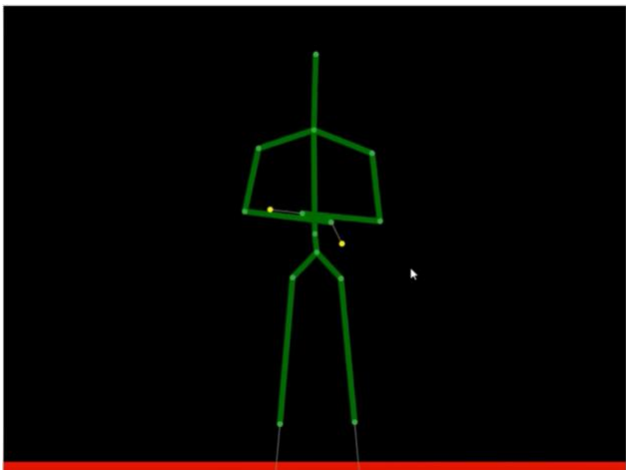


Figure 6 Skeleton picture by the detected 20 joints of a standing person.

VI. Conclusion

Virtual reality has become widespread worldwide and has become a technology gaining serious use in more and more areas. Alba Regia Technical Faculty of Óbuda University wanted to join this process. The method was to develop the appropriate tools and to set the related educational and research goals. In the article the development process was presented, tools, the purpose of the development and the usability of the developed laboratory for the educational and research purposes of the University of Óbuda Virtual Reality Laboratory was showed. The completed laboratory can be used in almost all training programs, in addition, in project-based education, development projects, dissertations and diploma theses can be applied to it. In the field of research projects, the intelligent data analysis research group can be primarily connected to the research and they want to use the tools in the future in data analysis and data visualization projects. The formulated and planned VR system will also be applicable in basic research, applied research and R&D projects cooperation with industrial companies.

Acknowledgment

I would like to thank the help and the effort of colleagues and students. The financial support of this paper was from Arconic 2019 grant.

REFERENCES

[1] T. Alldieck, M. Magnor, W. Xu, C. Theobalt, and G. Pons-moll, “Detailed Human Avatars from Monocular Video.” 2018.

[2] S. Jayaram, H. I. Connacher, and K. W. Lyons, “Virtual assembly using virtual reality techniques,” *CAD Comput. Aided Des.*, vol. 29, no. 8, pp. 575–584, 1997.

[3] E. S. Richir, J. S. Munoz-arango, C. Cruz-neira, and S. Beer, *VRIC 2019*. 2019.

[4] M. R. Deva, N. Bissett, B. Gb, W. Au, M. Duursma, and W. P. Hills, “(12) United States Patent,” vol. 2, no. 12, 2016.

[5] M. Slater, “Enhancing Our Lives with Immersive Virtual Reality,” vol. 3, no. December, pp. 1–47, 2016.

[6] J. Achenbach, T. Waltemate, M. E. Latoschik, and M. Botsch, “Fast Generation of Realistic Virtual Humans,” 2017.

[7] M. Gonzalez-franco and J. Lanier, “Model of Illusions and Virtual Reality,” *Front. Psychol.*, vol. 8, no. June, pp. 1–8, 2017.

[8] U. S. Ci, “United States Patent,” vol. 2, 2018.

[9] M. Speicher, “Shopping in Virtual Reality,” no. March, 2018.10] B. Szendrey, I. Járomi, G. Nagy, B. Hartal, Á. Rozgonyi, and É. N. Hajnal, “Depth Cameras and their Usefulness in Social Places,” in *AIS 2019 : 14th International Symposium on Applied Informatics and Related Areas organized in the frame of Hungarian Science Festival 2019 by Óbuda University*, 2019, pp. 17–20.

[11] N. Zhou, “Tracking of Deformable Human Avatars through Fusion of Low-Dimensional 2D and 3D Kinematic Models,” 2019.

[12] V. Wanick, J. Castle, and A. Wittig, “APPLYING GAMES DESIGN THINKING FOR SCIENTIFIC DATA VISUALIZATION IN VIRTUAL REALITY ENVIRONMENTS Resumo,” in *XIII International Conference on Graphics Engineering for Arts and Design*, 2019.

[13] P. Zheng *et al.*, “Smart manufacturing systems for Industry 4.0: Conceptual framework, scenarios, and future perspectives,” *Front. Mech. Eng.*, vol. 13, no. 2, pp. 137–150, 2018.

[14] “SideQuest.” [Online]. Available: https://sidequestvr.com/setup-howto. [Accessed: 20-Oct-2020].

Accuracy of UAV based data collection (case studies)

Balázsik Valéria – Dr. Tóth Zoltán*
* Óbuda University, Alba Regia Technical Faculty of Technology of the Institute of Geoinformatics, Székesfehérvár, Hungary
toth.zoltan@amk.uni-obuda.hu

Abstract — Nowadays many new data acquisition technologies have been appeared. Due to the increasingly advanced digital devices, high resolution cameras and utility of fast processing, Unmanned Aerial Vehicles (UAV) have become a prominent feature of the various remote sensing procedures. In the past, UAVs were only available for military purposes but over the last decade the 'drone' equipment has become easily accessible to everybody. While in the past photogrammetry was only used by qualified professionals, it is now applied successfully in many fields. Even though UAVs are widespread in surveying practice too, the related accuracy issues are not determined regularly. The manufacturing companies may calibrate the instruments of the device one by one, e.g. camera, GNSS receiver, in which way the nominal precision of an UAV, as an integrated system can be derived considering the error propagation law. However, in the practice the actual precision may notably differ from the nominal one, caused by different internal and external effects, such the flying stability of the device (platform), weather conditions, features of the observation object, etc. Because the benefits of the technology may justify the use of data in topographic mapping, cadastral mapping or even engineering geodesy, we have conducted studies to determine the accuracy that can be achieved using UAV as an integrated data acquisition tool. First, we developed a test field in rural area for investigations. The results of the test flights conducted there were determined that what conditions influence the accuracy of the survey. In addition, cadastral measurements were performed in the populated area using UAV and GNSS technologies and the results were compared with digital cadastral map data. Investigating the feasibility of engineering geodetic applications, we have performed an artificial surface survey using UAV technology. The results and conclusions of these studies are presented in this article.

I. INTRUDUCTION

Nowadays, UAV technology has become part of everyday practice for certain applications. In our previous articles we have provided examples in archeology [4] and mining [6]. After analysing the test measurements, we present two additional practical examples in which UAV technology has been applied. The studies are based our students' thesis works [2, 5].

II. CREATING A LARGE SCALE ENGINEERING MAP

In many engineering applications, the high relative reliability of height data is important. These technologies are typically slow and costly measurements such as levelling or conventional surveying, and usually deliver low resolution map. The purpose of our study is to determine the data accuracy and reliability of data acquisition using UAV, and to answer the question that the short range and high number of overlaps UAV technology may replace the traditional surveying for engineering geodetic tasks. In the example shown on Figure 8, measurements were made on a test field in a horizontal parking lot. From 4 positions (red dots), 6 photogrammetry GCP (yellow dots) and 35 test points were determined conventionally by a total station survey in a local network, which after has been adjusted and

inserted into the national reference system of Hungary (Fig. 1).

Figure 1. Test field with point marks in a parking lot [2]

Each test point was determined by as many redundant measurements as possible. The results of a purely total station survey were determined by adjustment as an independent network. Network-specific statistics are listed in the Table 1.

	ΔY[mm]	ΔX[mm]	ΔH[mm]
average	2,4	2,5	0,7
maximum	3,8	3,7	1,3

Table 1. Residual errors after the adjustment [2]

The data in Table 1 show that the network is one order of magnitude more accurate than one would expect from a photogrammetric survey, so it may be suitable for examination. The images were taken with a vertical camera axis, 80% overlapping in-line and inter-row directions, with a low, 29 m flight altitude and an estimated 0.9 cm/pixel geometric resolution. In addition, oblique images (with 60-degree camera axis) and lower altitude images were taken with a smaller, approximately 60% overlapping (Fig. 2). Processing was done with Agisoft Photoscan, using the conventional steps (image alignment, point cloud generation, absolute orientation, surface generation, orthophoto generation). 6 GCPs were added to process. The residual errors of GCP on the georeferenced model were less than 1 cm. In order to compare the results of the traditional and the photogrammetric surveys, the point cloud was displayed in a general-purpose commercial CAD program supplemented with orthophoto mosaic of the area (Figure 3). Test points were manually selected for comparison, thus, of course, the results are also influenced by the quantization of the point cloud.

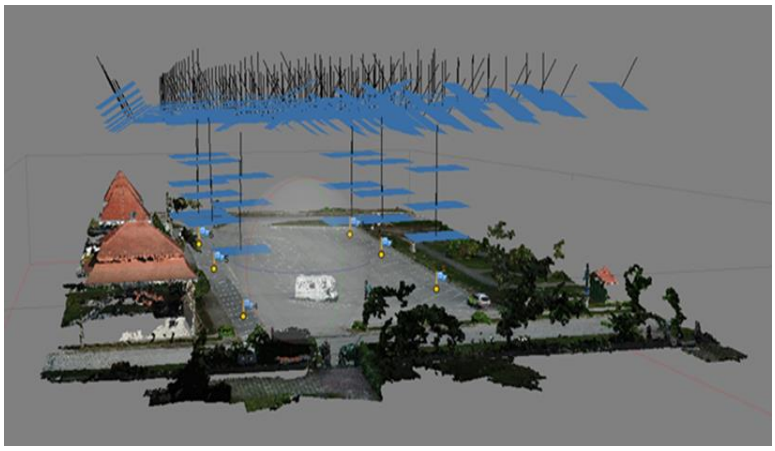


Figure 2. The location and position of images [2]

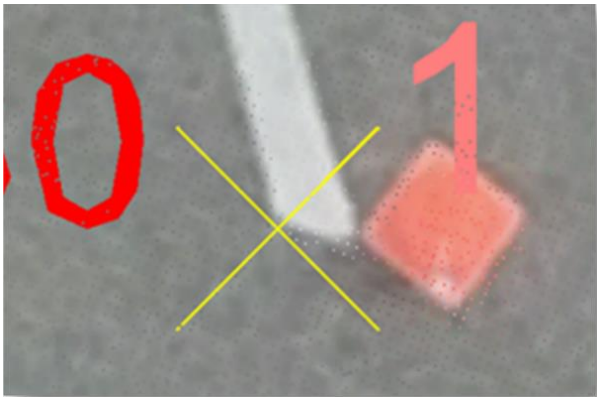


Figure 3. The location and position of images [2]

Table 2. summarizes the typical coordinate deviations of the test points measured by the ground survey and determined by UAV technology.

	ΔY[m]	ΔX[m]	ΔH[m]
average	0,012	0,012	0,007
RMS	0,009	0,008	0,004
maximum	0,044	0,031	0,018

Table 2. Deviations on the test points by processing all images [2]

The processing of the photogrammetric survey was repeated using only images with vertical axis (nadir images) and significantly less favourable results were obtained (Table 3).

	ΔY[m]	ΔX[m]	ΔH[m]
average	0,049	0,061	0,063
maximum	0,098	0,199	0,103

Table 3. Deviations on the test points by processing only nadir images [2]

The results obtained with photogrammetric processing are clearly improved by the favourable arrangement of the images. Table 2 shows the reliability achieved with UAV technology, which practically determines its reliable applications. Based on the values obtained, general-purpose surveys can definitely be replaced by this technology. The further advantages of the UAV technology are that the resolution of survey is better, and the time required for the survey is significantly less than conventional (see polar survey) methods. For high-accuracy surveys, additional oblique images can improve the accuracy to achieve the expected quality.

III. UAV TECHNOLOGY IN CADASTRAL MAPPING

In many countries of the world, including Hungary, no digital cadastral maps for the entire country, made by a new survey is available [7, 8]. At those regions of the country where a new survey for digital mapping has not been conducted, older analog maps have been digitized. However, older maps were made in different projection systems, so besides the digital transformation, they had to be transformed into the currently used national projection system. Thus, in addition to errors of drawing and digitizing of the analog maps, the transformation procedure also resulted errors in the data of the digital map. While the reliability of the parcels of the new survey digital cadastral maps is appr. ±9 cm, the reliability of the corner points of the buildings is appr. ±15 cm; nevertheless, in the case of digitized, projected old analog maps these values can be up to meters. UAV photogrammetry can produce not only high-resolution orthophotos but also 3-dimensional surface models. This can be used to measure the walls of buildings, which were not usually seen on orthophotos produced using traditional aerial photogrammetry, due to the overhang the roof.

Figure 4. Dense cloud of the area [5]

Our investigations were carried out on a rural block containing 20-30 land parcels. At low flight altitudes of 35 m were taken images, with overlapping 75% in-line and 65% inter-row directions. As in the previous example, in addition to vertical axis images, oblique images also were made with ~65°. With a GNSS receiver, we measured 9 GCP for photogrammetric processing and also determined the boundaries of the parcels. The dense spot cloud of the area (Figure 4) was produced by the usual processing method and its orthophoto mosaic (Figure 5) for further examination.

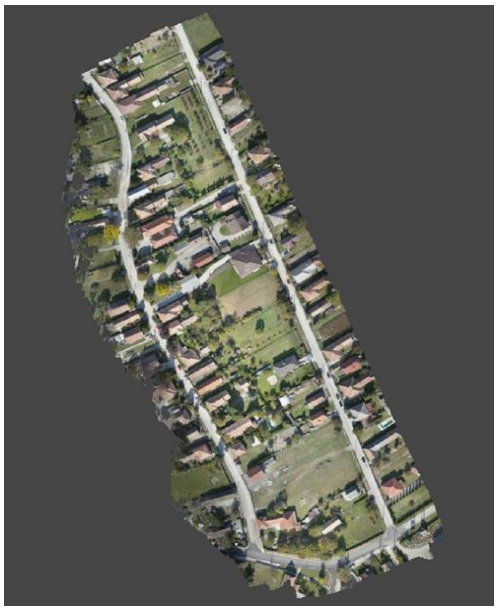


Figure 5. Orthophoto mosaic of the block [5]

Our aim was to investigate that the resulting orthophoto is suitable to replace the cadastral survey. We studied how reliably the elements to be mapped (eg. fence corners) can be identified and how accurate and efficient the technology is. To do this, parcel by parcel we compared the map elements measured by the conventional method and identified from the orthophoto (Fig. 6). The Table 4, 5 and 6 show some typical values.

Figure 6. Cadastral map displayed on orthophoto [5]

	GNSS		UAV			
id 112	**Y**	**X**	**Y**	**X**	**ΔY**	**ΔX**
239	638738,80	196847,22	638738,70	196847,22	0,10	0
231	638748,77	196822,08	638748,67	196822,07	0,10	0,01
237	638711,63	196798,56	638711,55	196798,49	0,08	0,07
238	638697,28	196821,32	638697,20	196821,35	0,08	-0,03

Table 4. [5]

	GNSS		UAV			
id 129/1	**Y**	**X**	**Y**	**X**	**ΔY**	**ΔX**
131	638603,50	196973,68	638603,58	196973,61	-0,08	0,07
132	638604,52	196959,52	638603,58	196959,53	-0,07	-0,01
138	638646,62	196981,67	638646,70	196981,62	-0,08	0,05
140	638639,59	196993,46	638639,62	196993,54	-0,03	-0,08

Table 5. [5]

	GNSS		UAV			
id 131	**Y**	**X**	**Y**	**X**	**ΔY**	**ΔX**
102	638603,29	197032,63	638603,23	197032,61	0,06	0,02
119	638655,92	197060,91	638655,84	197060,96	0,08	-0,05
125	638606,77	196992,04	638606,69	196992,06	0,08	-0,02
120	638668,68	197026,59	638668,75	197026,60	-0,07	-0,01

Table 6. [5]

According to the Tables, it can be concluded that the magnitude of the deviations is within the tolerance expected from cadastral surveys. Note that these values are significantly influenced by the reliability of the detail point identification. Due to the nature of some elements, this technology is not, or only partially applicable.

CONCLUSIONS

In this article we have presented accuracy estimates of UAV survey based on image matching photogrammetry. Accuracy estimate has also been presented for two relatively less investigated applications (surveying for cadastral, and large scale engineering map). Beyond the known accuracy results, the technology can provide a good basis for cost requirements of the survey. Based on the results of the test measurements and it's low cost, the technology is expected to spread in many areas.

REFERENCES

[1] V. Balázsik, Gy. Busics, P. Engler, R. Farkas, L. Földváry, T. Jancsó, A. Kiss, Z. Tóth, M. Verőné Wojtaszek, Establishment of an UAV calibration field on Iszka Mountain and the first results of accuracy tests. Remote Sensing Technologies and GIS Online, ISSN: 2062-8617, eISSN: 2062-8617, 6(3), P. 448-454 (in Hungarian). (2016)

[2] Á. Kleszky, Large Scale Mapping Traditionally and with UAV Technology. Thesis Work. Székesfehérvár (in Hungarian). (2020)

[3] National report on the state of land resources of the Republic of Uzbekistan. State Committee of the Republic of Uzbekistan on Land Resources, Geodesy, Cartography and State Cadastre, Tashkent. P. 64. (2019)

[4] K. Pokrovenszki, B. Vágvölgyi, Z. Tóth, Practical Experience with the 3D Photogrammetric Methods used at the Excavation of Csókakő Castle. Magyar Régészet P. 20-27. (2016)

[5] T. Schneider, Application of UAV-borne Data Acquisition for Land Registry, Thesis Work. Székesfehérvár. (in Hungarian). (2020)

[6] Z. Tóth, Remote-Sensing-Methods-for-Open-Pit-Mine-Surveying. In: Orosz G. (ed.) AIS 2018 - 13th International Symposium on Applied Informatics and Related Areas, Székesfehérvár. P. 4. (2018)

[7] J. Katona, M. Gulyas Horoszne, Determination of factors modifying land value based on spatial data In: Drótos, Dániel; Vásárhelyi, József; Czap, László; Ivo, Petráš (szerk.) Proceedings of the 19th International Carpathian Control Conference (ICCC 2018) Piscataway (NJ), Amerikai Egyesült Államok : IEEE, (2018) pp. 625-628.

[8] János, Katona, Parcel analysis for the general application of remote sensing monitoring. In: Orosz, Gábor Tamás (szerk.) AIS 2019 : 14th International Symposium on Applied Informatics and Related Areas organized in the frame of Hungarian Science Festival 2019 by Óbuda University Székesfehérvár, Magyarország : Óbudai Egyetem, (2019) pp. 69-70.

Calibrations of digital aerial cameras

Gabriella BOR*, Gergely LÁSZLÓ*,
* Óbuda University, Alba Regia Technical Faculty, Institute of Geoinformatics, Székesfehérvár, HUNGARY
borgabriella97@gmail.com, laszlo.gergely@amk.uni-obuda.hu

***Abstract*— Aerial photography is one of the most effective sources of information to obtain data and examine the condition of different areas of the earth's surface. Their most important role is in the production of maps, as a fresh source of data is essential for their production. Remote sensing and photogrammetry have become increasingly intertwined to this day. The aim of the paper is to describe the calibration process of digital aerial cameras. Calibration is essential for aerial photography for engineering purposes, as without it we would get inaccurate, error-laden results.**
Three cameras were calibrated, two with an Hasselblad A6D-100c sensor with an R 50 (visible in the visible light range) and one with an IR 50 II lens with an IR (infrared in the range). AgiSoft Metashape software was used to examine the geometric calibration. For radiometric calibration, Hasselblad's Phocus image processing program.
My work is part of a "live" task, as I spend my internship at Envirosense Hungary Kft.

I. Digital photography

Digital photography is performed using digital cameras. The cameras can be divided into two groups, measuring cameras and amateur cameras. Amateur cameras are found in almost every household, although nowadays telephone photography is starting to take their place. Digital measuring cameras can be divided into two groups, ground measuring cameras and aerial measuring cameras. The internal data of the cameras are known. These cameras retain this data permanently, unlike amateur cameras.

But what is internal data? The imaging beam is restored, modeled optically during calibration. The imaging beam is the set of projection rays that create the image. The image is created by central projection on the surface of the digital sensor.

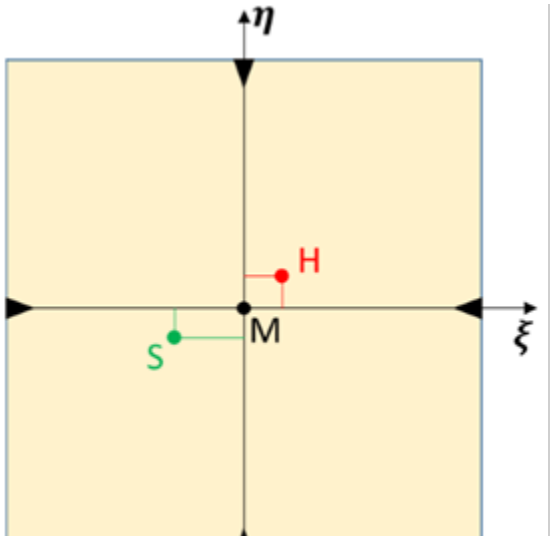


Figure 1. Reference points [1]

H – autocollimation main point
M – center of the image coordinate system
S – center of symmetry
η, ξ - image coordinates of the main point
c_k – camera constant (OH)

These data are determined during the calibration process and can be found in the calibration protocol. If the main point (H) and the center of the image (M) coincide, the camera is aligned. (In the above case, this is not true, apparently does not coincide)

II. Aerial measuring cameras

Parts of aerial measuring cameras [1]:

1. Camera body
2. Negative storage
3. From the cameras

(4.) Other ancillary equipment (GPS, navigation search telescope, overlap control, etc.)

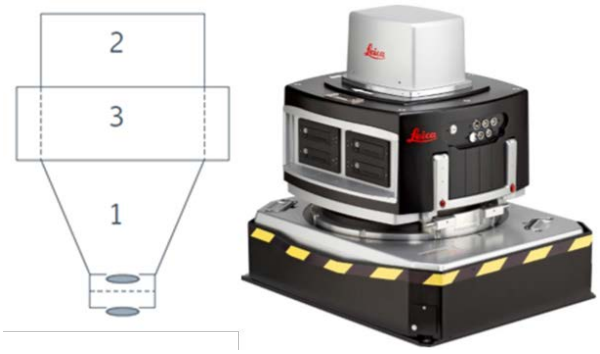


Figure 2. Leica DMC III digital measuring camera [1]

The lens produces the image. The aperture controls how much light it transmits to the sensor. The aperture ring can be used to adjust the gap through which light is allowed in, and to control how much of the incoming light passes through. How long you receive this light is determined by the shutter speed. If light enters through a small aperture, the shutter speed should be increased to allow enough light to reach the negative or the sensor so that the resulting image is not too dark. The settings can be called good if they show a state close to reality. The metering device measures the amount of light on the subject to be photographed. If set to automatic, the appropriate shutter speed and shutter speed will be set. The locking ring is located in the camera body. The aperture affects the depth-of-field range of the image.

Images are formed on a lens system, so in an optical sense, the lens creates the image and this is captured. The point where the camera axis intersects the image plane is called the main point. If the image plane is not perpendicular to the optical axis, the mapping is not ideal. The lens distorts the image. The degree of distortion can be measured. Distortion is the defect of the lens, the distortion resulting from its centering, which increases towards the edge of the image. If light does not arrive on the optical axis, it will be distorted, resulting in a geometric error. The distortion ($\Delta\xi$) is constant over a given radius.

radius (radial) - A_1, A_2, A_3

tangential (perpendicular to the radial) - B_1, B_2

affine - C_1, C_2

Lenses can be grouped according to their viewing angles:

- small $\alpha < 50°$
- medium $50° < \alpha < 60°$
- large $60° < \alpha < 90°$
- very large $90° < \alpha$

The diameter (d) and constant of the objective pupil together are the angle of view (α)

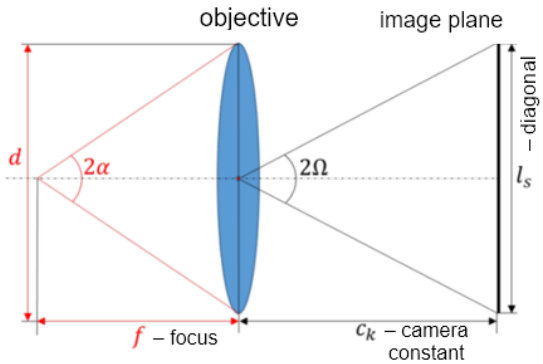


Figure 3. Viewing angles [1]

In the case of the digital image, we speak of surfaces. Digital images are made up of pixels, which are surfaces of extent. We can assign coordinates to a point on the surface. To get these, we need to know the size and extent of the pixels. The pixel coordinate shows how many rows and how many columns a given pixel is in. The higher the number, the finer, more detailed the image. This is characterized by the pixel value. For a camera, how many pixels should be understood as the number of tiny sensor elements located on the surface where the image is captured. In practice, since the number of pixels is in the order of millions, the megapixel unit (mega = 106; million) is used for the number of pixels.

III. Calibrations

The appropriate Hungarian word for calibration is authentication; term related to measuring instruments. In everyday life, scales are calibrated on the market, measuring tapes in geodesy, and all this to get the right, accurate quantity on the measuring instrument. Photogrammetry is associated with the calibration of cameras, measuring cameras and amateur cameras. Camera calibration provides factory data that can be considered authentic, but these are usually checked from time to time to ensure that the nominal data matches the actual data.

One of the purposes of the calibration is to determine the internal data of the camera (η_0, ξ_0, c_k), the other is to eliminate distortion errors by the lens. A camera also requires calibration radiometrically to ensure that the images they take are color-correct.

Calibration can be performed in four ways: In the laboratory, with an control point field, based on asterisks and markers.

A. Camera calibrations in lab

In the laboratory, calibration can be performed in two ways. One option is a rotatable gonimeter, another option is an optical bench rotated at specific angles. For the latter, it is important that the collimators are accurately attached to each other.

In the case of a goniometer, the binoculars (T_1, T_2) must be reset and the glass plate placed in the image plane of the camera, and then the camera is placed in the appropriate position on the turntable, satisfying several aspects. Care should be taken to ensure that the center of the camera's incoming pupil falls on the axis of rotation, this will allow us to detect the crosshairs of the telescope in the mirror image. At the same time, the main point H can be determined.

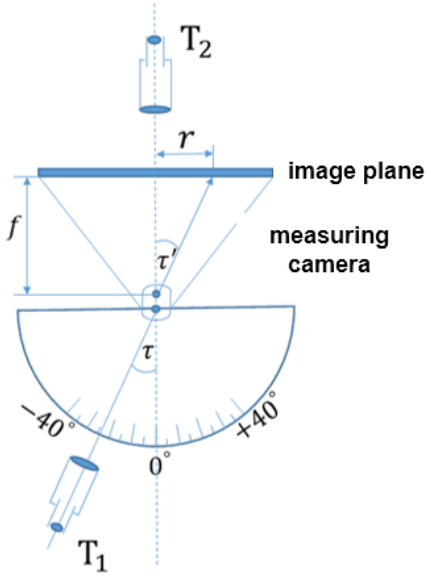


Figure 4. Camera calibration in the laboratory [1]

During the measurement, the τ -angles are measured, so the difference between the angles of the autocollimation main point (H) and the directed grid points is formed. To reduce the errors resulting from the measurement result, we calculate the plotting values and construct their curves, for which first a middle curve and then a smoothing line can be edited. This gives the distortion values.

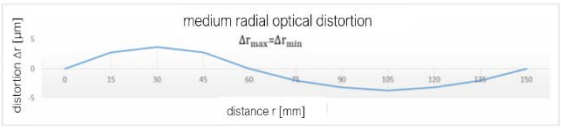


Figure 5. Medium radial optical distortion [1]

B. Camera calibrations with control point field

Field points should be marked, so that they are evenly distributed and have a depth difference of at least twenty percent of the average shooting distances. The coordinates of the control points can be considered known (X_i, Y_i, Z_i)) and are ideal if they can be measured on at least two recordings (ξ_i, η_i)

Relationship between image and object coordinates:

$$\xi = \xi_0 - Ck\frac{r_{11}(x-x_0)+r_{21}(y-y_0)+r_{31}(z-z_0)}{r_{13}(x-x_0)+r_{23}(y-y_0)+r_{33}(z-z_0)} + \Delta x$$

$$\eta = \eta_0 - Ck\frac{r_{12}(x-x_0)+r_{22}(y-y_0)+r_{32}(z-z_0)}{r_{13}(x-x_0)+r_{23}(y-y_0)+r_{33}(z-z_0)} + \Delta y$$

The collination equation is modified by Δx, Δy, considering the optical plot and the affine distortion of the sensor. These corrections can be grouped as radial plot (A_1, A_2, A_3), tangential plot (B_1, B_2), and affine distortion (C_1, C_2) parameters.

Radial distortion:

$\Delta r_{rad} = A_1 \cdot r^3 + A_2 \cdot r^5 + A_3 \cdot r^7$

r: the distance of a given pixel from the main point

$r = \sqrt{(\xi - \xi_0)^2 + (\eta - \eta_0)^2}$

$$\Delta x_{rad} = (\xi - \xi_0) \cdot \frac{\Delta r_{rad}}{r}$$

$$\Delta y_{rad} = (\eta - \eta_0) \cdot \frac{\Delta r_{rad}}{r}$$

Tangential distortion:

$$\Delta x_{tan} = B_1 \cdot (r^2 + 2 \cdot (\xi - \xi_0)^2) + 2 \cdot B_2 \cdot (\xi - \xi_0) \cdot (\eta - \eta_0)$$

$$\Delta y_{tan} = B_2 \cdot (r^2 + 2 \cdot (\eta - \eta_0)^2) + 2 \cdot B_2 \cdot (\xi - \xi_0) \cdot (\eta - \eta_0)$$

Calculating the unknowns: The control point field is created, evenly distributed, which can be measured with a geodetic instrument. The values of η and ξ are derived from the measurement.

1. Installation and measurement of control points → $X_i; Y_i;\ Z_i$

2. Take at least three shots

3. Measurement of alignment points (ξ; η) on the recordings

4. Determining the preliminary value of unknowns

5. Perform equalization with gradual approximation

a) $Adx + l = v$ system of correction equations

b) $A^T A dx + A^T l = A^T v \rightarrow N \cdot dx + n = 0$

c) $dx = -N^{-1} - n$

d) $dx + x^{(0)} = x^{(1)} \rightarrow dx^{(1)} + x^{(1)} = x^{(2)} \leftarrow$ *gradual approximation*

6. Error calculation

$\sigma_0 = \frac{\sqrt{\sum Ni}}{\mathrm{f}}$

$f = n - u$

$\sigma_n = \sigma_0\sqrt{qnn}$

C. Camera calibration based on stars

When we calibrate our camera based on stars, the field's control points are replaced by the stars. For the calibration to be successful, the position of the images relative to the celestial coordinate system must be known and the image coordinates (ξ, η) of the photographed stars must be measured. If image coordinate corrections and camera internal data are used, the relationship between the sky and image coordinate system can be written.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = K \cdot \begin{bmatrix} cos\alpha \cdot cos\delta \\ sin\alpha \cdot cos\delta \\ sin\delta \end{bmatrix} = K \cdot \begin{bmatrix} U \\ V \\ W \end{bmatrix} \qquad R\begin{bmatrix} \xi - \xi_0 - \Delta x \\ \eta - \eta_0 - \Delta y \\ -c_k \end{bmatrix}\begin{bmatrix} U \\ V \\ W \end{bmatrix}$$

$$K = \sqrt{(\xi - \xi_0 - \Delta x)^2 + (\eta - \eta_0 - \Delta y)^2 + c_k^2}$$

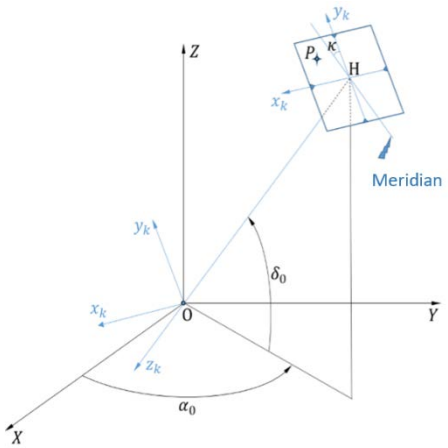


Figure 6. Camera calibration based on stars [1]

ξ, η – image coordinates of points
c_k, ξ_0, η_0 – camera internal data
Δx, Δy – image coordinate corrections
O – projection center
H – main point
α – right ascension
δ – declination angles
K – pixel vector length

$$(\xi - \xi_0) \cdot (1 - A_1 \cdot r^2 - A_2 \cdot r^4 - A_3 \cdot r^6 = -c_k \cdot \frac{r_{11} \cdot U + r_{21} \cdot V + r_{31} \cdot W}{r_{13} \cdot U + r_{23} \cdot V + r_{33} \cdot W}$$

$$(\eta - \eta_0) \cdot (1 - A_1 \cdot r^2 - A_2 \cdot r^4 - A_3 \cdot r^6 = -c_k \cdot \frac{r_{12} \cdot U + r_{22} \cdot V + r_{32} \cdot W}{r_{13} \cdot U + r_{23} \cdot V + r_{33} \cdot W}$$

To be able to calibrate by stars, a minimum of five stars must be measured. Atmospheric refraction significantly affects the recordings - it is important to eliminate this during processing - therefore photography in the zenith direction is recommended because the effect of refraction is small there. As with other different measurements, it is advisable to record the measurement data (e.g. exact time, temperature, air pressure, camera setting data, etc.). Using various astronomical programs, the precession, nutation, motion, daily and annual aberration of the star can be calculated, and then the exact coordinates of these values can be calculated from these values, with the knowledge of which the calibration can be performed.

D. Camera calibration with coded signals

In this calibration procedure, the test diagrams act as control points. Images should be taken with a fixed focus value and calibration should be performed with the same focus value. It is a good idea to take shots from multiple points of view, thus increasing reliability when the points in the test area are in one plane. During the measurement, images were taken from three positions where the cameras were rotated in the plane of the sensor. The cameras we examined consist of a Hasselblad A6D-100c sensor and an HC 50 II lens, medium format cameras. In terms of resolution, 100 megapixels. Its sensor surface is 53.4 × 40.0 mm, where the image is created. In general, for digital cameras, the size of the medium format sensors is 43.8 × 32.9 mm and 53.7 × 40.2 mm. By geometric resolution we mean the size of the smallest image unit captured by a sensor that actually corresponds. This is better the lower its value. This type contains 11600 × 8700 sensors.

Figure 7. HC 50 II lens

They can be classified as medium format cameras that use a film size of 120 or use a digital imaging sensor that produces this size 120. This format produces slightly smaller images than large format films (102x127 mm), but larger than images taken with full-frame sensors or 135 films. There are no standard sizes, they usually vary from manufacturer to manufacturer. Another feature of medium format cameras is that they capture outstandingly high quality images and also provide high accuracy in terms of color accuracy. In addition, they are larger and heavier than most cameras due to the large sensors. Its great advantage is that it is modular, ie its components can be replaced according to your own needs.

Figure 8. Hasselblad A6D-100c

IV. DESCRIPTION OF THE TEST

The measurement was performed on the ground floor of the Institute, in front of the Council Hall. Due to the size of the site and the black and white flooring, it was ideal for the task. Fifty-two control points were selected that covered the test area. 16 markers were glued to the wall opposite the cameras and the rest were placed away from the cameras and placed there. The remaining signs were placed on both sides of the wall, to door openings, and to pegs set up in the middle of the area. We made sure that most of the points were visible from all three positions. Four images were taken from one position, rotating the camera clockwise by 90 ° in the plane, i.e., a total of twelve images were taken with one camera. The points were also measured with a measuring station. The markers used in the study could be downloaded depending on the software.

Figure 9. Markers

RGB and NIR cameras were calibrated in Agisoft. Each image was checked separately. Because we used the program's own punctuation marks, the software was able to recognize them automatically. We checked to see if there were any markers in the image that look good and that they were in the right place. The algorithm was not flawless, where markers did not show the entire surface, there it did not recognize them, and there was one of the markers reflected on the glass that confused it. The partially visible points where the signal center was well identifiable were marked manually. There were times when only after calibration, the results showed (the example below) that a point had been identified, they were in the wrong place. After running the repair and calibration again, the correct values have been reached.

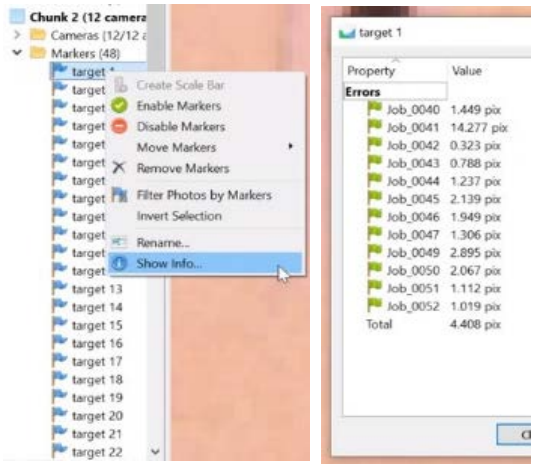


Figure 10. Extent of inconsistencies in pixels per signal and per image

Figure 11. Location of markers

Larger defects were typically at points closer to the camera. They were faulty not only because the distance changed from more distant points, but also because they were not sharply formed, they were blurred because of their proximity.

We obtained the fixed focus, the coordinates of the main image point, the parameters of three radial plots and two tangential plots, the metric size of the sensor surface.)

The following results show the error of the X, Y, Z components, expressed in meters for RGB and IR cameras:

#Label	X	Y	Z	Error_(m)	X_error	Y_error	Z_error
target 1	21,352	-0,177	1,236	0,006549	0,001648	-0,006012	-0,002007
target 2	21,234	-1,387	1,224	0,006018	0,002323	-0,005526	-0,000529
target 3	21,101	-2,778	1,27	0,005879	0,001453	-0,005623	-0,000913
target 4	21,006	-3,855	1,25	0,003525	-0,001949	-0,002865	0,000646
target 5	21,351	-0,218	0,594	0,006276	0,000921	-0,006204	-0,000208
target 6	21,236	-1,411	0,568	0,005094	0,001313	-0,004922	0,000019
target 7	21,107	-2,788	0,552	0,003232	-0,00017	-0,003218	-0,000244
target 8	21,009	-3,882	0,551	0,003647	-0,000183	-0,003632	-0,000275
target 9	21,348	-0,185	-0,452	0,005748	-0,000054	-0,005747	0,000093
target 10	21,227	-1,503	-0,404	0,005869	-0,000832	-0,005769	0,000682
#Total Error				0,007268	0,003156	0,006428	0,001242

#Label	X	Y	Z	Error_(m)	X_error	Y_error	Z_error
target 1	21,352	-0,177	1,236	0,006255	0,001745	-0,00572	-0,00184
target 2	21,234	-1,387	1,224	0,005401	-0,00025	-0,00534	-0,00076
target 3	21,101	-2,778	1,27	0,006107	-0,00075	-0,0059	-0,0014
target 4	21,006	-3,855	1,25	0,005127	0,002141	-0,00452	0,001142
target 5	21,351	-0,218	0,594	0,005949	0,000446	-0,00593	-0,0001
target 6	21,236	-1,411	0,568	0,004887	-0,00112	-0,00476	-0,00017
target 7	21,107	-2,788	0,552	0,00348	0,000017	-0,00347	-0,00031
target 8	21,009	-3,882	0,551	0,00408	0,00005	-0,00408	-0,00018
target 9	21,348	-0,185	-0,452	0,00545	0,000397	-0,00544	-0,000063
target 10	21,227	-1,503	-0,404	0,006174	-0,00263	-0,00556	0,000577
#Total Error				0,006648	0,002556	0,006044	0,001061

Figure 12. RGB, IR camera results

Because of clarity, of the 52 points, I illustrate only the first ten measurements per camera, as these were the points that appeared from almost every position without exception. In the bottom row, the mean error values are to be interpreted for all points.

	Value	Error	F	Cx	Cy	B1	B2	K1	K2	K3	K4	P1	P2
F	10860.8	27.9465	1.00	0.04	-0.21	-0.39	-0.02	-0.99	0.97	-0.94	0.88	0.05	-0.17
Cx	1.67084	5.57294		1.00	-0.02	0.00	0.33	-0.04	0.04	-0.04	0.04	0.98	-0.01
Cy	-83.665	6.25971			1.00	0.32	-0.04	0.20	-0.18	0.17	-0.15	-0.02	0.99
B1	0.288139	1.39263				1.00	-0.01	0.40	-0.38	0.33	-0.27	0.01	0.34
B2	-2.22266	0.533042					1.00	0.02	-0.03	0.03	-0.03	-0.34	-0.04
K1	0.0813215	0.0406388						1.00	-0.99	0.97	0.92	-0.05	0.16
K2	-0.467382	0.241445							1.00	-0.99	0.96	0.05	-0.15
K3	1.44	0.646567								1.00	-0.99	-0.05	0.13
K4	-1.77949	0.663881									1.00	0.05	-0.11
P1	-0.000140808	0.000274598										1.00	0.00
P2	-0.00318663	0.000306481											1.00

Figure 13. Distortion matrix

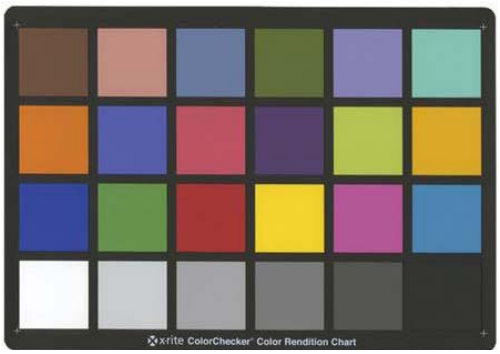


Figure 14. Color calibration table

For radiometric correction, a color calibration table was used, which was imaged at Papkutapuszta Airport. The closer we take pictures of it, the easier it will be for Phocus to recognize the board, but the focal length had to be taken into account.

ΔE avg.: 1.73 max.: 3.94

ΔE avg.: 1.82 max.: 4.47

The recorded images were calibrated in Hasselblad's Phocus image processing program. In the *Color Calibratio*n menu, *Auto Locate* - which automatically recognizes the color scheme - did not work because it did not recognize the target, meaning you had to manually attach the grid to the color calibration board. Then, by clicking on *Calibrate*, he calibrated the image and calculated the ΔE values shown above and as a result we obtained the radiometrically corrected images. The visual difference between ΔE two colors, which can take a value between 0 and 100, in our case the average value below 2 should be interpreted so that the deviation from the reference can only be noticed by a very experienced eye on a monitor calibrated to a better value. Detection of the difference below 1 is only possible with a target instrument test. In the

description of the figures, the maximum value is around 4, which is already a visible difference for some colors. Strangely, the biggest difference is in black in both cases. [2]

V. Sensor types

A. CCD and CCD

CCD and CMOS are the best known digital image capture modes. On the surface of these sensors, the photosensitive elements are arranged in a linear or matrix manner. The CCD (Charge Coupled Device) is considered the first digital camera. The chip transmits the incoming light in front of the lens as a digital signal to another device that can be read and processed. The result will be an image of near-perfect quality, with minimal distortion, even at low light requirements. One of the disadvantages is that the manufacturing cost of the CCD chip is very high, in contrast to CMOS. The other downside is that it gets very hot due to high energy consumption, so the thermal noise will also be higher. CMOS (Complementary Metal-Oxide-Semiconductor) is considered the "little brother" of CCD. It uses a photoelectric effect in the same way as its "big brother". The camera uses less power to extend its life. It is more sensitive to electromagnetic interference than the CCD.

Where should the image capture happen? Obviously where my optically produced image is sharp. The optical law of sharp imaging determines where the image is sharp. A sharp image is obtained if the reciprocal of the focal length (f) is equal to the sum of the reciprocal of the subject distance (t) and the reciprocal of the image distance (k).

$$\frac{1}{f} = \frac{1}{t} + \frac{1}{k}$$

B. Multispectral cameras

Digital multispectral satellite remote sensing systems are based on multispectral cameras. Multispectral sensors are those that are able to capture electromagnetic energy in several wavelength ranges. It is possible to display the image for data recorded in different spectral ranges. The simultaneous visual representation of combinations of these data is called composite. As a result, the recordings can be studied thoroughly, as data can be discovered on them that the human eye alone cannot perceive.

C. LIDAR

LIDAR (Light Detection and Ranging), which can be classified as active systems, measures reflected electromagnetic energy in the visible, infrared and ultraviolet ranges. The recorded data does not depend on the humidity and shade of the air, so it does not depend on the weather or the time of day.

D. RGB sensors

Sensors that detect three ranges of visible light and have different spectral sensitivities are located next to each other. The spectral ranges of these sensors are as follows:

R (red)	650 nm – 750 nm
G (green)	490 nm – 575 nm
B (blue)	420 nm – 490 nm

We see these colors in the pictures the same way we see them with the naked eye.

E. NIR sensors

The infrared range can be divided into three parts, near, medium and far infrared. Our camera is able to capture images in the near-infrared wavelength range from 760 nm to 900 nm, on which field objects appear in false colors.

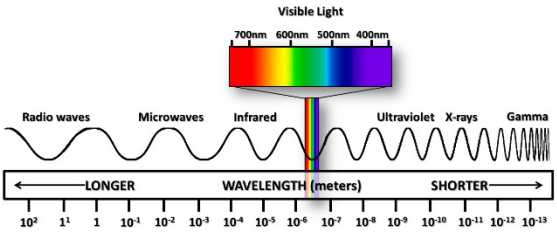


Figure 15. Electromagnetic spectrum ranges [3]

VI. Softwares which was used for calibration

Agisoft

Agisoft is one of the best known point cloud software. The program automatically detects the points in the test field, then measures and assigns the numbers. This is possible because, according to a geometric arrangement, the markers are unique, carry code in themselves, or because the coded signals complement the grid or markers of the test field and recognize their orientation. [5]

Phocus

Hasselblad's image processing software is Phocus, which works with all Hasselblad cameras, but can of course be used for photos taken with any camera. [6]

VII. DIGITAL ORTHOPHOTO PRODUCTION

The cameras examined in my dissertation are used to produce digital orthophotos. Orthophoto is a special photo transformation that is free from distortion. There are two types of distortion. One is perspective distortion due to the fact that the image plane is not parallel to the object plane. The other is altitude distortion, which is caused by elevation differences in the terrain.

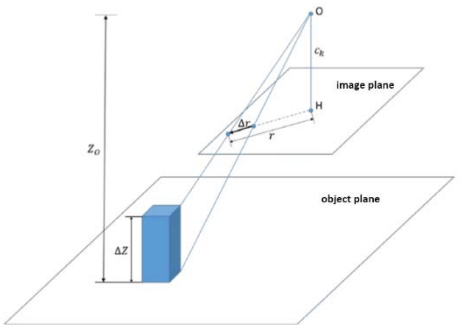


Figure 16. Δr distortion [1]

Δz: that can be extracted from the topography model

r: distance of point from nadir

$$\Delta r = \Delta z \cdot \frac{r}{C_k \cdot M_k} \qquad M_k = \frac{h_k}{C_k}$$

Two types of digital elevation models (DMM) are known. The digital terrain model (DDM) and the digital terrain model / surface model (DTM / DFM0).

If only topography is depicted, then we are talking about a traditional orthophoto, if "everything is in place" (buildings, vegetation, etc.), then we are talking about a real orthophoto.

Before taking an orthophoto, we need to know how accurate the terrain model is.

For the terrain model, how inaccurate it can be, i.e. the allowable error, can be calculated as follows:

$$mz_{max} = \frac{\Delta r \cdot C_{K \cdot m_0}}{r}$$

Orthophoto production process:

Step 1: Data Preparation

- original image: in digital or analogue form and scanning
- image orientation
  elements of internal information (c_k, ξ_0, η_0)
  elements of external information $(X_0, Y_0, Z_0, \varphi, \omega, \kappa)$
- digital elevation model: acquisition or production
  if we produce: based on contour maps contour lines; by field measurement; aerial laser scanning (LIDAR); possible by photogrammetry, which requires a pair of stereo images.

Step 2: Select an orthophoto area

- enter corner coordinates (usually EOV);
- we adjust the orthophoto to the resolution of our original image

Rule 1: The resolution of the orthophoto is always better than the original image

Rule 2: Up to twice as good

Step 3: The required values must be assigned to the orthophoto pixels

$$0 \leq g_i \leq 255$$

Basic equation of central projection:

$$\xi = -C_k \frac{r_{11}(x - x_0) + r_{21}(y - y_0) + r_{31}(z - z_0)}{r_{13}(x - x_0) + r_{23}(y - y_0) + r_{33}(z - z_0)}$$

$$\eta = -C_k \frac{r_{12}(x - x_0) + r_{22}(y - y_0) + r_{32}(z - z_0)}{r_{13}(x - x_0) + r_{23}(y - y_0) + r_{33}(z - z_0)}$$

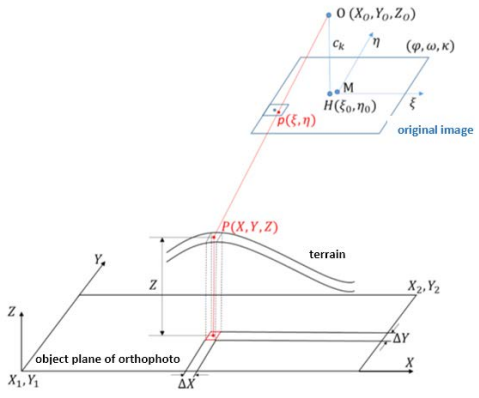


Figure 17. Digital orthophoto production [1]

1. The principle of immediate neighborhood
2. Bilinear interpretation
(4g value eg 1 | weighted average)

$$g' = \frac{\sum g_i \cdot s_i}{\sum s_i}$$

3. Tertiary collination
   - For 4 × 4 pixel cubes
   - The contrast of the image decreases in the orthophoto, ie the image "becomes lifeless", therefore it is recommended to contrast enhancement

4. Making an orthophoto mosaic
   - Marking the cutting edges and joining them along it
   - Several orthophotos cover an area
   - Rule 1: along lines (road, rail, land border)
   - Rule 2: Lines should be cut perpendicularly
   - color balance is required for matching

5. Orthophoto map alignment
   - Scale, projection, coordinate, mapping writings (north direction, coupon number, scale, etc. writing)
   - Directly suitable for mapping

VIII. Summary

The purpose of this article was to describe the calibration process of digital aerial cameras. We calibrated three cameras, two RGB (visible in the visible light range) and one IR (infrared in the range) Hasselblad A6D-100c sensor, equipped with an HC 50 II lens, which is owned by Envirosense Hungary Kft.

Calibration of the cameras is essential to produce a high-precision end product.

The test shows why calibration should be performed in all cases. In theory, the cameras we examined should arrive at the customer with appropriate calibration values for both the goods and their category, however, the distortion diagrams show that the distortion is significant even when moving towards the edges of the lenses. Which might not be a big problem with ground-based photogrammetry, but for aerial photography at 2,000 meters above the surface, the errors can be so significant that it makes it difficult to perform aerial triangulation of flights designed primarily with less overlap. This is because the software we use (Trimble OrthoVista) is a specifically rigid-wing mainframe, so it expects high-precision data as input, so if the distortion at the image edges exceeds a certain level, the automatic alignment search algorithm will not find a match within those error limits. points, so the number of areas with multiple overlaps will also be significantly reduced, which will also negatively affect ultimate accuracy. In addition, there may be a problem in making orthomosaics that while the accuracy on the inside of the images is good, after constructing the boundaries, the final product does not appear to have been composed of different images; already clearly shows that the geometric accuracy is not uniform within the images.

In addition to the numerical accuracy, the real, visible results of the calibrations presented in the dissertation will be revealed during the flights to be performed in the near future.

In the future I will also cover these results, and I will try to further improve the geometric accuracy with the help of BINGO software and possibly other calibration procedures.

IX. References

[1] Dr. habil Jancsó Tamás: Digitális fotogrammetria jegyzet (2017)

[2] Delta E 101 by Zachary Schuessler http://zschuessler.github.io/DeltaE/learn/?fbclid=IwAR1WxtC7JeEb72rV0Db1Qpz6o-hp2uNAqMlKGzOWDe7Sc2sbeubeORNX3zQ

[3] Verőné Wojtaszek Malgorzata: Fotointerpretáció és távérzékelés: A távérzékelés fizikai alapjai (2011)

[4] Bácsatyai László: Magyarországi vetületek (2005)

[5] Agisoft Metashape Manual: https://www.agisoft.com/pdf/metashape-pro_1_5_en.pdf

[6] Phocus by Hasselblad Manual http://static.hasselblad.com/2014/12/Phocus-User-Manual_v14_-ENG_2016.pdf

[7] Katona, Janos ; Gulyas, Margit Horoszne Determination of factors modifying land value based on spatial data
In: Drótos, Dániel; Vásárhelyi, József; Czap, László; Ivo, Petráš (szerk.)
Proceedings of the 19th International Carpathian Control Conference (ICCC 2018)
Piscataway (NJ), Amerikai Egyesült Államok : IEEE, (2018) pp. 625-628.

[8] Alhusain, Othman ; Tóth, Zoltán ; Rakusz, Ádám ; Almási, László ; Farkas, Bertalan Vision-based System for Quality Control of Some Food Products ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING XXXV pp. 477-482. , 6 p. (2004) Nyilvános idézők összesen: 3 Független: 3 Függő: 0

[9] Barsi, Árpád ; Fi, István ; Mélykúti, Gábor ; Lovas, Tamás ; Tóth, Zoltán Úthibák detektálása - Mobil felmérő rendszer fejlesztése a BME-n MÉLYÉPÍTŐ TÜKÖRKÉP: A SZAKMA LAPJA

2005 : 2 pp. 32-33. , 2 p. (2005) Matarka Zárolt Nyilvános idézők összesen: 3 Független: 0 Függő: 3

[10] Tóth, Zoltán ; Mélykúti, Gábor ; Barsi, Árpád Digitális videokamera kalibrációja GEOMATIKAI KÖZLEMÉNYEK / PUBLICATIONS IN GEOMATICS 8 pp. 297-302. , 6 p. (2005) Folyóiratcikk/Szakcikk (Folyóiratcikk)/Tudományos

Simulation of Epidemic Dynamics Using a Modified SEIRS Model

Árpád Elekes
0000-0002-5878-6566
Budapest, Hungary
elekes.a@freemail.hu

***Abstract*—A modified SEIRS epidemic model is introduced and then used to simulate some scenarios, such as possible recurrence of the disease and the impact of restrictive control measures on the dynamics. Also spatial spreading investigated among neighboring regions.**

***Keywords*—epidemic modeling, epidemic simulation, compartment models, SEIRS model**

I. Introduction

This work was motivated obviously by the recent Covid-19 pandemic (and an accidental encounter with an introduction [1],[2]). Initially the main questions of interest were the conditions for a possible recurrence and the spatial spreading of the disease among neighboring interconnected regions. It turned out however, that the model parameters presently cannot reliably be assessed to address these issues realistically enough. This is partly the reason why the treatise here has become more general. So, what follows is a straightforwardly extended SEIRS (still rather simplified) model of a contagious disease and the investigation of some interesting scenarios (so not a full extent examination) based on the simulation of this model.

The most common quantitative method of describing epidemic dynamics is the use of compartment models, represented by a system of differential equations. Here the SEIRS model from this family was choosen as an initial model to extend, and solved numerically for different initial conditions and parameter settings.

II. THE SEIRS BASE MODEL

A. *Features of the SEIRS model without vital dynamics*

- Compartments of the model: susceptible *S*, exposed *E*, infectious *I*, recovered *R*.
- The parameters: transmission rate β, virulence rate σ, recovery rate *γ*, resusceptibility rate *ρ*.
- The progression in the model: a susceptible individual by contacting the disease can be infected, first not contagious yet (in the exposed group), then after the incubation period can be infectious, and even later can recover. If the immunity only temporary, re-susceptibility is possible (Fig. 1.).

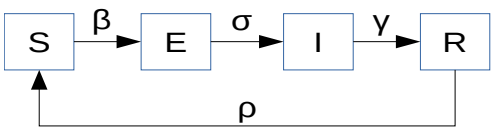


Fig.1.The SEIRS model.

B. *The governing equations of the system*

$$\dot{S} = -\beta SI + \rho R \quad (1)$$

$$\dot{E} = \beta SI - \sigma E \quad (2)$$

$$\dot{I} = \sigma E - \gamma I \quad (3)$$

$$\dot{R} = \gamma I - \rho R \quad (4)$$

III. THE PROPOSED EXTENDED SEIRS MODEL

A. *The additional features and differences of the new model over the base SEIRS model are:*

- In order to be able to take into account the mortality from the disease a new *D* compartment is introduced for the deceased part of the population (Fig. 2.).

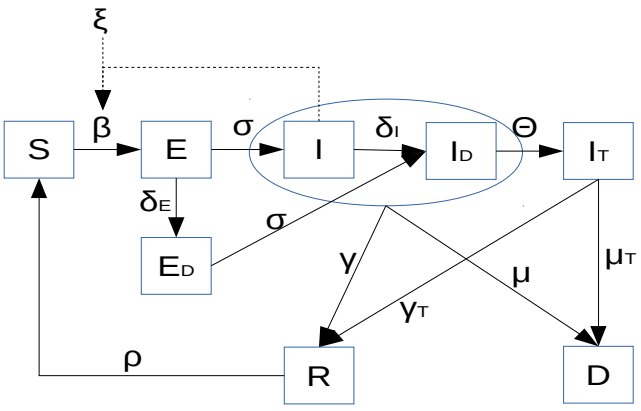


Fig. 2. The SEIRSDx model.

- Counter measures (i.e. quarantine, hygiene, social distancing...) conditional on some signal (i.e. the number of fatalities or new cases rises above a limit) have a decreasing effect on the transmission rate: $0 \le \xi \le 1$ in (5).
- Here the original SEIRS infectious group separated for three distinct compartments: the non-diagnosed *I*, the diagnosed and non-treated I_D , and the diagnosed and medically treated I_T . The treatment is supposed to be beneficial on the recovery, that is why the *γ* recovery and *μ* mortality rates of the treated and not treated class can be different.
- A susceptible individual by contacting the disease and becoming infected, first not yet infectious (in the

exposed group), then after the incubation period (can be zero) can be virulent. If diagnostication possible after the undiagnosability period then according to the diagnostic capacity rate the infected can be diagnosed and in this case, automatically separated from the rest of the population. The diagnosed then, depending on the actual health service capacity, can be medically treated, and depending on the recovery rate and mortality rate later can recover or decease. If the immunity only temporary, a recovered person can become susceptib again depending on the re-susceptibility rate.

- The diagnostic capacity rate δ (number of diagnosis/ time) is assumed to be constant (11).
- The actual medical treatment capacity $\theta(t)$ is the number of new cases could be provided with medical care at time t, which is the available capacity remaining from an initial (θ_{MAX} constant) value diminished by the already treated group $I_T(t)$ (12).
- The infectious groups and the recovered group R stratified by delay (from a reference) time (while the S, D groups are not). This allows treating the transitions between these groups (delay) time dependent, not just constants: recovery rate γ_{NT}, mortality rate μ_{NT}, treated recovery rate γ_T, treated mortality rate μ_T, resusceptibility rate ρ are all can be functions of time (starting from the joining time to the given compartment).
- In the simplest (constant) case (which is used here in the later case studies) it is possible to set the starting point in time of the:
  - virulence (latency, incubation period),
  - diagnosability,
  - recovery,
  - mortality,
  - resusceptibility.

B. The major equations

$$\dot{S} = -\beta\xi S I + \rho R \tag{5}$$

$$\dot{E} = \beta\xi S I - \sigma E - \delta_E \tag{6}$$

$$\dot{E}_D = \delta_E - \sigma E_{ND} \tag{7}$$

$$\dot{I} = \sigma E - \delta_I - (\gamma+\mu) I \tag{8}$$

$$\dot{I}_D = \sigma E_{ND} + \delta_I - (\gamma+\mu) I_D - \theta \tag{9}$$

$$\dot{I}_T = \theta - (\gamma_T+\mu_T) I_T \tag{10}$$

$$\delta = \text{const} = \delta_E + \delta_I \tag{11}$$

$$\theta = \theta_{MAX} - I_T \tag{12}$$

$$\dot{R} = \gamma(I+I_D) + \gamma_T I_T - \rho R \tag{13}$$

$$\dot{D} = \mu(I+I_D) + \mu_T I_T \tag{14}$$

$$\delta_E = \delta \frac{E}{(S+E+I)} \tag{15}$$

$$\delta_I = \delta \frac{I}{(S+E+I)} \tag{16}$$

C. What is still missing

What is not considered here, but for some disease could be an important factor too:

- vaccination
- vital dynamics (demography)
- age differences
- gender differences
- no distinction is made between the $E \rightarrow E_D$ and $I \rightarrow I_D$ transitions (14),(15), but obviously should. (symptomatic vs. asymptomatic case) $\delta = \delta_{Testing}$ (=const) + no. of symptomatic cases (for diseases which have fully diagnostic symptomes)

IV. SOME RESULTS

A. Epidemic recurrence

One way of how cyclic disease wave can develop is when the immunity is not final, so after some time the members of the recovered population can become susceptible again.

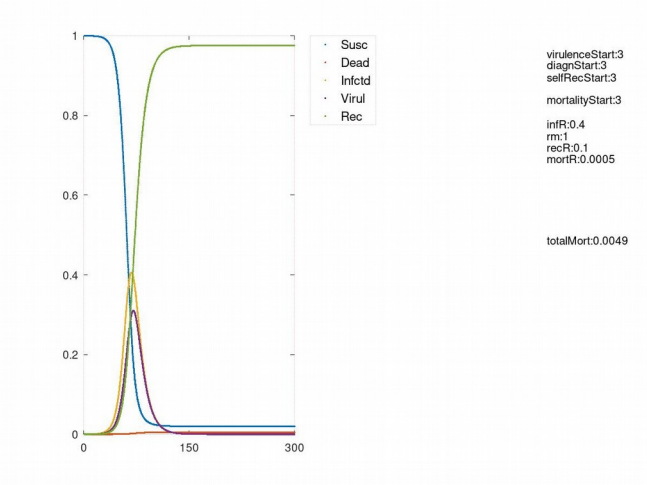


Fig.3. A possible scenario without recurrence

An epidemic second and third wave can develop also when the restriction measures introduced and applied alternately, even without re-susceptibility Fig.8. .

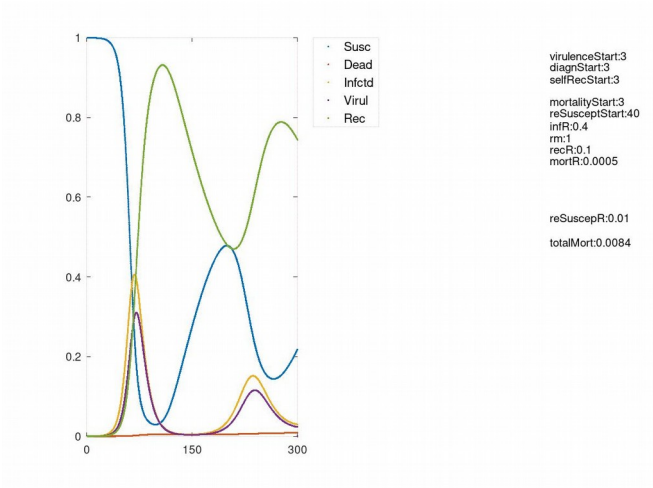


Fig.4. Recurrence scenario with the same settings except now recurrence starts after 40 days with re-susceptibility rate 1%.

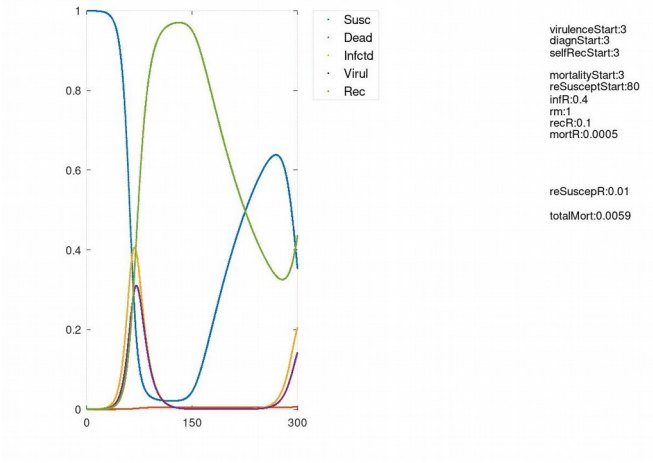


Fig.5. Recurrence scenario with the same settings except now recurrence starts after 80 days.

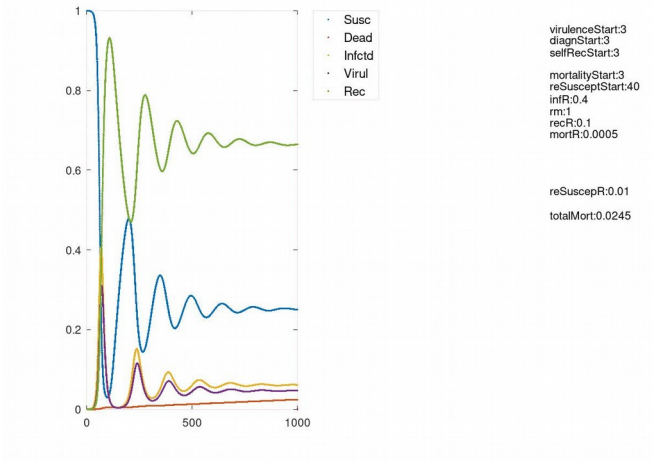


Fig.6. The longer run of Fig.4. scenario.

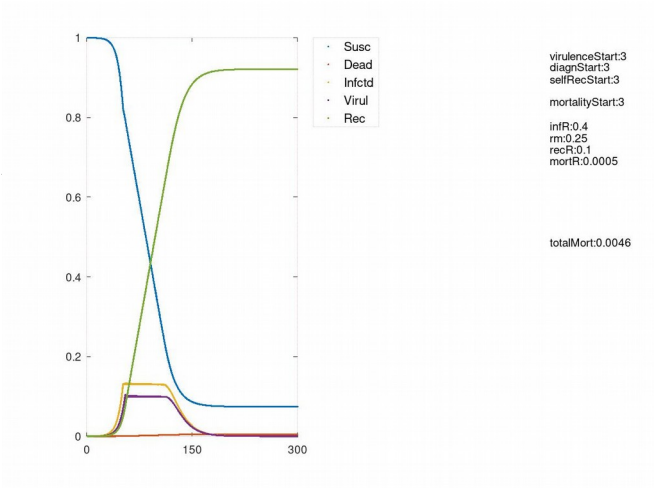


Fig.7. The scenario of Fig.3. with control measure effect $\xi = 0.25$. The restriction is applied is when the number of infected cases exceeded the 15% of the non-infected population.

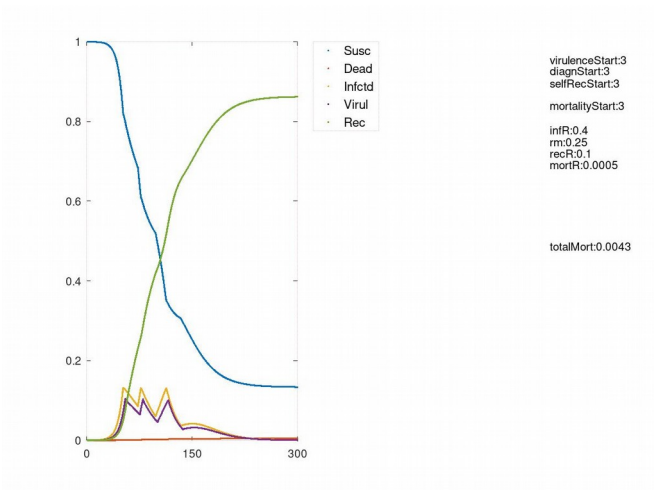


Fig.8. Recurrence scenario of Fig.4. settings (still without re-susceptibility) but here the restriction last for a prefixed 3 weeks (the restriction starting signal is the same).

B. Spatial spreading of an epidemic

One of the basic assumptions of compartment models is the homogeneous mixing of the population, which is not realistic over the entire course of an epidemic and on the larger territories considered. That is why it can be interesting to divide a major territory to subareas with possibly different population densities and even maybe some parameters Fig.9. Here some population exchange (only the non-diagnosed, so non-quarantined) is allowed between connected areas.

Also the parameters of real life epidemics have a rather random nature, so it can be interesting to include this feature in the compartment models too (Fig.10.).

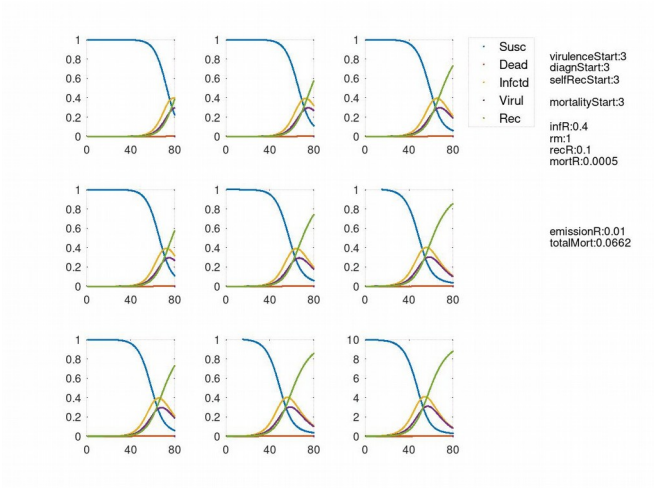


Fig.9. Epidemic spreading on a grid of 3x3 areas, starting from the bottom right.

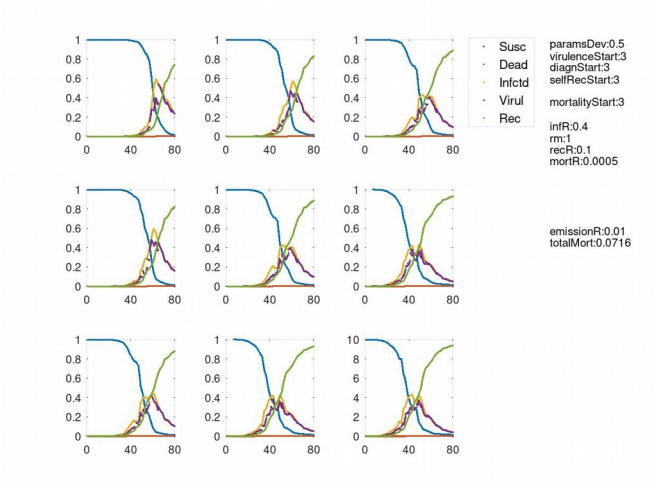


Fig.10. A more realistic looking Fig.9. case with parameters randomly changing (normal distribution, with standard deviation 0.5).

CONCLUSION

The difficulty of prediction of course lies not in the math but in the biology, namely: how to interpret the highly random real life statistics to obtain or measure the relevant parameters of the models.

REFERENCES

[1] Trefor Bazett, “The MATH of Epidemics | Intro to the SIR Model”, https://www.youtube.com/watch?v=Qrp40ck3WpI

[2] Trefor Bazett, “The MATH of Epidemics | Variants of the SIR Model” https://www.youtube.com/watch?v=f1a8JYAixXU

WTO's Environmental Goods: Technical and Legal Scope

Dr. Ezgi Ediboğlu
Department of Inernational Relations
Adana Alparslan Türkeş Science and Technology University
Adana, Turkey
ezgiediboglu@gmail.com

***Abstract*—The World Trade Organization introduced negotiations for an Environmental Goods Agreement. The aim is to reduce tariffs for "important" environmental-related products. There are however contentious debates on what environmental goods are and what makes some of those goods important. Adding the relevance of international environmental law on environmental goods, the issue of defining WTO's environmental goods becomes even more challenging. In this paper, the author aims to analyse these discussions, and set the boundaries for the scope of environmental goods.**

***Keywords*—environmental goods, the WTO draft agreement on environmental goods, international environmental law, international trade law**

I. INTRODUCTION

After a decade-long discussions under a committee of the World Trade Organization (WTO), 46 WTO member states started the negotiations for adopting an Environmental Goods Agreement (EGA) in 2014 [1]. The EGA aims "to eliminate tariffs on a number of important environment-related products" [2]. The expected outcome of this aim is to increase the use and investment potential of such goods. As a result, the EGA can provide a win-win situation for both international trade and environment regimes.

Understanding the term "environmental goods" of the WTO is crucial as the mitigation of, and adaptation to, transboundary environmental problems partially relies on using goods that are environmentally sound [3]. If the EGA comes into force, it is estimated that green-house gas concentration would decrease with the more use of environmental goods [4]. Likewise, a study on the effect of the EGA coming into force in the United States claims that "greater usage of energy-efficient bulbs is estimated to save 238 million kilowatt hours in the United States each year. These savings correspond to 124,000 tons of coal each year" [5]. In other words, the EGA could support international environmental regimes -such as the international climate change regime- to achieve their environmental goals [6].

The negotiating parties of the EGA so far convened 18 times to discuss environmental goods but there is still no clear guidance regarding the term environmental goods. One argues that negotiating parties deliberately avoid a definition [7]. The negotiating parties use a listing method which means that a list is prepared that only includes environmental goods that are considered important by the negotiating parties. However, there is no agreement on what important environmental goods are. Contentious debates were present on this point which resulted in the cessation of meetings since December 2016.

This article aims to set the boundaries for the scope of environmental goods. For this purpose, the negotiating parties' criteria to choose environmental goods and to list the important ones will be discussed. Subsequently, an international environmental law concept of environmentally sound technologies will be analysed. Environmentally sound technologies are considered critical for the mitigation of human-caused environmental problems and for the increase of the resilience of human beings to the pressing environmental issues [8]. It will be questioned whether the concept of environmentally sound technologies could shed light on the scope of environmental goods.

This article will conclude that environmental goods do not constitute a clearly definable concept. The author argues that it is rather a list of goods that are identified by the EGA parties. As a result, if the EGA comes into force, environmental goods would be goods that are exempted from WTO tariffs for a period of time until new products are identified as environmental goods. The author suggests that the concept of environmentally sound technologies could provide guidance on the negotiating parties of the EGA to identify important environmental goods.

II. ENVIRONMENTAL GOODS WITHIN THE EGA

A. Understanding the Listing Method and Current Environmental Goods

The trading system of the WTO is complex. The analysis of the list of environmental goods within the EGA therefore requires an introduction to the WTO.

The WTO oversees the implementation of international trade agreements that guide the trade policies of the vast majority of countries [9]. One of the main principles of the WTO is non-discrimination between products falling under the same product group to give equal opportunity to similar products [10]. For deciding what similar products are, the WTO uses the Harmonized System of the World Customs Organization [11]. The Harmonized System broadly categorises products and the same tariffs apply to the products falling under the same category. Adding that the Harmonized System influences 98 percent of world trade and has over 5,000 product groups, these categories are crucial for international trade [12].

For instance, "machinery and mechanical appliances" is one category given in the Chapter 84 of the Harmonised System. Sub-categories are given to better define each broad category while sub-sub-categories are not allowed internationally. As an example, "milking machines and dairy machinery" is a sub-category meaning that all machinery under this category are subjected to same tariffs. As

environmental goods potentially encompass a range of technologies with vastly different applications, they are not feasible as a category to the Harmonized System. As a result of this situation, the negotiating parties to the EGA adopted a listing method that aims to eliminate tariffs on certain product groups of the Harmonized System that are listed within the EGA.

This listing method was adopted from a regional economic forum, the Asia-Pacific Economic Cooperation which listed 54 product groups of the Harmonised System for tariff cuts in 2012 [13]. The negotiating parties of the EGA adopted the list of the Asia-Pacific Economic Cooperation with the aim of expanding it [14]. During the negotiations, there were times where 650 product groups were identified but the final draft of the EGA had 300 product groups [15]. It is important to emphasise that these are not products themselves but product groups. These 300 product groups are related to the following 10 categories: air pollution control, solid and hazardous waste management, energy efficiency, wastewater management and water treatment, environmental remediation and clean-up, environmental monitoring analysis and assessment, cleaner and renewable energy, noise and vibration abatement, resource efficiency, and environmentally preferable products [16].

After clarifying this listing method, the issues of this method are discussed below.

B. Issues of the Listing Method

During the negotiations, the negotiating parties of the EGA submitted their lists containing product groups that they consider as environmental goods [17]. Every country and the European Union -as a party to the WTO and a negotiating party to the EGA- individually decided what environmental goods are and which of those goods are sufficiently important to be exempted from WTO tariffs. It is not clear how and with which criteria and method the "importance" of environmental goods is defined. A study points out that the evaluation of a product as environmental depends on socio-economic conditions and the perception of the actor [18]. It is hence rather unsurprising that there were differences between the choices of negotiating parties.

Relying on the submissions of the negotiating parties, draft lists of product groups were decided. Contentious debates emerged on each draft of the EGA list. Some claimed that there are product groups that are not environmental and unacceptable, and some alleged that key product groups are missing from the list [19]. Despite the presence of technical experts in the negotiations, a consensus could not be reached for an agreed list of product groups. Leaving the scope of environmental goods solely to parties leaves a great discretion on parties and it becomes nearly impossible to arrive at a joint list. Adding that the list requires updates for keeping up with the technological changes, each updating process would be subjected to the similar discussions amongst parties [20]. It is not clear how the WTO aims to address this problem.

Another issue regarding the listing method is that it does not correspond with the aim of the EGA to allow more trading and the use of goods that are environmentally essential. Indeed, even though the EGA would promote the trade of listed product groups, it would disadvantage others which may be key for environmental protection. For example, a study investigating at the coverage of the final draft list of the EGA reveals that some solar panel technologies are covered by the EGA list and some are not [21].

Another issue that shows that the listing method does not correspond to the aim of the EGA is that the negotiating parties already have very low tariff rates for the listed product groups in the EGA draft list [22]. All EGA negotiating parties are either developed or major developing countries and these countries do not impose high tariffs for product groups of the EGA [23]. For example, a study analysing the potential effect of a draft list of the EGA on greenhouse gas emissions amongst its parties shows that greenhouse gas emissions would decrease between 0.1 and 0.9 percent of 2030 total greenhouse gas emissions [24]. A much higher percentage of decrease on emissions is needed for the EGA to contribute environmental protection [25].

High tariffs are in place for EGA product groups mainly in developing and least-developed countries. Luring developing and least-developed countries into the EGA is however challenging. Most high technologies are produced by EGA negotiating parties [26]. The remaining countries are hence wary that the EGA would mainly benefit the EGA negotiating parties [27]. As a result of this situation, the EGA does not reach its main target countries that have high tariff rates on product groups listed in EGA. In other words, the EGA does not reach those countries for which its tariff elimination would be most effective.

Due to these issues, there is no further date for EGA negotiations. Nevertheless, trading goods that are playing a role in environmental protection is essential and many multilateral environmental agreements have been promoting and in some cases have given obligations on their parties for this purpose [28]. The support of the international trading regime by cutting tariffs would be a step forward in this regard. Thus, the author questions whether disagreements would be minimised and more countries would join the EGA negotiations in case that the scope of environmental goods are better defined. For this purpose, the following section looks at a better developed international environmental concept of environmentally sound technologies for comparing and finding criteria for the term environmental goods.

III. An International Environmental Law Approch to Environmental Goods

A. The Relevence of International Environmental Law

Since the Declaration of the United Nations Conference on the Human Environment in 1972, international environmental law puts an importance on environment related technologies as they are considered to play a decisive role for combatting, and for the adaptation to the adverse effects of, environmental issues [29].

The term technology refers both to hardware components such as tangible products, and to software components such as know-how or computer programmes. Furthermore, there is the component of orgware which is about the organisation of the technology [30]. For instance, when a country is implementing a capacity building project for improving its local technological conditions, this capacity building is considered to be about the organisation of the technology and hence is counted as technological support [31]. The concept of technology under international environmental law is restricted in relation to its purpose such the protection of the ozone-layer [32]. In any case, however, the technology in

question must be broadly related to environmental protection and to the adaptation to the adverse effects of environmental issues.

As mentioned before, many multilateral environmental agreements have been promoting and in some cases have given obligations on their parties for the use of technologies for environmental purposes. This is where the WTO becomes relevant. Indeed, the access of countries to these technologies requires their transfers between nations in compliance with the international trade rules of the WTO. However, as explained, the WTO principles and the Harmonised System do not allow technologies to be differentiated according to their benefit to the environment. To address this problem, the WTO started the negotiation for the EGA which is experiencing problems with identifying environmental goods. Understanding the scope of technologies referred under international environmental law could be useful for clarifying the WTO term of environmental goods. This could help the negotiating parties of the EGA to agree on a list of product groups.

There has been a host of different references to technologies under international environmental law such as environmental technologies, green technologies, clean technologies, environmentally friendly technologies, and environmentally sound technologies. Even though their scope changes according to the context where they used, the common aspect of all these technologies is that they play a role in environmental protection and the adaptation to environmental issues [33]. This article prefers the term environmentally sound technologies as it became a widely used concept. The section below discusses environmentally sound technologies to see whether they could illustrate some criteria for identifying environmental goods.

B. The Comparison of Environmentally Sound Technologies and Environmental Goods

Environmentally sound technologies are defined by their function. They "protect the environment, are less polluting, use all resources in a more sustainable manner, recycle more of their wastes and products, and handle residual wastes in a more acceptable manner than the technologies for which they were substitutes" [34].

This definition gives an overall idea but does not provide clear criteria on how to decide whether a technology is environmentally sound. For instance, energy efficient technologies are emitting less greenhouse gases than older technology substitutes, but their energy efficiency varies drastically. It is hard to define what level of energy efficiency makes them environmentally sound. This situation seems to resemble the issue of defining the "environmental" aspect of the term for environmental goods. Nevertheless, the author argues that there are essential differences between them.

The concept of environmentally sound technologies is a very broad concept that in practice refers to any kind of hardware, software, and orgware that provides any kind of improvement for the environment [35]. A look at the diversity of projects on environmentally sound technologies illustrates the vast range of activities. There were projects for transferring renewable energy technologies, for overcoming public opposition to a technology, or for training engineers to develop their skills on a new technology [36]. When a project is undertaken for the transfer of environmentally sound technologies, it is hard to claim that this project is not related to environmentally sound technologies. This is because the term environmentally sound technologies is a broad concept that includes a normative value of environmental protection [37].

On the other hand, the WTO term "environmental goods" is not necessarily incorporating environmental protection as a normative value. For example, according to the European Union's understanding of environmental goods, "most of the goods in the WTO list are not 'environmental' goods" [38]. Whereas the concept of environmentally sound technologies has an inclusive quality in that it includes technologies according to a normative standard, environmental goods are limited to the identified product groups.

Furthermore, since there will be an economic advantage for listed product groups due to tariff cuts, it is possible that proposals may partially focus on economic rather than environmental considerations. For instance, a study observed that "half of the environmental goods likely to be in use within the coming decade do not currently exist" [39]. This observation brings about the question whether new products are aimed to be exempted from tariffs for their potential environmental benefits or for their value in trade.

In this light, the "environmental" aspect of the WTO's term of environmental goods is mainly a selection criterion for the EGA list, the content of which is left to the discretion of the negotiating parties. It is possible that the final EGA list will not include crucial environmentally sound technologies and that these technologies will thus be economically disadvantaged. Even though it is realistically not possible to create a list in which all product groups are considered equally important for the environment, it is still imperative that the list takes environmental concerns into account.

Having said this, it becomes apparent that the concept of environmentally sound technologies and the term environmental goods are not compatible in their scope. Extracting criteria from the concept of environmentally sound technologies for defining environmental goods is hence not feasible. However, international environmental regimes conducting studies or projects on environmentally sound technologies could provide guidance for the negotiating parties of the EGA to select important environmental goods. Indeed, project experiences with environmentally sound technologies would provide the negotiating parties of the EGA with crucial data and can be very influential for the decision of which environmental goods are considered "important". For example, institutions of the climate change regime and the United Nations Environment Programme are respected institutions in this regard.

At the same time, since the EGA has the potential to influence international environmental regimes by contributing to an increased transfer of environmentally sound technologies, more communication amongst the institutions of international environmental regimes and the negotiating parties of the EGA is needed. There is however no initiation of the WTO in this regard.

The last section below will revisit the main discussions and will conclude the analysis on environmental goods.

IV. CONCLUSION

International trading principles and rules have been evolving since the 1940s whereas international environmental laws have developed after the 1970s [40]. In this regard it was not surprising to see that WTO trading principles and the

system it uses for the categorisation of goods did not, at least not initially, provide the necessary flexibility to incorporate environmental concerns into the international trading regime. After half a century and a hitherto unexperienced level of serious global environmental problems, the WTO needs to find a way to change this situation.

With the EGA, the WTO aims at making tariff exceptions on a range of environmental product groups that will be economically advantaged for widespread transfers amongst nations without the WTO having to amend its principles and laws. The EGA negotiations are however suspended since the negotiating parties would not agree on product groups that they consider important environmental goods. This paper therefore aimed at analysing this contentious WTO term - environmental goods- and provide a better understanding.

After the analysis in this article, it is claimed that environmental goods are not a concept but rather a categorisation label of the EGA. It is rather a circular definition: what environmental goods are what is listed in the EGA and what is listed in the EGA are environmental goods. The author therefore suggested to look at international environmental law and the concept of environmentally sound technologies since they have an "environmental" perspective which is considered necessary by the author for the negotiating parties of the EGA to understand to identify production groups. Indeed, the subject matter is as related to the environment as it is related to trade. This environmental perspective would also inspire the negotiating parties to choose essential environmental technology groups as environmental goods.

REFERENCES

[1] The negotiations started with: WTO, "Doha ministerial declaration," WT/MIN(01)/DEC/1, November 2001, para. 31(iii). See also, "Environmental goods agreement," available https://www.wto.org/english/tratop_e/envir_e/ega_e.htm [accessed: 25.10.2020].

[2] "Environmental goods agreement," available https://www.wto.org/english/tratop_e/envir_e/ega_e.htm [accessed: 25.10.2020].

[3] Intergovernmental Panel on Climate Change, Methodological and technological issues in technology transfer: A special report of IPCC working group III, Cambridge Uni. Press, 2000, p. 3.

[4] Trade Partnership Worldwide, "Value of the environmental goods agreement: Helping China meet its environmental goals," 2016, available https://www.uschina.org/sites/default/files/Value%20of%20the%20Environmental%20Goods%20Agreement%20on%20China%20%28English%29.pdf [accessed: 25.10.2020].

[5] K. Mahlstein, and C. McDaniel, "The environmental goods agreement: How would US households fare?," International Centre for Trade and Sustainable Development, March 2017, p. vi.

[6] J. Monkelbaan, "Using Trade for Achieving the SDGs: The Example of the Environmental Goods Agreemen," Journal of World Trade, vol. 51(4), pp. 575-604, 2017.

[7] M. Sugathan, "Lists of environmental goods: An overview," International Centre for Trade and Sustainable Development, 2013.

[8] See the source in [3].

[9] Marrakesh Agreement Establishing the World Trade Organization, 1867 UNTS 14, 1994, art. III(1).

[10] General Agreement on Tariffs and Trade, 1867 UNTS 187, 1994, art. III. General Agreement on Trade in Services, 1869 UNTS 183, 1994, art. 17. Agreement on Trade Related Aspect of Intellectual Property Rights, 1869 UNTS 299, 1994, art. 3.

[11] World Customs Organization, "What is the Harmonized System (HS)?," available http://www.wcoomd.org/en/topics/nomenclature/overview/what-is-the-harmonized-system.aspx [accessed: 25.10.2020].

[12] D. Yu, "The harmonised system – amendments and their impact on WTO mebers' schedules", World Trade Organization Economic Research and Statistics Division Staff Working Paper ERSD-2008-02, 2008, p.1.

[13] Asia-Pacific Economic Cooperation, "Annex C - APEC list of environmental goods," September 2012, available http://www.sice.oas.org/TPD/EGA/Negotiations/54goodsAPEC_e.pdf [accessed: 25.10.2020]. R. Steenblik, "Environmental Goods: A Comparison of the APEC and OECD Lists," OECD Trade and Environment Working Paper No 2005-04, COM/ENV/TD(2003)10/FINAL, 2005, available https://www.oecd.org/tad/envtrade/35837840.pdf [accessed: 25.10.2020].

[14] Joint statement regarding trade in environmental goods, January 2014, para. 4.

[15] For briefs of each round, see Global Affairs Canada, "Environmental Goods Agreement (EGA)," available http://www.international.gc.ca/trade-agreements-accords-commerciaux/topics-domaines/env/plurilateral.aspx?lang=eng [accessed: 25.10.2020].

[16] European Commission, "Trade Sustainability Impact Assessment on the Environmental Goods Agreement: Final Report," 2016, p. 3.

[17] For briefs of each round, see the source in [15].

[18] UNEP International Environmental Technology Centre, Phytotechnologies: A Technical Approach in Environmental Management, 1st ed., United Nations Publications Freshwater Management Series No 7, , 2003, pp. 7-9.

[19] European Commission, The environmental goods and services sector: a data collection handbook, Eurostat Luxembourg: Office for Official Publications of the European Communities, 2009, pp. 102-106.

[20] Ibid, p. 105 points the need for updates. See also, Submissions by New Zealand, "Environmental Goods," TN/TE/W/46, February 2005, paras. 13-18.

[21] I. Wind for International Centre for Trade and Sustainable Development, "HS Codes and the Renewable Energy Sector," 2010, pp. 3-4. T. Hu, "Redefining EGS to solve severe climate and other global environmental problems," in Trade in the Balance: Reconciling Trade and Climate Policy Report of the Working Group on Trade, Investment, and Climate Policy, F. S. Pardee, 2016, pp. 62-63.

[22] P. Wooders, "Greenhouse Gas Emission Impacts of Liberalizing Trade in Environmental Goods," International Institute for Sustainable Development, 2009.

[23] Ibid.

[24] Ibid, p. 19.

[25] IPCC, Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Geneva, 2014.

[26] K. E. Maskus, "Encouraging International Technology Transfer," UNCTAD-ICSD Project on IPRs and Sustainable Development, Issue Paper No 7, 2004, p. 7. M. Araya, "The Relevance of the Environmental Goods Agreement in Advancing the Paris Agreement Goals and SDGs: A Focus on Clean Energy and Costa Rica's Experience," International Centre for Trade and Sustainable Development Issue Paper, December 2016.

[27] WTO, "Trade and Environment at the WTO," 2004, p. 34, available https://www.wto.org/english/tratop_e/envir_e/envir_wto2004_e.pdf [accessed: 25.10.2020]. M. Wu, "Why Developing Countries Won't Negotiate: The Case of the WTO Environmental Goods Agreement," Trade Law & Development ,vol. 6, pp. 93-176, 2014.

[28] UNCTAD, "Compendium of International Arrangements on Transfer of Technology: Selected Instruments," UNCTAD/ITE/IPC/Misc.5, 2001.

[29] United Nations Conference on the Human Environment, 11 ILM 1416, 1972, principle 20.

[30] M. L. Larsson, The Business of Global Energy Transformation: Saving Billions Through Sustainable Models, Palgrave Macmillan, 2012, pp. 16-104.

[31] See the source in [3].

[32] Vienna Convention for the Protection of the Ozone Layer, 1513 UNTS 293, March 1985, art. 1(3).

[33] A similar approach is taken in United Nations Conference on Environment & Development, [3 to 14 June 1992], para. 34.1 and in the source [3] p.3.

[34] United Nations Conference on Environment & Development, [3 to 14 June 1992], para. 34.1.

[35] This is the author's comment after analysing an extensive amounth of projects for her PhD thesis which will be published. Ezgi Ediboğlu, unpublished.

[36] UNFCCC Secretariat and UNEP DTU Partnership, "Enhancing Implementation of Technology Needs Assessments Guidance for Preparing a Technology Action Plan," 2016, p. 2, available http://unfccc.int/ttclear/misc_/StaticFiles/gnwoerk_static/TEC_column_M/33933c6ccb7744bc8fd643feb0f8032a/82af010d04f14a84b9d24c5379514053.pdf [accessed: 25.10.2020].

[37] See the source in [3], p. 3.

[38] See the source in [19], p. 105.

[39] See the source in [19], p. 105 that cites OECD, The global environmental goods and services industry, OECD Publications, Paris.

[40] P. Birnie, A. Boyle, and C. Redgwell, International law and the environment, 3rd ed., Oxford University Press, 2009.

Using Algorithms for the Prediction of Low-risk and High-risk Human Papillomavirus (HPV) in Males

Ogbolu Melvin Omone[1]
BioTech Research Center, EKIK,
Óbuda University,
Budapest, Hungary,
ogbolu.melvin@biotech.uni-obuda.hu

Alex Ugochukwu Gbenimachor[2]
Dipartimento di Ingegneria Informatica,
Automatica e Gestionale "A. Ruberti",
Sapienza Università di Roma,
Rome, Italy,
gbenimachorugochuwkwu@gmail.com

Miklós Kozlovszky[3]
BioTech Research Center, EKIK,
Óbuda University,
Budapest, Hungary,
kozlovszky.miklos@nik.uni-obuda.hu

***Abstract*— This study focus on approximating the prevalence of Human papillomavirus (HPV) among both male staff and students of Ibrahim Badamasi Babangida University in Nigeria, and to analyze both high-risk(hr) and low-risk(lr) factors associated with the persistence of the virus in men which is the leading cause of cervical cancer in women after sexual intercourse. When an HPV-infected man is left untreated for a long period of time, it raises the opposite sex (women) at a future risk of cervical cancer development in the future after the occurrence of sexual intercourse. This paper model high-level python algorithms to categorize, quantify, classify, and predict the male participant(s) who belong to HPV high-risk or low-risk group using Neural Network Algorithm (NNA), Random Forest Algorithm (RFA), and K-Means Clustering Algorithm (KMCA).**

Keywords*— *Cervical Cancer, Human papillomavirus (HPV), HPV high-risk group, HPV low-risk group, K-Means Clustering Algorithm (KMCA), Neural Network Algorithm (NNA), Random Forest Algorithm (RFA)

I. Introduction

Recently, many studies have focused on the prevalence of Human papillomavirus (HPV) and its effects among women only, thereby neglecting the fact that men are the leading cause of this persistent virus [1]. HPVs are a non-enveloped and icosahedral double-stranded DNA [2] kind of viruses which are known to be a common Sexually Transmitted Infection (STI) from the male to the female during sexual intercourse [3], [4]. This viral infection is classified into low-risk HPV (lrHPV) and high-risk HPV (hrHPV), which are assigned with numbers called subtypes and causing different types of physiological pathosis. The high-risk HPV subtypes are HPV–16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59, 68, 73 and 82, while the low-risk HPV subtypes are HPV–6, 11, 40, 42, 43, 44, 54, 61, 70, 72, 81, and 89 [5]. Many epidemiological studies have revealed that infections from low-risk HPV subtypes usually clear over a certain period of time without medical interventions while infections from high-risk HPV subtypes mostly result into cancerous diseases [5]. The resulting physiological pathosis mainly affects both the reproductive system of men and women; which is caused by high-risk HPV-infections (subtypes are mostly HPV-16 and 18 which are classified as oncogenic HPV subtypes) in women results into cancer of the cervix, cancer of the vulva, cancer of the vagina, and cancer of the anus, while in men, high-risk HPV-infections results into cancer of the anus and cancer of the penis [1], [6], [7]. However, low-risk HPV-infections (subtypes are mostly HPV-6 and 11 which are classified as non-oncogenic HPV subtypes) in men and women results into anogenital warts (i.e. benign epithelial tumors) [8], oral lesions(i.e. ulcer of the oral cavity/mouth ulcer), common warts, and plantar warts [9].

A. The Development of Human Papillomavirus (HPV) In Men

In 1996, a group of researchers revealed that due to the sexual behavior among men, the risk of cervical cancer development among women has been 5-times to 9-times on the increase in Spain. This means that men with multiple sexual partners and married men with extramarital sexual partners who are mostly infected by HPV-16, 18, 31, and 33 subtypes tend to put their wives at a higher risk of cervical cancer [10]. Hence, related studies have revealed that men are the utmost carriers of HPV infections [11] and it is likely to persist overtime in men than in women [1]. The persistence of HPV in men has resulted into both life-threatening and noncritical kind of diseases [12]. Certain life-threatening types of these diseases are the Anogenital Tract (AT) and Upper Aero-Digestive Tract (UADT) cancers and other kinds of carcinoma in-situ are minimally invasive cancers [4].

The most commonly observed pathway for HPV infection is sexual intercourse; HPV can only be transmitted during unsafe sexual practices through vaginal sex, oral sex, and anal sex [13]. Unsafe sexual practices is the act of having unprotected sexual intercourse [6] with an HPV-infected persons or person (mostly opposite sex) [14] which is a risk factor for the prevalence of HPV infection in men. However, it has been revealed that men who have sex with men (MSM) are at a higher risk of HPV-infection caused by anal sex. Although, men who are more at a higher risk are specifically associated with receptive anal sex and a multiple number of sexual partners [15]. When a man is HPV-infected with the HPV (subtypes such as 16 and 18), the persistent virus cannot directly lead to cancer. The virus is left untreated for a long period of time, develops into precursor lesions [15] and then begins to spread around the reproductive system. The development of HPV in men is likened to the development of an infected host cell-cycle. When a man is infected by HPV, the virus invades a host cell in the basal layer of epithelia (i.e. the innermost layer of the continuous sheets of the skin cells) through minor cuts and scrapes [16]. Overtime, the host cell begins to divide (this process is known as replication cycle), thereby replicating itself [17]. When the HPV DNA is replicated, it further leads to genetic mutations in the host cells that can possibly develop into cancer in future [18].

II. Materials and Methods

A. Study Population

This study includes over one hundred and fifty (150) male participants of Ibrahim Badamasi Babangida University in Nigeria, which were both students and staff of the university. All the participants who participated in the assessment were provided with a written consent that informed them about of the purpose of the study and the knowledge they would gain as a result of participating. However, the study remains

anonymous in order to respect and protect the participants privacy. Among many who volunteered to participate in the assessment, only 100 participants were qualified to be included in the study. The selection criteria were based on the following;

- Age: Participants who indicated to be between 18 – 60 years old.
- Demographic factor: Participants who are student/staff of Ibrahim Badamasi Babangida University in Nigeria.
- Record of sexual experience: Participants who indicated to have had sex in their lifetime, one/more sexual partners, and practice protected/unprotected sex.

B. Data Collection

An online questionnaire, known as Human Papillomavirus (HPV) Assessment Test (HAT) was developed with the use of Google forms as a tool for data collection which included complete anonymity of the participants data following the currently established General Data Protection Regulation (GDPR) privacy policies [19]. The online questionnaire was designed in form of an assessment with embedded scores/points which are automatically evaluated and summed up to show the level of HPV risk for each participant depending on each option selected while responding to the questions. The overall assessment was 70 points, whereby participants with ≤20 points are categorized as HPV low-risk carriers while participants with ≥21 points are categorized as HPV high-risk carriers. There were fifteen (15) questions included in the questionnaire which could provide answers about the participant's social and demographic data, safe/unsafe sexual practices, sexual intercourse activities/behavior, and knowledge, attitude, and perception (KAP) as related to HPV.

C. Data Analysis and Statistical Analysis

Data analysis was carried out using python programming algorithms and descriptive statistics. The data contains a detailed information of males who are likely to be infected with HPV based on the HPV-related risk factors contained in the questionnaire. These risk factors are most closely associated with HPV-infection, which systematically shows the level of risk according to their Age at First Sexual Intercourse (AFSI), Multiple Number of Sexual Partners (MNSP), Sexual Lifestyle and Practices (SLP), and Knowledge, Attitude and Perception/Practices (KAP) of HPV, HPV vaccine, and HPV-related infections. Males who tend to have a high-risk of HPV infection as categorized based on the risk factors are potential HPV carriers and are the cause of cervical cancer in females who get infected by them and are not treated for several years. As for data analysis, fifteen (15) questions were selected from the questionnaire, which was used to classify which participants are categorized as high-risk/low-risk HPV carriers which is in correlation with the total score of the assessment of each participants in the data. In order to achieve this classification, the Neural Network Algorithm (NNA), Random Forest Algorithm (RFA), and K-means Clustering algorithms were applied using python programming language.

1) Neural Network Algorithm (NNA)

The Neural Network Algorithm (NNA) is an algorithm that imitates the neuron in the human brain [20]. In this paper the NNA takes series of input variable from the male HPV dataset as input ($X_1, ..., X_n$) and then learns the dependent output data as output ($Y_1, ..., Y_n$) [21]. The input variables includes both low- and high-risk factors of HPV which are used to predict at random the output data (i.e. to predict participants who belongs to low/high-risk HPV group) [22]. The NNA takes in data as binaries (0 and 1). Hence, in order to learn multiclass variable in the case of the male HPV dataset with various scores (which helps to predict the output group: low/high-risk), the input results is taken from product operation [21].

2) Random Forest Algorithm (RFA)

Random Forest Algorithm (RFA) is an algorithm for supervised learning which is mainly used for classification problems/sample classification [23]. It can be applied to perform regression and classification for data analysis. RFA is known to have the attributes/characteristics related to decision trees (i.e. Information Gain, Gini Impurity, and Information Entropy); this algorithm will be applied to each of the features (i.e. risk factors contained in the data) in order to classify the participants into their classes (low/high-risk group). Hence, yields different decision trees [24], [25]. In this paper, RFA classifies the male HPV dataset by selecting which feature(s) mostly defines or contains a detailed information about all given classes.

3) K-Means Clustering Algorithm (KMCA)

Firstly, clustering is an unsupervised learning approach for classifying data; by partitioning/grouping the data into smaller clusters given the features represented in the data [26], [27]. Hence, K-Means Clustering Algorithm (KMCA) is the classification of *n* features/observation data and *k* integer clusters [28]. In this paper, the KMCA worked perfectly for features represented graphically as it is the most efficient kind of algorithm used for the male HPV dataset. Therefore, the high-risk HPV group were clearly separated from the low-risk group.

III. Results and Discussion

A. Sociodemographic Factors

TABLE I. Sociodemographic Factors

Factors	Count	Percentage
Age Group (years)		
18 – 25	23	23%
26 – 30	23	23%
31 – 40	41	41%
41 – 50	10	10%
51 – 60	3	3%
Marital Status		
Single	61	61%
Married	37	37%
Divorced	2	2%
Monthly Income (NGN)		
≤ 30,000	16	16%
30,000 – 60,000	23	23%
60,000 – 90,000	22	22%
90,000 – 120,000	13	13%
≥ 120,000	26	26%
Level of Education		
Elementary/Primary School	2	2%
High School/Secondary School	13	13%
Bachelor's Degree	54	54%
Master's Degree	24	24%
Doctorate Degree	7	7%

Using descriptive statistics, Table 1 reveals certain significant percentages for the population sociodemographic

factors. The population shows a higher number of middle-aged; thus 41(41%) of the participants indicated to be between the age of 31 – 40 years. Hence, the population contained few young-adults and very few older-adult, where 23(23%) indicated to be between 18 – 25 years, 23(23%) indicated to be between 26 – 30 years, 10(10%) indicated to be between 41 – 50 years and only 3(3%) indicated to be between 51 – 60 years.

Since the assessment was mostly taken by participants who work/study in an university (i.e. an academia environment), statistics reveals that most of the participants were singles; thus 61(61%) indicated to be singles, 37(37%) indicated to be married, and only 2(2%) indicated to be divorced.

This study shows that most of the participants do not earn enough income to care for their needs or pay for medical bills; thus, 16(16%) earned ≤#30,000, 23(23%) earned between #30,000 – #60,000, 22(22%) earned between #60,000 – #90,000, 13(13%) earned between #90,000 – #120,000, and 26(26%) earned ≥#120,000 above. Poorly, 16(16%) of the participants who indicated to earn ≤#30,000 monthly which is the same as the National Minimum Wage (NMW) for federal workers in Nigeria could be living in abject poverty. Since, statistics discovered that for an individual who resides in Nigeria, #43,200 is the estimated cost of living while for a family of four is estimated to be #137,600 [29], [30]. Our study shows that the participants of this assessment do not earn enough income to take care of their family in case of the married ones who earns ≤#30,000 – #120,000 which is not up to #137,600 which is estimated to carter for an entire family. Hence, due to lack of fund the population are likely to be HPV-infected but are not aware since the healthcare system of Nigeria is not free, and they are unable to pay for checkups, medications and medical bills.

The population under investigation were mostly academician, hence, statistics reveals a high level of education among the population. There were only 2(2%) who indicated to hold an elementary/primary school certificate, 13(13%) indicated to hold a high school/secondary school certificate, 54(54%) indicated to hold or currently undergoing a bachelor's degree, 24(24%) indicated to hold or currently undergoing a master's degree, and only 7(7%) indicated to hold a doctorate degree. Therefore, those who indicated to hold an elementary/primary and a high school/secondary school certificate are likely to be non-teaching staff who work in the university, 54(54%) who indicated to hold or currently undergoing a bachelor's degree and 24(24%) who indicated to hold or currently undergoing a master's degree are students of the university, and 7(7%) who indicated to hold a doctorate degree are likely lecturers/instructors of the university.

B. Data Training and Loss Function

1) Training and Validation Loss for Neural Network Algorithm (NNA)

The loss graph shows a learning curve in the Neural Network Algorithm (NNA). It computes the difference between the predicted label (for scores) and the actual label (HPV scores for entire population). The HPV Assessment Test (HAT) computes a score for each participant which can be used to predict if a participant belongs to the high-risk/low-risk group. Using the NNA, if the HPV scores for entire population computes to zero (0) [31], it means that the model prediction is perfect. In this study, where the epoch=150, computation=0 for the model [32] to predict the labels given the HPV features (risk factors contained in the data). The effect of the loss function is to make sure that each datapoint of the features are classified properly/perfectly.

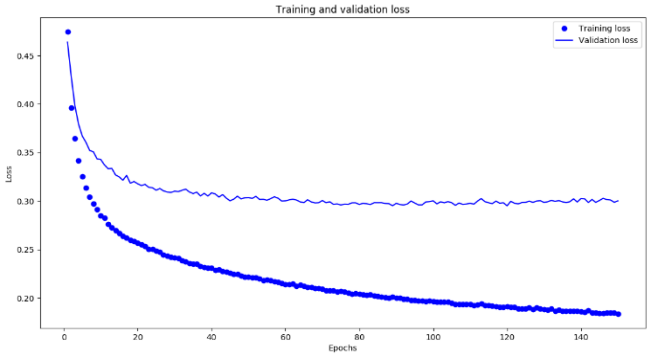


Fig 1: Loss Graph for Neural Network Algorithm (NNA)

2) Training and Validation Accuracy for Neural Network Algorithm (NNA)

The accuracy graph shows how well the Neural Network Algorithm (NNA) performed in classifying the scores (labels) given the number of input data. It shows the ratio of prediction correctness/efficacy [33] of the male HPV dataset. The Accuracy for validation and training is given as 95% and 96%, which shows that the NNA performs well in classifying each feature provided with the scores.

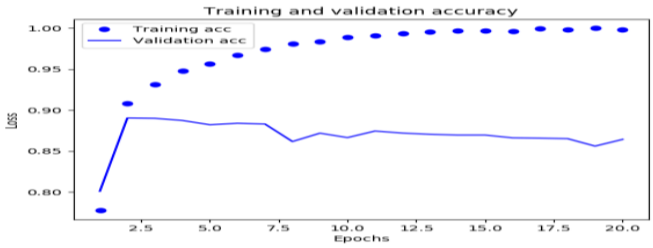


Fig 2: Accuracy Graph for Neural Network Algorithm (NNA)

3) Training and Validation Loss for Random Forest Algorithm (RFA)

In computing the loss graph for Random Forest Algorithm (RFA), the Out of the Bag (OOB) estimate was used to compute the average loss for each computed prediction from each boosted decision tree that do not contain in their respective bootstrap samples. The training of the male HPV dataset demonstrates how the OOB estimate can be measured at the addition of each new tree during the training [34].

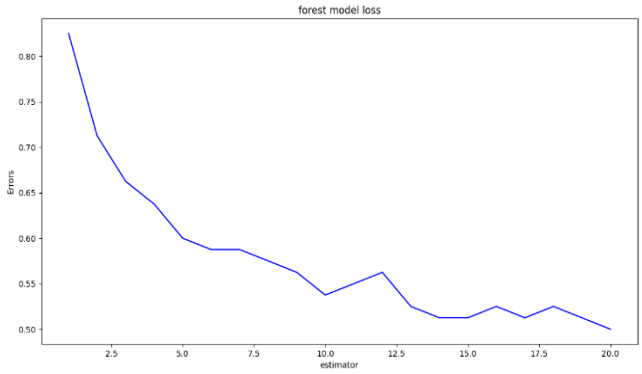


Fig 3: Loss Graph for Random Forest Algorithm (RFA)

4) Training and Validation Accuracy for Random Forest Algorithm (RFA)

The accuracy achieved in training the Random Forest Algorithm (RFA) on the male HPV dataset is 0.45/45% given 20 trees (was done using 20 estimates to classify the data).

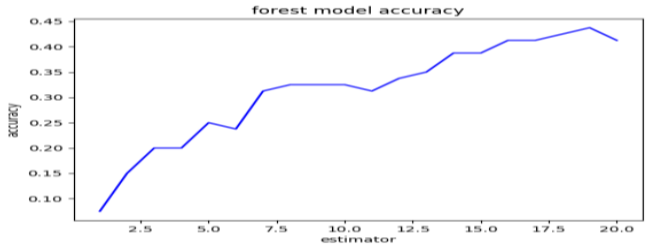


Fig 4: Accuracy Graph for Random Forest Algorithm (RFA)

5) Training and Validation Loss for K-Means Clustering Algorithm (KMCA)

In order to compute the loss function for K-Means Clustering Algorithm (KMCA) which basically tells the difference between the predicted labels and actual labels. The KMCA uses the 'elbow method' [35] which measures/computes the number of clusters in the male HPV datasets. The 'elbow method' performs a series of *k* iterations and record the Sum Square Errors (SSE) [36]. During the training of the male HPV dataset, there is a convergence from *n* clusters from *k-5* to *k-20*.

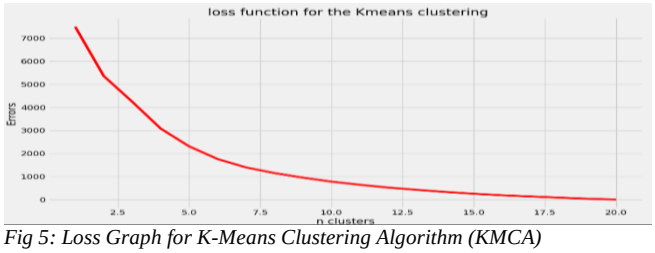


Fig 5: Loss Graph for K-Means Clustering Algorithm (KMCA)

6) Training and Validation Accuracy for K-Means Clustering Algorithm (KMCA)

The K-Means Clustering Algorithm (KMCA) gave an accuracy of 0.26 or 26% and the resultant effect on the algorithm being an unsupervised algorithm is to learn the similarities in features and group in their label clusters [28]. The result here showed that the KMCA performed poorly in classifying the features to their respective labels.

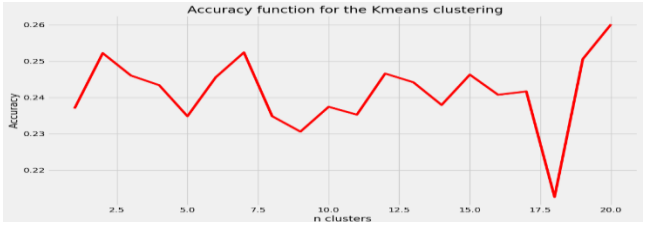


Fig 6: Accuracy Graph for K-Means Clustering Algorithm (KMCA)

C. Prediction for HPV High-risk/Low-risk Group

For classification, low-risk HPV (lrHPV) group/carriers are participants with ≤20 points while high-risk HPV (hrHPV) group/carriers are participants with ≥21 points.

1) Classification with NNA for MNSP and AFSI

As shown in Figure 7 below, NNA was used for the classification of the risk factors/features for HPV (i.e. Multiple Number of Sexual Partners (MNSP) and Age at First Sexual Intercourse (AFSI)). These features are classified into categories which are then identified by different colors given by each score labels. As mentioned above, lrHPV group are participants with ≤20 points while hrHPV group are participants with ≥21 points. The high-risk group have a higher tendency of infecting women with HPV; they are clearly identified by colors and shows participants with MNSP who started having sex at a young age (AFSI), scores are 20/70 – 25/70 points.

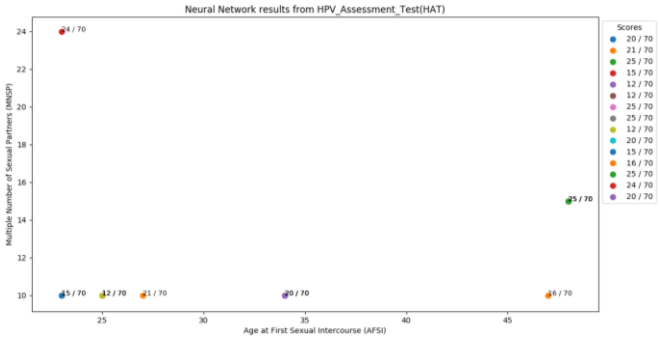


Fig 7: Classification with NNA for MNSP and AFSI

2) Clustering with KMCA for MNSP and AFSI

Using KMCA as shown in Figure 8 below, predicted labels reveals clustered features identified with each colored label. These colors are chosen at random to identify each of the labels. This tells which group given their labels have a higher tendency of infecting women with HPV and which group does not. The predicted labels are represented in the clusters of their own colors. The graph below shows a group of participants MNSP and AFSI; these features are grouped according to their scores and are described in the legend of the graph. Scores are 10/70 – 60/70 points. Each of the scores measures the risk each male-participant poses to their partners (women). The graph shows a good prediction with true labels.

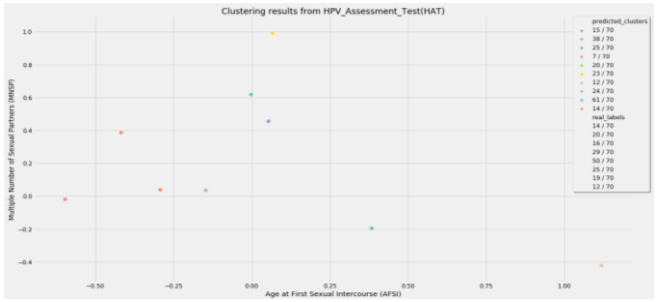


Fig 8: Clustering with KMCA for MNSP and AFSI

3) Clustering with KMCA HPV Vaccinated/Non-vaccinated Participants and HPV Infection History

Figure 9 shows the features for participants who have been vaccinated or not vaccinated against with HPV infection and those with medical history for HPV infection. Each predicted label shows the classification for each predicted cluster in the graph; it shows the cluster of features 7/70 – 61/70 point. Showing. The predicted hrHPV group in the graph are 22, 36, 38, and 61 points.

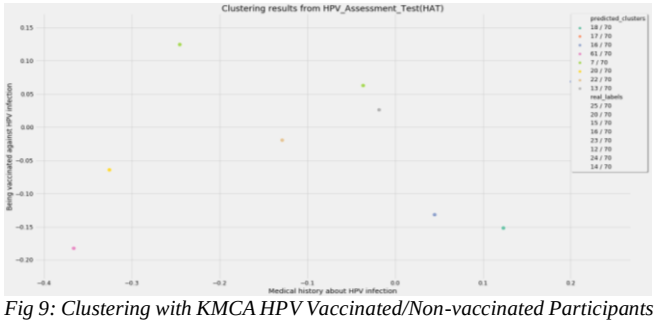


Fig 9: Clustering with KMCA HPV Vaccinated/Non-vaccinated Participants and HPV Infection History

4) Clustering with KMCA for Knowledge about HPV Genotypes and HPV Vaccinated/Non-vaccinated Participants

Using KMCA, Figure 10 shows that the participants who have no knowledge about genotypes and who have not been HPV vaccinated are at high-risk of HPV infection. The predicted clusters are used to classify and differentiate the low-risk group from the high-risk group. The data points are classified according to the predicted clusters. It also shows prediction for true labels and real labels. The legend shows predicted clusters are the high-risk HPV group.

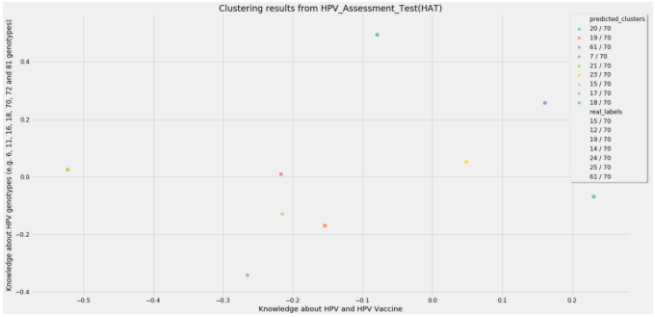


Fig 10: Clustering with KMCA for Knowledge about HPV Genotypes and HPV Vaccinated/Non-vaccinated Participants

5) Classification with NNA for SLP

Sexual Lifestyle and Practices (SLP) is used to check for the Sexual Lifestyle and Practices (oral sex, anal sex, vaginal sex, and protected/unprotected sex practices) of the participants. The graph below shows the classification between SLP for oral, anal, and vaginal sex and protected/unprotected sex practices using NNA. The male HPV dataset was well classified according to each label given the features set data points. However, some labels overlapped into single data point/features. High-risk participants have 21, 24, 29, and 50 points.

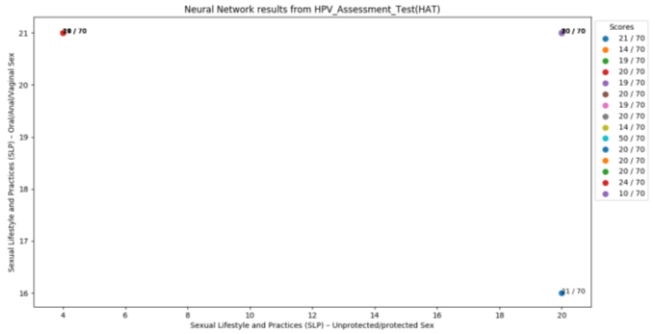


Fig 11: Classification with NNA for SLP

6) Classification with NNA for Knowledge about HPV Genotypes and HPV Vaccinated/Non-vaccinated Participants

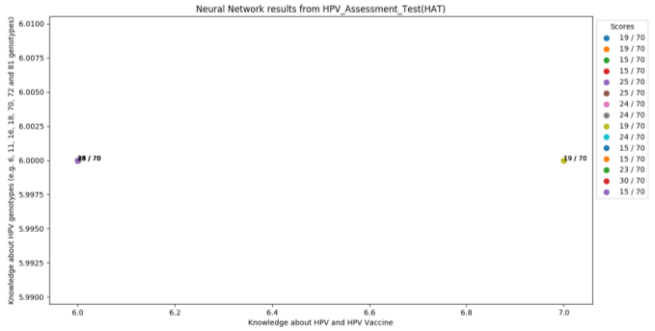


Fig 12: Classification with NNA for Knowledge about HPV Genotypes and HPV Vaccinated/Non-vaccinated Participants

Figure 12 above classifies participants who have no knowledge about HPV genotypes and who have not been vaccinated against HPV infection using data points. These data points are features sets and classified given the scores/labels. It classified participants who scored 15/70, 25/70, and 19/70 as those with high-risk tendencies of infecting women with HPV. However, it is obvious that NNA doesn't work perfectly for classifying these risk factors.

7) Classification with RFA for Knowledge about HPV Genotypes and HPV Vaccinated/Non-vaccinated Participants

As shown in Figure 13, the RFA was able to make same classification as NNA. It classified the data point features to labels as in the scores/labels for participants who have no knowledge about HPV genotype and those who have not been vaccinated against HPV infection.

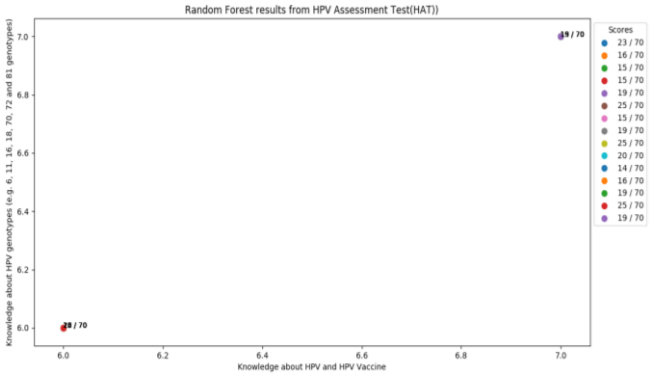


Fig 13: Classification with RFA for Knowledge about HPV Genotypes and HPV Vaccinated/Non-vaccinated Participants

8) Classification with RFA for AFSI and HPV Vaccinated/Non-vaccinated Participants

Figure 14 shows that participants who falls under the category of early Age at First Sexual Intercourse (AFSI) and those who have not been vaccinated against HPV infection belong to the HPV high-risk group. The classification for each data features into the given scores/labels are identified by different colors.

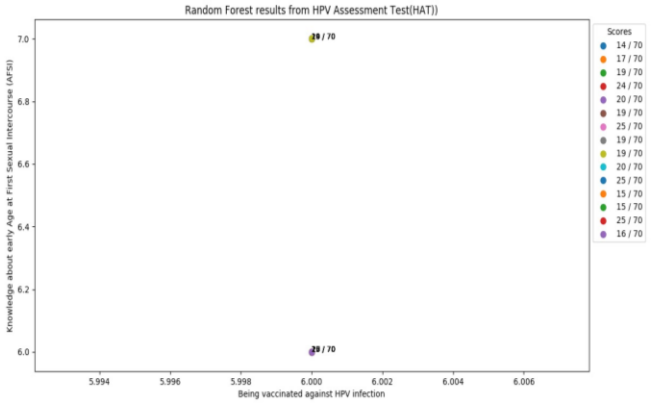


Fig 14: Classification with RFA for AFSI and HPV Vaccinated/Non-vaccinated Participants

9) Classification with RFA for SLP

Figure 15 shows result for participants with Sexual Lifestyle and Practices (oral sex, anal sex, vaginal sex, and protected/unprotected sex practices). The result shows data points with features of participants with risk given different scores/labels. Hence, the legend as shown in the figure below is the same as the classification in the NNA as in Figure 11.

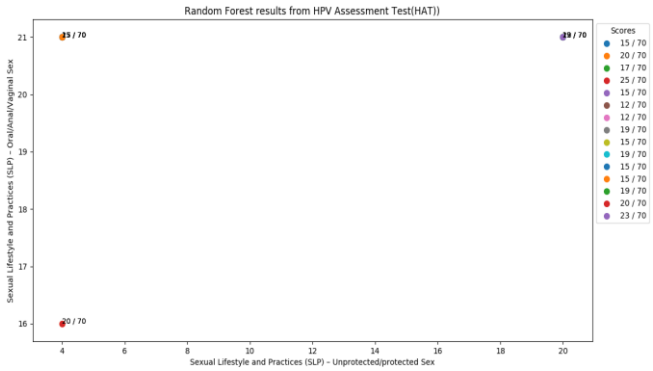


Figure 15: Classification with RFA for SLP

IV. Conclusion

The algorithms used in this paper shows good predictions with Figures 7, 8, 9, 10, 11, and 15 while Figures 12, 13, and 14 did not show good predictions. The limitation experienced in using Random Forest Algorithm (RFA) as shown in Figures 12, 13, and 14 is due to the amount of dataset used for analysis. However, Neural Network Algorithm (NNA) out-performed the K-Means Clustering Algorithm (KMCA) and the Random Forest Algorithm (RFA). Hence, it is believed that other python algorithms could do better if implemented and with the availability of a large-scale data to analyze or learn the features contained in the dataset. The purpose of this project was achieved, the algorithms classified the scores/labels given the features in the male HPV dataset that was learned during the training and validation process. In future, to get better accuracy for a better classification of the features, the Human Papillomavirus (HPV) Assessment Test (HAT) will be distributed to a larger population in order to collect large-scale data.

Acknowledgment

The authors thank the EFOP-3.6.1-16-2016-00010 project and the GINOP-2.2.1-15-2017-00073 named "Telemedicina alapú ellátási formák fenntartható megvalósítását támogató keretrendszer kialakítása és tesztelése, Hungary for their financial support.

References

[1] A. R. Giuliano, G. Anic, and A. G. Nyitray, "Epidemiology and pathology of HPV disease in males," *Gynecol. Oncol.*, vol. 117, no. 2 Suppl, pp. S15-19, May 2010.

[2] "Human Papillomavirus (HPV) Infection - Human Papillomaviruses - NCBI Bookshelf." https://www.ncbi.nlm.nih.gov/books/NBK321770/ (accessed Aug. 12, 2020).

[3] F. Cutts *et al.*, "Human papillomavirus and HPV vaccines: a review," *Bull World Health Organ*, vol. 85, no. 9, pp. 719–726, Sep. 2007.

[4] A. Castillo, "Human Papillomavirus and Carcinogenesis in the Upper Aero-Digestive Tract," in *Carcinogenesis*, K. Tonissen, Ed. InTech, 2013.

[5] O. M. Omone and M. Kozlovszky, "HPV and Cervical Cancer Screening Awareness: A Case-control Study in Nigeria," in *2020 IEEE 24th International Conference on Intelligent Engineering Systems (INES)*, Jul. 2020, pp. 145–152.

[6] M. Rodríguez-Álvarez, J. Gómez-Urquiza, H. Husein-El Ahmed, L. Albendín-García, J. Gómez-Salgado, and G. Cañadas-De la Fuente, "Prevalence and Risk Factors of Human Papillomavirus in Male Patients: A Systematic Review and Meta-Analysis," *IJERPH*, vol. 15, no. 10, p. 2210, Oct. 2018.

[7] H. Husein-El Ahmed and G. A. Cañadas-De la Fuente, "IMAGES IN CLINICAL MEDICINE. Squamous-Cell Carcinoma of the Penis with Human Papillomavirus," *N. Engl. J. Med.*, vol. 374, no. 2, p. 164, Jan. 2016.

[8] G. Ozaydin-Yavuz, S. G. Bilgili, H. Guducuoglu, I. H. Yavuz, S. Elibuyuk-Aksac, and A. S. Karadag, "Determinants of high-risk human papillomavirus infection in anogenital warts," *Postepy Dermatol Alergol*, vol. 36, no. 1, pp. 76–81, Feb. 2019.

[9] S. N. Adebamowo, O. Olawande, A. Famooto, E. O. Dareng, R. Offiong, and C. A. Adebamowo, "Persistent Low-Risk and High-Risk Human Papillomavirus Infections of the Uterine Cervix in HIV-Negative and HIV-Positive Women," *Front Public Health*, vol. 5, Jul. 2017.

[10] F. X. Bosch *et al.*, "Male sexual behavior and human papillomavirus DNA: key risk factors for cervical cancer in Spain," *J. Natl. Cancer Inst.*, vol. 88, no. 15, pp. 1060–1067, Aug. 1996.

[11] E. I. Svare, "Risk factors for genital HPV DNA in men resemble those found in women: a study of male attendees at a Danish STD clinic," *Sexually Transmitted Infections*, vol. 78, no. 3, pp. 215–218, Jun. 2002.

[12] H. A. Cubie, "Diseases associated with human papillomavirus infection," *Virology*, vol. 445, no. 1–2, pp. 21–34, Oct. 2013.

[13] N. Osazuwa-Peters *et al.*, "Understanding of risk factors for the human papillomavirus (HPV) infection based on gender and race," *Sci Rep*, vol. 9, no. 1, p. 297, Dec. 2019.

[14] Zitkute V and Bumbuliene Z: Risk Factors Affecting HPV Infection, Persistence and Lesion Progression in Women and Men, Clin Res Infect Dis 3(2): 1026, 2016.

[15] Marra, E.: Anal HPV Infection and Disease: Common and preventable, but hard to treat, 2018.

[16] A. Castillo,"HPV infection and carcinogenesis in the upper aero-digestive tract,"*Colombia Médica*, vol.42, no.2, pp.233–242, Jun. 2011.

[17] S. V. Graham, "The human papillomavirus replication cycle, and its links to cancer progression: a comprehensive review," *Clinical Science*, vol. 131, no. 17, pp. 2201–2221, Sep. 2017.

[18] "New Clues About How HPV Spreads In the Body." https://www.cancer.org/latest-news/new-clues-about-how-hpv-spreads-in-the-body.html#for_researchers (accessed Aug. 12, 2020).

[19] "General Data Protection Regulation (GDPR) – Official Legal Text." https://gdpr-info.eu/ (accessed Mar. 02, 2020).

[20] E. Lin and S.-J. Tsai, "Machine Learning in Neural Networks," in *Frontiers in Psychiatry*, vol. 1192, Y.-K. Kim, Ed. Singapore: Springer Singapore, 2019, pp. 127–137.

[21] J. Patel and R. Goyal, "Applications of Artificial Neural Networks in Medical Science," *CCP*, vol. 2, no. 3, pp. 217–226, Sep. 2007.

[22] J. A. M. Sidey-Gibbons and C. J. Sidey-Gibbons, "Machine learning in medicine: a practical introduction," *BMC Med Res Methodol*, vol. 19, no. 1, p. 64, Dec. 2019.

[23] D. Denisko and M. M. Hoffman, "Classification and interaction in random forests," *Proc Natl Acad Sci U S A*, vol. 115, no. 8, pp. 1690–1692, Feb. 2018.

[24] D. P. Bhukya and S. Ramachandram, "Decision Tree Induction: An Approach for Data Classification Using AVL-Tree," *IJCEE*, pp. 660–665, 2010.

[25] C. Kingsford and S. L. Salzberg, "What are decision trees?," *Nat Biotechnol*, vol. 26, no. 9, pp. 1011–1013, Sep. 2008.

[26] M. Z. Rodriguez *et al.*, "Clustering algorithms: A comparative approach," *PLoS ONE*, vol. 14, no. 1, p. e0210236, 2019.

[27] M. R. David and S. Samuel, "Clustering of PubMed abstracts using nearer terms of the domain," *Bioinformation*, vol. 8, no. 1, pp. 20–25, Jan. 2012.

[28] H. Alashwal, M. El Halaby, J. J. Crouse, A. Abdalla, and A. A. Moustafa, "The Application of Unsupervised Clustering Methods to Alzheimer's Disease," *Front Comput Neurosci*, vol. 13, May 2019.

[29] "Nigeria National Minimum Wage | 2018-2020 Data | 2021-2022 Forecast | Historical." https://tradingeconomics.com/nigeria/minimum-wages (accessed Jul. 02, 2020).

[30] "• Nigeria: minimum wage 2018-2020 | Statista." https://www.statista.com/statistics/1119133/monthly-minimum-wage-in-nigeria/ (accessed Sep. 05, 2020).

[31] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights Imaging*, vol. 9, no. 4, pp. 611–629, Jun. 2018.

[32] A. R. Asif *et al.*, "Performance Evaluation of Convolutional Neural Network for Hand Gesture Recognition Using EMG," *Sensors (Basel)*, vol. 20, no. 6, Mar. 2020.

[33] L. Bertolaccini, P. Solli, A. Pardolesi, and A. Pasini, "An overview of the use of artificial neural networks in lung cancer research," *J Thorac Dis*, vol. 9, no. 4, pp. 924–931, Apr. 2017.

[34] S. Janitza and R. Hornung, "On the overestimation of random forest's out-of-bag error," *PLoS One*, vol. 13, no. 8, Aug. 2018.

[35] E. Demidenko, "The next-generation K-means algorithm," *Stat Anal Data Min*, vol. 11, no. 4, pp. 153–166, Aug. 2018.

[36] T. Weißer, T. Saßmannshausen, D. Ohrndorf, P. Burggräf, and J. Wagner, "A clustering approach for topic filtering within systematic literature reviews," *MethodsX*, vol. 7, Feb. 2020.

Synthesis and analysis of AGV systems

Rozalia Lakner
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
lakner.rozalia@amk.uni-obuda.hu

Botond Bertók
Faculty of Information Technology
University of Pannonia
Veszprém, Hungary
bertok@dcs.uni-pannon.hu

***Abstract*—Based on the similar structure of P-graphs and Petri nets, it is beneficial to merge their capabilities in solving specific problems. In the current work, the algorithmic synthesis of Petri nets by the P-graph framework is introduced and illustrated by an example of route planning of automated guided vehicle (AGV) systems.**

Keywords—colored Petri nets, P-graph, automated guided vehicle.

I. Introduction

The purpose of the automated guided vehicle (AGV) systems is to transfer objects between given points automatically, efficiently and safely.

To achieve this, the synthesis of feasible pathways is an important task in the AGV system design, where the synthesized models are used for simulation and analysis of the system.

II. Synthesis of Feasible Pathways of AGV System

Essentially, the synthesis is the determination of the process structure, that behave as required. For highly effective algorithmic solution of the synthesis problem, a mathematically rigorous framework, so called P-graph was introduced [1]. This method based on the combinatorial mathematics of graph theory, and a set of axioms expressing the properties of the feasible structures [2]. The P-graph framework generates the best and all of the alternative structures of the system, where the structures are represented by bipartite directed graph.

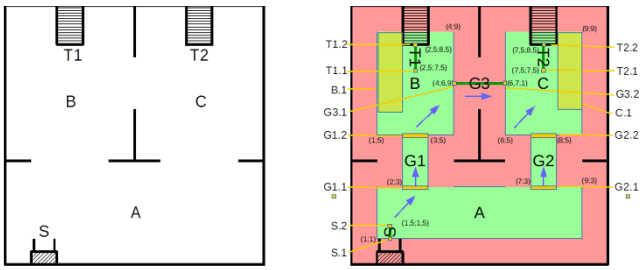


Fig. 1. The layout of the examined AGV system.

Fig. 1 illustrates the layout of the investigated AGV system, that consists of three rooms (A, B and C) and three gates connecting the rooms. There are three special stations: the charging station (S) for the AGVs in room A, the pick-up station (T1) in room B and the delivery station (T2) in room C. The AGV-s can travel among the various points of the layout for picking up and delivering packages on predefined paths. For safety movements of the AGVs, safety zones must be determined considering the dimension of the vehicle. These zones can be seen on the right-hand side of Fig. 1 with green colors, together with their special points. Based on the delivery request and its transportation order, there can be several feasible pathways involving movements of the AGV among the predefined locations.

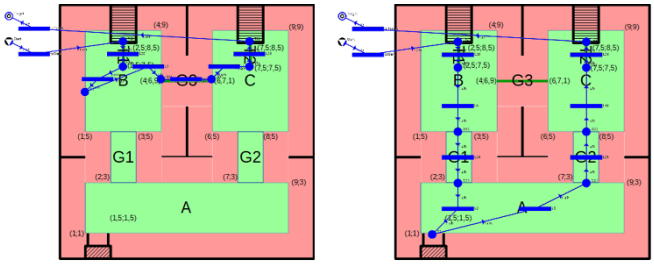


Fig. 2. Feasible pathways of the AGV system.

In Fig. 2 two of the eight feasible pathways of the delivery request is depicted, where the synthesis can be done by the publicly available software P-graph Studio [3]. The nodes of P-graph are the special places of the investigated system and the possible actions or vehicle movements between two neighboring places [4]. The starting and target points of AGV are at T1.2 and T2.2, respectively. The entity type nodes represented by solid circles in the P-graph represents potential vehicle locations, while activity type nodes represented by horizontal bars the possible vehicle movements between two neighboring locations.

Besides all the feasible pathways, the P-graph algorithms generate the maximal structure, that is the union of the feasible solution structures. The maximal structure of the AGV example is shown in Fig. 3.

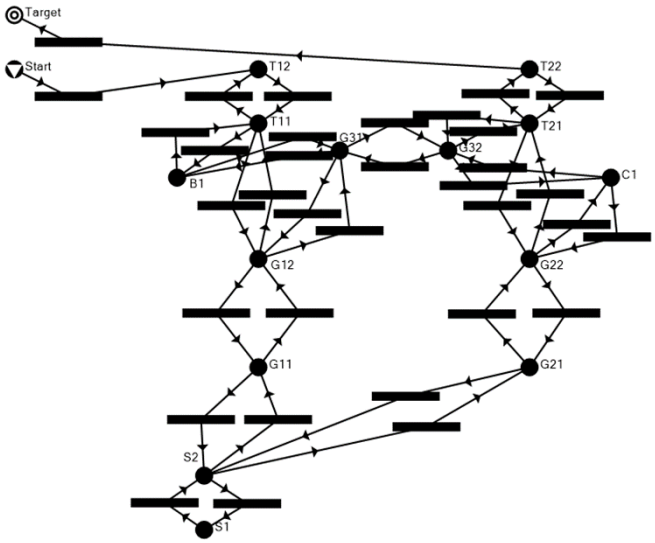


Fig. 3. The maximal structure of the AGV system.

P-graphs and Petri nets [5] have similar structures based on rigorous formal mathematical definitions and graphical notations. The similarity is observed in Fig 4., where the Petri

net model of the layout of the AGV system constructed by CPN Tools [6] is shown.

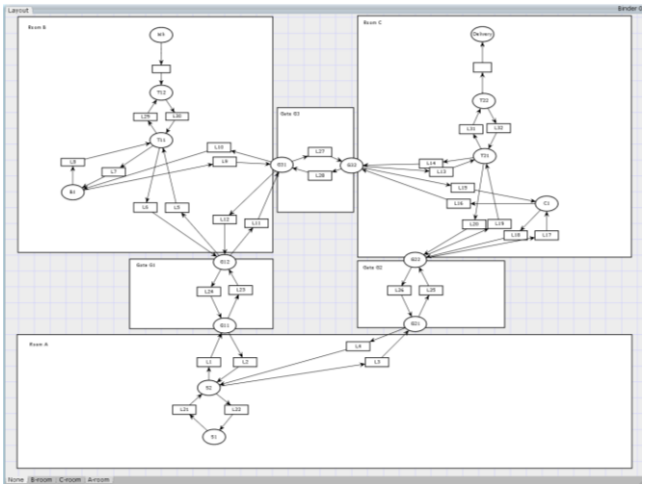


Fig. 4. The layout of the AGV system represented by Petri net.

The two frameworks aim at different objectives of systems design. While Petri nets are introduced primarily for modeling and analysis, P-graph framework has been developed for process synthesis. Even though the objectives are quite different, it is possible to merge their capabilities in solving specific problems, e.g., synthesizing Petri nets by the P-graph framework [7].

III. Simulation and Analysis of AGV System with Colored Petri Nets

Based on the similar structure of P-graph and Petri nets, the synthesized feasible structures serve as structures of Petri net models, where the special places and the possible actions of AGV-s in P-graph correspond to the places and transitions in Petri net, respectively.

The Petri net model of the AGV system with given delivery requests and their designated transportation pathways, defined by the initial state is shown in Fig. 5. In the examined case study, two AGVs are in the test environment at the same time. In the initial state one of the two AGVs is in position T12, and it delivers a package from pick up station (Wh) to delivery station (Delivery) through the pathway T11, B1, G31, G32, T21, T22. The other AGV starts from place T22 and travels on the route T21, G22, G21, S2, S1. The pathways assigned to the AGVs are shown in Fig. 6, where route of the first AGV is marked in red and route of the second AGV is marked in blue. The highlighted points in the environment affected by both AGVs are shown in purple in the figure.

In CPN model, the preconditions and consequences of AGV actions described with arc inscriptions and guards, the capacity limits of the places represented by anti-places.

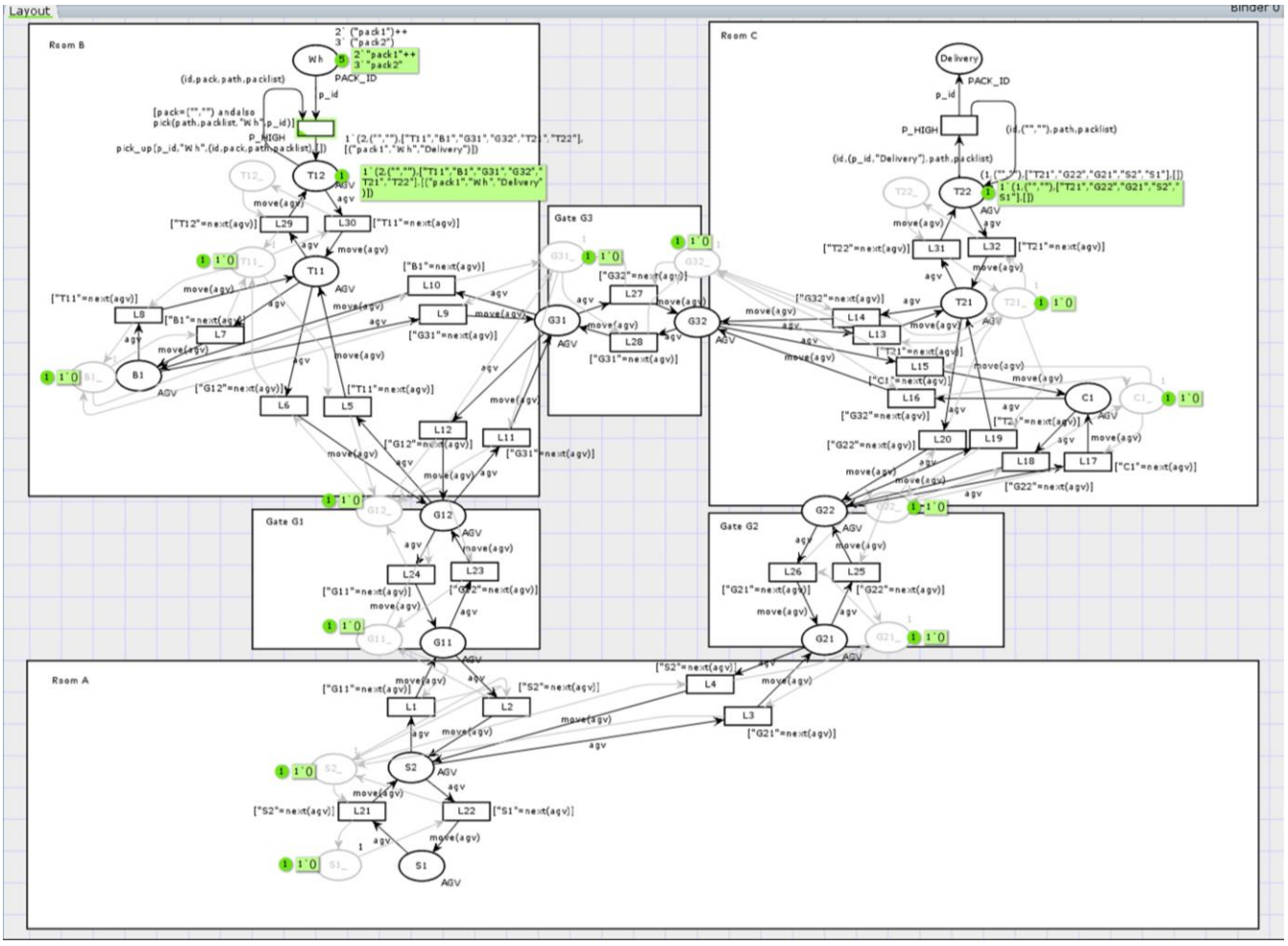


Fig. 5. The CPN model of AGV system with initial state.

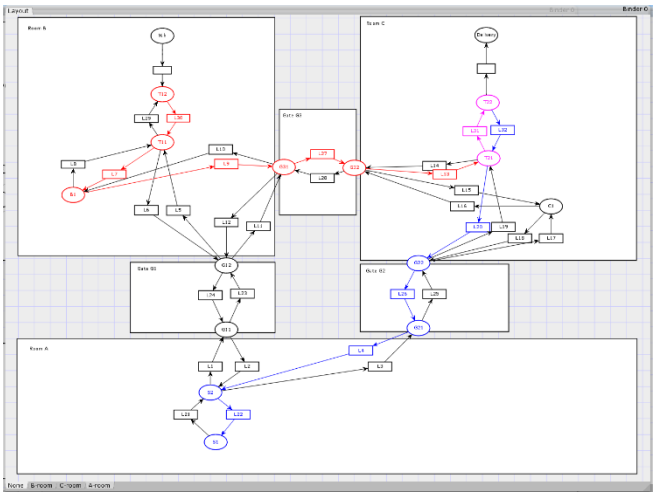


Fig. 6. The designated pathways of two AGVs in Fig. 5.

Besides simulation of AGV system, important properties, like reachability and deadlock are investigated by reachability graph analysis [8] with State Space Tool integrated into CPN Tools.

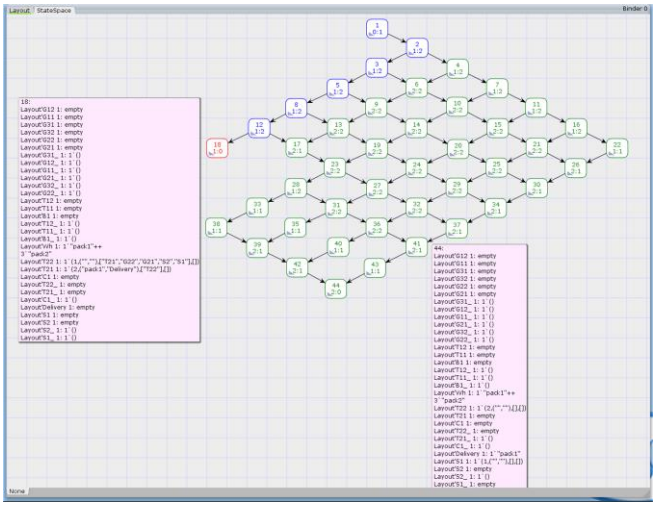


Fig. 7. The reachability graph of CPN model in Fig. 5.

The reachability graph of the system under study is shown in Fig. 7, where the states represented by rectangles. State 1 corresponds to the initial state, and states 18 and 44 are deadlocks. In state 44, one of the AGVs is located at S1 (filling station) and the other is at T22 (delivery station). The additional route to be traversed by both AGVs are empty, i.e. they have traversed the route assigned to them and transported the designated packages. The resulting deadlock is the expected final state of the system.

In state 18, one of the AGVs is at T22 (delivery station) and the other is at T21 (place right next to delivery station). The state is a deadlock, which is the unexpected final state of the system.

According to the expected and unexpected final states of the system, the partitions of the nodes of the reachability graph are as follows:

- nodes from which directed path leads only to expected final states (green nodes)
- nodes from which directed path leads only to unexpected final states (red nodes)
- nodes from which a directed path leads to both expected and unexpected final states (blue nodes)

Partitions are characterized by the fact that if we are in one of the states of the partition marked in green, we will definitely remain in the green partition even after performing any further transition. The same applies to the partition marked in red. Accordingly, the states of the red partition are to be avoided by the system, and the states of the green partition are the states to be reached by the system.

In the case study on Fig. 8, two AGVs move on two different, intersecting routes in the study environment. In the initial state, one AGV is at location S1 (filling station) and the other is at location T22 (delivery station), where their designated pathways are as follows: S1, S2, G11, G12, T11, T12, T11, B1, G31, G32, T21, T22, T21, G22, G21, S2, S1 and T22, T21, G32, G31, B1, T11, T12, T11, B1, G31, G32, T21, T22, T21, G22, G21, S2, S1. The designated route of the first AGV is from the filling station, it touches the pick-up station and the delivery station back to the filling station, while a package has to be transported from pick-up station (Wh) to delivery station (Delivery). The route assigned to the other AGV leads from the delivery station to the pick-up point and then through the delivery point to the filling station, while it has to transport a packet from Wh to Delivery.

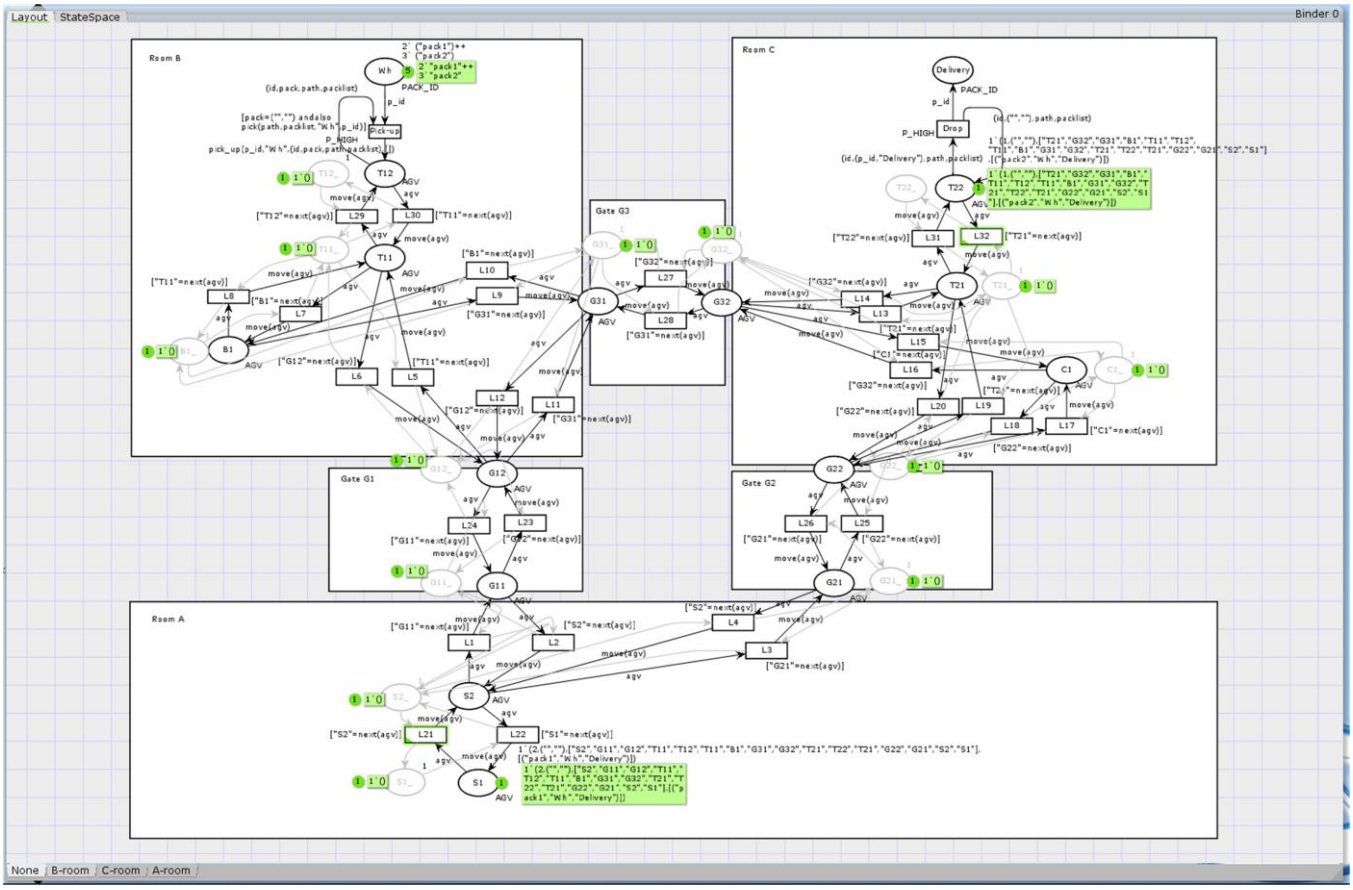


Fig. 8. The CPN model of AGV system with another initial state.

The designated pathways of the AGVs are shown in Fig. 9, where the places affected only by the first and the second AGV is marked in blue and red, respectively. The places in the environment affected by both AGVs are shown in purple.

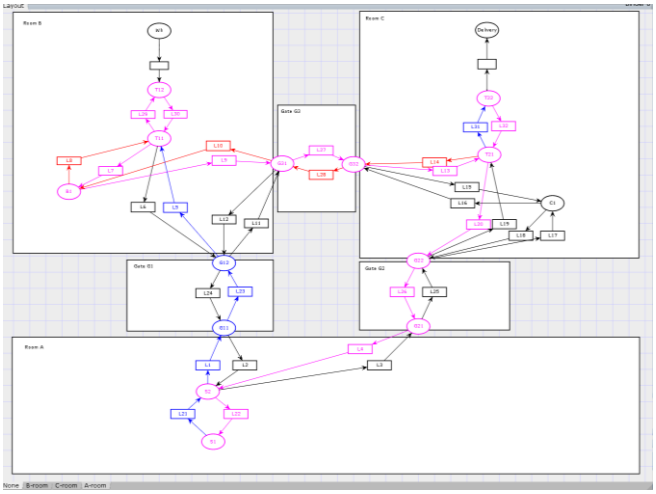


Fig. 9. The designated pathways of two AGVs on Fig. 8.

The reachabilty graph for the initial state of the examined system is shown in Fig. 10. It can be seen that the system has nine deadlocks, of which one is a expected final state (state number 205) and eight are unexpected final states. It can be seen from the partitions of the reachability graph that unexpected final states 68 and 172 can be easily avoided. However, before reaching the other six unexpected end states (states 69-74), we can enter a red partition 4-8 steps earlier, from where we arrive at an unexpected final state regardless of the execution order of the subsequent transitions. The "top" state of the red partition at the top right of the figure, i.e. state number 15, is shown in Fig. 10, from which the expected final state is certainly not reached, but the system reaches a deadlock only after seven steps.

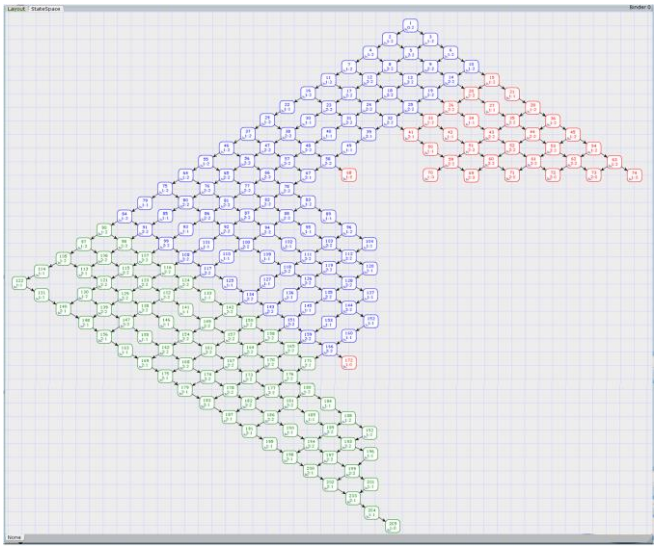


Fig. 10. The reachability graph of CPN model in Fig. 8.

CONCLUSIONS

The algorithmic synthesis of Petri nets by the P-graph framework is used for route planning of automated guided vehicle (AGV) systems. The resulted colored Petri net model appropriate to the simulation and formal analysis of the system. Analysis of the reachability graph have shown that in case of planning AGVs routes, it can be critical to identify and avoid deadlocks. Therefore, important to identify unexpected system states at an early stage and determine if there is a sequence of steps (and if so, which ones) that the AGVs can reach the expected final states.

REFERENCES

[1] F. Friedler, K. Tarjan, Y. W. Huang and L. T. Fan, "Graph-Theoretic Approach to Process Synthesis: Axioms and Theorems", Chem. Eng. Sci. 47, pp. 1973-1988, 1992

[2] F. Friedler, K. Tarjan, Y. W. Huang and L. T. Fan, "Combinatorial Algorithms for Process Synthesis", Comp. Chem. Eng. 16, pp. 313-320, 1992.

[3] Software P-Graph Studio, www.p-graph.com, Last accessed 01/09/2020.

[4] Barany, M., Bertok, B., Kovacs, Z., Friedler, F., & Fan, L. T. "Solving vehicle assignment problems by process-network synthesis to minimize cost and environmental impact of transportation", *Clean Technologies and Environmental Policy*, *13*(4), 637-642, 2011.

[5] J. L. Peterson, "Petri Net Theory and the Modeling of Systems", Prentice Hall, New Jersey. 1981.

[6] CPN Tools: A tool for editing, simulating, and analyzing Colored Petri nets, cpntools.org, Last accessed 01/09/2020.

[7] R. Lakner, F. Friedler, B. Bertok, "Synthesis and Analysis of Process Networks by Joint Application of P-graphs and Petri Nets", Lecture Notes in Artificial Intelligence, 10258, pp. 309-329, 2017.

[8] T. Murata, "Petri Nets: properties, analysis and applications", In: Proc IEEE, vol. 77. pp. 541-80. 1989.

Appearance of End User Needs to Increase Energy Efficiency and Product Reliability

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

***Abstract* — The aim of the study - based on domestic and foreign sources - is to show the development possibilities of technical equipment which used in industrial and civil areas in terms of energy efficiency. During the introduction of the topic, this paper will introduce the emerging needs of end users and then present the results of a questionnaire-based needs survey. In a separate subsection, the paper focus on the user-controllable and modifiable modes in a particular operation of technical equipment and the possibilities of cost-effective implementation. After summarizing the results at the end of the study, the paper will discuss the monitoring possibilities of a centralized energy consumption system.**

I. INTRODUCTION

The globalized economy of the 21st century is flooding the world with countless cheap products. Many of these products have the property of ever-decreasing quality and longevity. For many, this high level of consumption is meaningless as the resources of our planet are finite. Somewhere we all feel deep inside that this is far from correct. Behind the luxurious way of life of the present society lies a vast mound of garbage which discarded prematurely.

What is the reason that once expensive status symbols end up in a junkyard due to their failure? What is the reason that a new, unused raw materials end up in the garbage dump? How much raw material and energy did they cost to produce. Questions can also be answered in another form of a new question: How long can this wasteful practice continue? It is time to rethink the general approach to modern energy use.

It is very difficult to change the current economic operation, but proposing options for an optional use of resources could be very useful. Many users would be happy to give up the performance of the device if it would increase the life of the equipment they use. There are few appliances with this option - none of the household appliances - although there is a market demand for it and it could be provided at low cost.

II. ENERGY EFFICIENCY

How do we know how energy efficient is a device, a household or a country? The energy intensity of a country's economy is often used as an indicator of energy efficiency - mainly because, at an aggregate level, this indicator is relatively easily available for evaluating and comparing countries. However, in a country with lower energy intensity, high energy efficiency is not necessarily required. For example, a small, service-based country with a mild climate would have a lower intensity than a large industrial-based country with a cold climate, even if the latter country uses energy more efficiently. Similarly, trends towards lower intensity are not necessarily driven by efficiency gains. Therefore, it is important to perform a more detailed analysis that provides insight into the factors influencing final energy consumption trends.

In 2009, the IEA (International Energy Agency) recognized the need for better monitoring of energy efficiency policies. This includes a country-specific analysis of end-uses in the largest sectors: construction, services, industry and transport. Energy efficiency policies are essential to achieve key energy policy goals, such as reducing energy bills, tackling climate change and air pollution, improving energy balance security and increasing energy efficiency. Nevertheless, global political coverage (~ 35%) leaves many opportunities untapped.

Reliable data and indicators on energy efficiency are key to learning about and monitoring the effectiveness of energy efficiency policies, as they show energy demand.

The IEA's statistical analyses show how the final energy consumption of IEA member states is evolving, and the development directions and implementation efficiencies of national energy efficiency policies can also be monitored. [1] The planned and agreed 3% annual efficiency increase has not yet been achieved.

III. CHANGES IN THE ENERGY EFFICIENCY

The declining momentum of global energy efficiency developments is a matter of serious concern. The IEA Energy Efficiency 2019 report examines the reasons for the slow-down, which has serious consequences for consumers, businesses, governments and the environment. [2]. Since 2015, the improvement in global energy intensity has been weakening every year. Demands for heating, cooling, lighting, mobility and other energy services are constantly increasing. The improvement in the energy intensity of the global economy (the amount of energy used per unit of economic activity) is slowing down. [3] The 1.2% improvement in 2018 was around half the average observed since 2010 (Figure 1). The indicator is well below the desired average of 3%. This reflects the relative lack of new energy efficiency policies and the need to strengthen existing measures. [4]

IV. RESULTS

The present study examines household and industrial electrical appliances. Although the presented results do not bring about direct global change, but they show the validity of the research with real data. In this way, it would act as a basis for initiating real improvements.

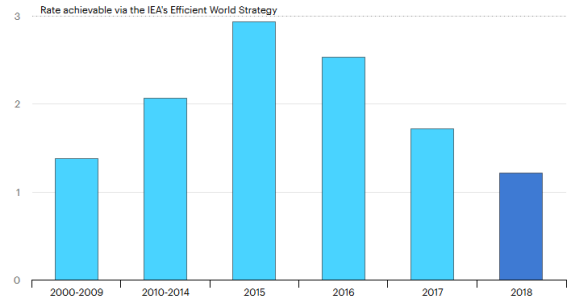


Figure 1. Global improvement in primary energy intensity, 2000-2018 [5]

The respondents to the questionnaires serving as the basis of the research are typically women and men, aged 18-27, who have participated in technical education, or participated before, who are working in technical jobs or have a technical interest. This is the group that will have the greatest influence and value on the selection and / or development of the technical equipment to be procured.
In the questionnaire, respondents were able to answer the questions with integer numbers on a scale from 1 to 5. 1 had the meaning "not important" and 5 had the meaning "very important". There was a gradual transition between the two values.
Respondents were rated for household appliances, consumer electronics, industrial electrical equipment, and electronic systems found in critical infrastructures. The first column (orange) of each category in Figure 2 shows how important respondents con-sider the reliability of electronic systems in a given area; the second column (blue) shows how reliable electronic systems are in the area according to respondents; the third column (lemon yellow) shows how important it is for respondents to increase the reliability of electronic systems in a given field.

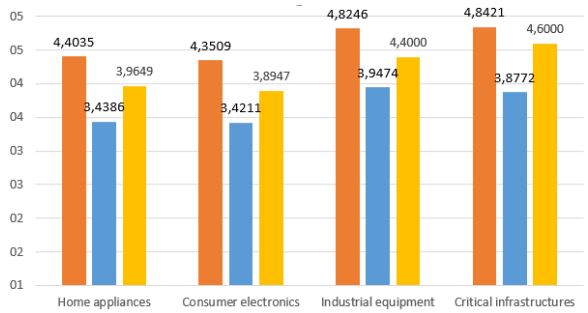


Figure 2. Reliability of electronic systems

The data clearly show that users, in both civil and industrial equipment, need more reliable electronic equipment and systems. This would also be reflected in the solvent de-mand shown in the figure below. End-users would also undertake a price increase of ~ 20-30% in the studied areas if they could use more reliable devices. (It was possible to answer the questionnaire with 10% accuracy.)

Increasing the reliability of civil and industrial electronic equipment is possible at minimal cost. For example: installation of higher quality components, making design changes, changing design and testing principles, installation of minimal cost hardware and software modules, etc. The discussion of long-term sales numbers and sales strategies is not the purpose of this article, but clearly visible the possibility of developing a new optimum that can greatly reduce the environmental impact.

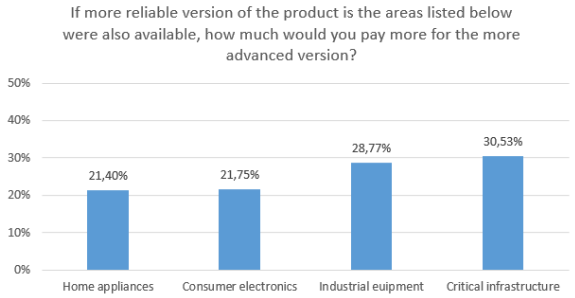


Figure 3. Willingness to bear the extra cost of more reliable electronic equipment

It will be appreciated that further scaling of different manufacturers, different product families, different products, different versions and vintages is not necessarily a viable path in terms of reliability (including manufacturer software, possibly licensed options). It is an option to give the end user the ability to fine-tune the product.

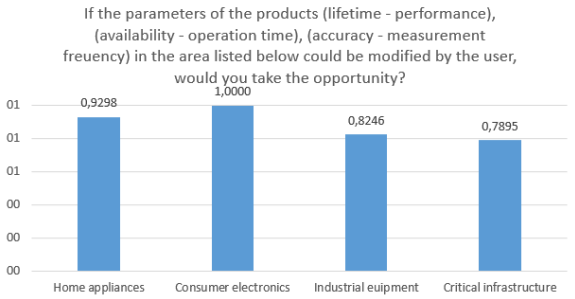


Figure 4. Users' demand for scalable electronic equipment

Figures 4 and 5 show that this would be in high demand by users (yes or no answer for those who completed the questionnaire), especially in civilian areas. Due to their quantity, the consumption of raw materials and environmental impact of civil electronic equipment is much higher than that of industrial equipment.

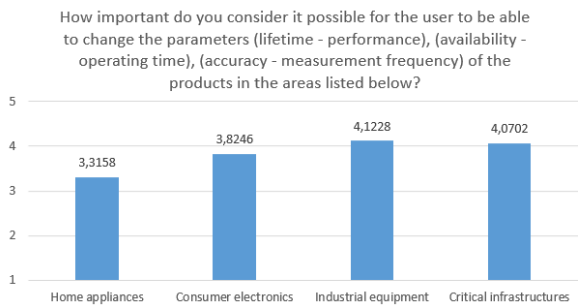


Figure 5. User scalable electronic devices

In the case of civil electronic equipment, the issue of lifetime (in terms of the use of raw materials and resources and the generation of hazardous waste) is also important.

In the case of industrial electronic equipment, performance (in terms of operational reliability), downtime, service time and other reliability parameters are the main indicators of performance.

For island-powered equipment, the user can decide on the balance between availability, uptime, and performance. The power consumption (wake time, calculation time and power) as well as the accuracy and frequency of the measurements also offer adjustment options. Software / firmwares can be used to automate - according to user settings - energy consumption, reduce the frequency of measurements (reducing wake time) and the accuracy of measurements (reducing measurement and calculation time) if the measured value does not change or only changes very slowly. If the rate of change of the measured parameter increases, the measurement frequency and measurement accuracy also follow it. Figure 6 shows that solvent demand would also incur an additional cost of ~ 20-30% to have the option.

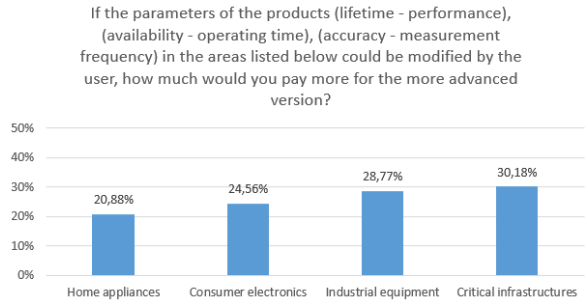


Figure 6. Willingness to bear the extra cost of user-scalable electronic equipment

V. MODULAR STRUCTURE

The ever-expanding reservoir and declining service life of our technical tools used in everyday life means more and more waste generation and resource use. The expanding technical knowledge of users and their willingness to intervene in technical equipment is increasing. With the help of the Internet, the information needed for repair is also widely and easily accessible.

It is important that only users with the right technical training modify electrical equipment, and it is important to increase the number of users with the right technical training. It is recommended to make the acquisition of basic technical knowledge an important part of general secondary education, with the aim of enabling end-users to prevent or eliminate everyday technical problems on their own.

Modular electronic equipment can be repaired in the event of a breakdown with even less resource usage and waste generation at the cost of some extra volume, hardware, software and engineering.

The following example is a model of an island-powered modular power supply system. Figure 7 shows a block diagram of the model. The modules can be easily fixed with screws or flexible lugs, and the releasable galvanic connection between them can be established with ribbon cables. With board-to-board connectors, the modules can be fastened together, and the electrical connection can be done with one device.

The photovoltaic module provides the power supply for island operation. For optimal efficiency, the battery charging module provides impedance matching for the solar panel. The purpose of the battery charging module is to properly charge and discharge the battery, to overcharge, to deep discharge, to charge current, to load current, and so on. supervision. Power and battery charging can also be supplied from an external power source. The DC / DC converter module provides the stable DC voltage required for the load.

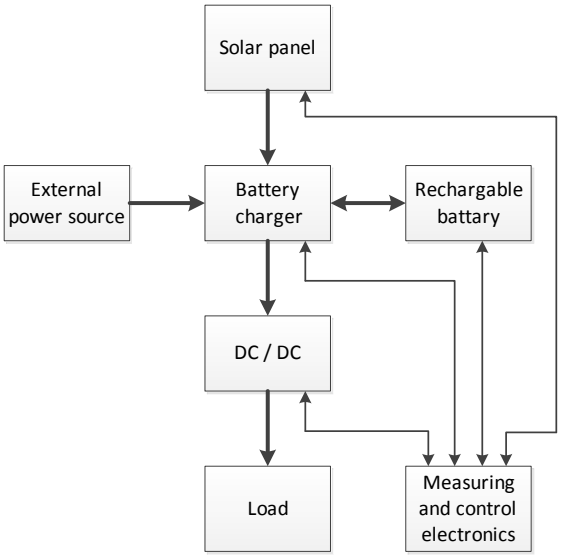


Figure 7. Block diagram of a modular off-grid power supply

The measuring and control electronics monitor the operating modes of each module and monitor their status and store the measured values. The modules can be measured easily with a voltage and current measuring circuit placed in front of and behind them, with minimal component costs. From the measured data, the input and output power for the module can be calculated, and from these, an efficiency for the module can be determined. The modules age and wear out over time. The deterioration of the effects and its speed can be determined from the stored data, and the failure can be predicted for the user with a simple algorithm - with a good approximation. "Appropriate / in need of replacement / inoperative" type feedback is sufficient for the user.

The measuring and control electronics are in practice a microcontroller and hybrid elements built around it, in practice a few euro cents can be realized. Most products have a microcontroller, the firmware / BIOS expansion of the device control and the addition of passive electronics are enough to achieve the new functions.

The control or regulation tasks of the control module could be easily changed by the user within the allowable limits. The user can connect to the device wirelessly or use the submenu of the pre-designed interface on the device.

The device may allow the user to adjust the power-life balance, for example, on a 5-position scale. In the case of a battery (such as a powerbank), in practice, this means that if the user chooses power, it can mean higher charging current (faster charging), changed charging diagram, higher charging voltage, lower deep discharge voltage, higher maximum load current. As a result, a larger amount of charge will be stored in the battery and the electrochemical batteries will also be subjected to a

higher load. The price of higher performance, reduced battery life. If the balance tilts towards service life, settings in the opposite direction to the above will be entered into the control system.

In the case of electronic energy converters (eg battery charging electronics, DC / DC converter, etc.), common points of failure are power semiconductor switching elements, high frequency loaded diodes, inductors, filter capacitors. The higher load current means a higher load on said components. The channel resistance (saturation) of MOSFETs increases, resulting in greater heat dissipation, which accelerates component aging. Electrolytic capacitors also dry out under the influence of heat, their internal equivalent resistance increases. These effects lead to a decrease in the efficiency of the module and an increase in energy consumption.

It is also possible to weight the power-efficiency-life values in the case of electric motor control. The operating current for a given speed can be maximized, and the current peaks that occur when the starting current or speed changes, can be reduced by reducing the rate of change of speed or the maximum value of pulsed currents that can be absorbed by the motor. As a result, the mechanical and thermal load and energy consumption of the electric motor are reduced. In the case of a commutator motor, the service life of the commutator and the carbon brushes can be extended (for example, in the case of universal motors also found in a washer or dryer). The motor control electronics also include semiconductor switches, the service life of which can be extended under lower loads.

When controlling heating elements, we can follow a similar line of reasoning. When the heaters are turned on, soft start can be greatly extended by software-controlled or hardware-implemented softstart.

[12 - 25]

VI. CENTRALIZED CONTROL OF ELECTRICITY CONSUMPTION

The ~ 30% of the world's total electricity consumption is for households, ~ 25% for commercial and public services, and ~ 40% for industrial operators. Reducing the cost of renewable energy and developing digital technologies open up enormous opportunities while creating new energy security dilemmas. Electricity from wind and solar energy will provide more than half of the additional electricity generation by 2040 in the established energy policy scenario and an almost complete increase in the sustainable development scenario.

Policy makers and regulators need to move fast to keep pace with the pace of technological change and the growing need for flexible operation of energy systems. Issues such as storage market planning, the interface between electric vehicles and the grid, and data protection can all expose consumers to new risks.

The task is further complicated by the uneven use of electricity within the 24-hour interval.

The power consumption and dynamics of end-user equipment could be modified by the electricity supplier, if justified. [8], [9] The use of a wireless IoT technology may be a good choice for the communication channel. [10], [11] The cost of equipment for the storage of electricity produced from the planned high share of renewable energy sources could be reduced by this solution. The extra cost of hardware and engineering for the product required could be many times higher, resulting in significant resources emerging globally.

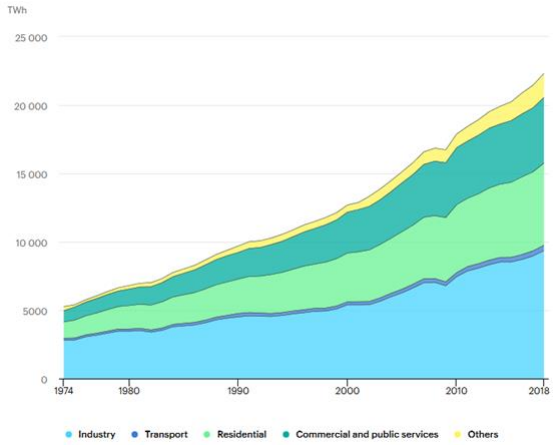


Figure 8. World electricity consumption by sector, 1974–2018 [6]

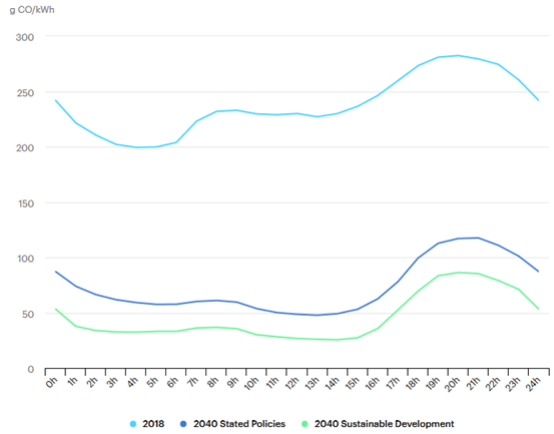


Figure 9. Intensity of average CO2 emissions from 24-hour electricity supply in the European Un-ion, 2018-2040 [7]

CONCLUSION

The authors consider the most important result of this study to be that, based on the presented research, the generation of technical education, which will soon leave technical education and is expected to work for about 45 active years, is definitely important and requires a reasonable surcharge, for the possibility of adjusting the performance-efficiency-lifetime ratios by the end user.

The authors are convinced that the addition of the outlined system component has a relevant market, economic and environmental impact.

REFERENCES

[1] IEA (2019), „World Energy Outlook 2019”, IEA, Paris https://www.iea.org/reports/world-energy-outlook-2019 (downloaded: 2020.08.05.)

[2] IEA (2019), „Energy Efficiency 2019”, IEA, Paris https://www.iea.org/reports/energy-efficiency-2019 (downloaded: 2020.08.05.)

[3] IEA (2020), „Global Energy Review 2019”, IEA, Paris https://www.iea.org/reports/global-energy-review-2019 (downloaded: 2020.08.05.)

[4] IEA (2020), „Energy Efficiency Indicators 2020”, IEA, Paris https://www.iea.org/reports/energy-efficiency-indicators-2020 (downloaded: 2020.08.05.)

[5] IEA, „Global improvements in primary energy intensity, 2000-2018”, IEA, Paris https://www.iea.org/data-and-statistics/charts/global-improvements-in-primary-energy-intensity-2000-2018 (downloaded: 2020.08.05.)

[6] IEA, „World electricity final consumption by sector, 1974-2018”, IEA, Paris https://www.iea.org/data-and-statistics/charts/world-electricity-final-consumption-by-sector-1974-2018 (downloaded: 2020.08.05.)

[7] IEA, „Average CO2 emissions intensity of hourly electricity supply in the European Union, 2018, and by scenario, 2040”, IEA, Paris https://www.iea.org/data-and-statistics/charts/average-co2-emissions-intensity-of-hourly-electricity-supply-in-the-european-union-2018-and-by-scenario-2040 (downloaded: 2020.08.05.)

[8] A. Szűts, „Developing a Complex Decision-Making Framework for Evaluating the Energy-Efficiency of Residential Property Investments”. ACTA POLYTECHNICA HUNGARICA 12 : 6 pp. 231-248. , 18 p. (2015)

[9] I. Czibere, I. Kovách, G. B. Megyesi. „Environmental Citizenship and Energy Efficiency in Four European Countries (Italy, The Netherlands, Switzerland and Hungary)”. SUSTAINABILITY 12 : 3 Paper: 1154 , 18 p. (2020)

[10] A. Szűts, I. Krómer: „Estimating Hungarian Household Energy Consumption Using Artificial Neural Networks”, Acta Polytechnica Hungarica, Vol. 11, No. 4,pp. 155-168,2014

[11] A. Szűts, I. Krómer: „Developing a Fuzzy Analytic Hierarchy Process for Choosing the Energetically Optimal Solution at the Early Design Phase of a Building”, Acta Polytechnica Hungarica, Vol. 12, No. 3, pp. 25-39, 2015

[12] GYÖRÖK, György, et al. Multiprocessor application in embedded control system. In: 2012 IEEE 10th Jubilee International Symposium on Intelligent Systems and Informatics. IEEE, 2012. p. 305-309.

[13] Gergely, G., Menyhard, M., Sulyok, A., Orosz, G. T., Lesiak, B., Jablonski, A., ... & Varga, D. (2004). Surface excitation of selected conducting polymers studied by elastic peak electron spectroscopy (EPES) and reflection electron energy loss spectroscopy (REELS). Surface and Interface Analysis: An International Journal devoted to the development and application of techniques for the analysis of surfaces, interfaces and thin films, 36(8), 1056-1059.

[14] Gergely, G., Menyhard, M., Orosz, G. T., Lesiak, B., Kosinski, A., Jablonski, A., ... & Varga, D. (2006). Surface excitation correction of the inelastic mean free path in selected conducting polymers. Applied surface science, 252(14), 4982-4989.

[15] Orosz, G. T., Sulyok, A., Gergely, G., Gurbán, S., & Menyhard, M. (2003). Calculation of the surface excitation parameter for Si and Ge from measured electron backscattered spectra by means of a Monte-Carlo simulation. Microscopy and Microanalysis, 9(4), 343-348.

[16] G. Györök and L. Simon, “Programozható analóg áramkör megszakításos alkalmazása mikrovezérlő környezetben,” in Alkalmazott informatika és határterületei, Székesfehérvár, Magyarország: Óbudai Egyetem, 2009, p. 1–4.

[17] G. Györök and L. József, “The effect of cyclic thermal stresses in the life of PET shrink’s heating body,” in 8th IEEE International Symposium on Applied Machine Intelligence and Informatics (SAMI 2010), A. Szakál, Ed. Herlany, Szlovákia, 2010, p. 1–4.

[18] G. Györök and L. József, “A Special Case of Electronic Power Control of Induction Heating Equipment,” in Acta Polytechnica Hungarica, vol. 11, no. 5, 2014, pp. 235–246.

[19] Ladislav Fozo, Rudolf Andoga and Radovan Kovacs, "Thermo-dynamic cycle computation of a micro turbojet engine," in: CINTI 2016. - Danvers : IEEE, 2016 P. 000075-000079. - ISBN 978-1-5090-3909-8

[20] Rudolf Andoga, Ladislav Madarász, Ladislav Főző, Tobiáš Lazar and Vladimír Gašpa, "Innovative approaches in modeling, control and diagnostics of small turbojet engines," in: Acta Polytechnica Hungarica. Vol. 10, no. 5 (2013), p. 81-99. - ISSN 1785-8860

[21] Maroš Komjáty, Ladislav Főző and Rudolf Andoga, "Experimental identification of a small turbojet engine with variable exhaust nozzle," in: CINTI 2015. - Danvers : IEEE, 2015 P. 65-69. - ISBN 978-1-4673-8519-0

[22] György Györök, Számítógép perifériák, Óbudai Egyetem, OE AREK 8003 ISBN 978 615 5018 57 2, Budapest, 2013.

[23] 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.

[24] R. Hemza, G. Péter, K. T. József, és S. Károly, „The Effects of Simultaneous Noise and Missing Information in Fixed Point Iteration-based Adaptive Control”, in Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, 2019, o. 1430–1435.

[25] I. K. Artúr, E. Renáta, H. Tamás, S. Károly, és G. Péter, „Optical flow-based segmentation of moving objects for mobile robot navigation using pre-trained deep learning models”, in Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), Bari, Italy, 2019, o. 3060–3066.

Calculating the parameters of 2D Helmert transformation by robust method

G. Nagy*
*Óbuda University, Alba Regia Technical Faculty, Institute of Geoinformatics
nagy.gabor@amk.uni-obuda.hu

***Abstract*—Helmert transformation is an essential tool of the geodetic calculations. In most cases we use less squares method to calculate the parameters of the transformation, but this way is sensitive when the control points contains outlier data, for example a point has wrong coordinates. This paper describes a robust method to calculate the parameters of Helmert transformation, and analyze the result, compare with the less squares method.**

***Index Terms*—Helmert transformation, coordinate transformation, 2D transformation**

I. Introduction

Many coordinate systems are used in the geodesy practice, these systems may be standardized grids or local coordinate systems. When we would like to make a connection between two coordinate system, we need functions which provides the coordinates from the coordinates of the another system.

There are many functions for this task, one of the simplest is the linear connection, which is called affine transformation. The equation of the affine transformation, from system A to system B:

$$\begin{aligned} x_1^B &= t_1^{AB} + l_{11}^{AB} x_1^A + l_{12}^{AB} x_2^A \\ x_2^B &= t_2^{AB} + l_{21}^{AB} x_1^A + l_{22}^{AB} x_2^A \end{aligned} \tag{1}$$

The equations can be written by vectors and matrices:

$$\mathbf{x}_B = \mathbf{t}_{AB} + \mathbf{L}_{AB}\mathbf{x}_A \tag{2}$$

Where $\mathbf{t}$ is the vector of translation, and $\mathbf{L}$ is an arbitrary matrix ($|\mathbf{L}| \neq 0$) of the linear transformation.

The coordinate transformation can be calculated by one matrix multiplication:

$$\begin{bmatrix} x_1^B \\ x_2^B \\ 1 \end{bmatrix} = \begin{bmatrix} l_{11}^{AB} & l_{12}^{AB} & t_1^{AB} \\ l_{21}^{AB} & l_{22}^{AB} & t_2^{AB} \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1^A \\ x_2^A \\ 1 \end{bmatrix} \tag{3}$$

A. Helmert transformation

The Helmert transformation is a special case of the affine transformation, when $\mathbf{L} = \mu\mathbf{R}$, where $\mathbf{R}$ is a rotation matrix, and μ is a scale factor. In this case $l_{11} = l_{22} = a = \mu\cos\alpha$ and $l_{12} = -l_{21} = b = \mu\sin\alpha$, where a and b is parameters of the Helmert transformation (with the vector of translation), and α is the rotation angle.

The Helmert transformation by one matrix multiplication:

$$\begin{bmatrix} x_1^B \\ x_2^B \\ 1 \end{bmatrix} = \begin{bmatrix} a & b & t_1^{AB} \\ -b & a & t_2^{AB} \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1^A \\ x_2^A \\ 1 \end{bmatrix} \tag{4}$$

The scale factor and the rotation angle can be calculated from a and b linear parameters:

$$\begin{aligned} \mu &= \sqrt{a^2 + b^2} \\ \alpha &= \arccos\left(\tfrac{a}{\mu}\right) = \arcsin\left(\tfrac{b}{\mu}\right) = atan2\,(b, a) \end{aligned} \tag{5}$$

The coordinates and the parameters may be handled as complex numbers:

$$\left(x_1^B + x_2^B i\right) = \left(t_1^{AB} + t_2^{AB} i\right) + \left(x_1^A + x_2^A i\right) \cdot (a - bi) \tag{6}$$

or shortly:

$$\mathbf{x}^B = \mathbf{t}^{AB} + (a - bi)\,\mathbf{x}^A \tag{7}$$

If we have $\mathbf{c}$ and $\mathbf{d}$ points in A and B coordinate system, the equations are:

$$\mathbf{c}^B = \mathbf{t}^{AB} + (a - bi)\,\mathbf{c}^A \tag{8}$$

$$\mathbf{d}^B = \mathbf{t}^{AB} + (a - bi)\,\mathbf{d}^A \tag{9}$$

The $\mathbf{t}^{AB}$ parameter can be eliminate, if we subtract (9) from (8):

$$\mathbf{d}^B - \mathbf{c}^B = \left(\mathbf{d}^A - \mathbf{c}^A\right) \cdot (a - bi)$$

and the a and b parameters of the Helmert transformation can be calculated by a complex expression:

$$(a - bi) = \frac{\mathbf{d}^B - \mathbf{c}^B}{\mathbf{d}^A - \mathbf{c}^A} \tag{10}$$

B. Calculate parameters

The affine transformation has 6 (two elements of the vector of the translation, and four elements of the matrix of the linear transformation), and the Helmert transformation has 4 parameters (two elements of the vector of the translation, and two independent values – a and b – of the matrix of the linear transformation). These parameters can be calculated if we know some points coordinates in both system, which are

called control points in the following. The equations of (1) may be elements of a system of linear equations in each point, where the variables are the parameters of the transformation. (The coordinates are known in both systems.) Each control point provides two equations.

The affine transformation needs minimum 3 control points, because these three control points provides $3 \cdot 2 = 6$ equations for the 6 parameters of the transformation. The Helmert transformation needs minimum 2 control points, because these two control points provides $2 \cdot 2 = 4$ equations for the 4 parameters of the transformation. If we have more point than the minimum requirement, the system of equations will be over-determined. The most popular method for solving these over-determined systems of equations is the least square method. The least squares method determines the parameters of the transformation so, that the sum of the squares of the errors (the difference between the control point position in the system B and the transformed position from the system A) will be the least.

C. Linear regression by robust methods

Least squares method provides the most likely result, if the errors of the measurements have normal distribution. In practice, the measurements may have outliers, where the error is larger than the expected. We usually qualify a measurement as an outlier if the error is larger than the triple of the standard deviation. The outlier measurements make an bad influence to the result of the least squares method.

The robust methods can calculate good result with outliers. For example, the impact of a wrong measurement is not too remarkable in the median, but the mean may be wronger. There is an set of numbers: 1.012, 1.008, 10.25, 1.002, 0.9998, 10.007, 0.9993, 0.9988, 1.003, 0.9990. The mean of this data set is 2.83419, but the median is 1.0025. The mean without the outlier is 1.0027375 (the median is 1.001). The outlier values (10.25, 10.007) make wrong the mean result, but the median produces a value around the most of the dataset numbers (except the outliers).

There are many robust method for calculating a linear regression. [1], [2] For example RANSAC (Random Sample Consensus) [3], [4], [5] or the Theil-Sen estimator [6]. The SBLR [7] is also such method, which is a multidimensional extension of the Fitting Disc Method [8]. These methods provide a good regression when the dataset contains outlier points.

Any linear regression method may be suitable for calculating the parameters of the affine transformation. The explanatory variables are the coordinates of the system A, and the dependent variables are the coordinates of the system B; both coordinates have a separately linear regression.

II. A robust method for 2D Helmert transformation

The 2D Helmert transformation has four parameters: two translation, one rotation (the angle of the 2D rotation, denoted α) and one scale factor (denoted μ). Instead of α and μ, the equation (4) uses a and b, the element of the matrix of the linear transformation.

The median is a simple and robust method on the one dimensional datasets, but it is not work in multiple dimensions. This paper describes a method, that divides the calculation more steps, and in these steps uses median calculation.

We can calculate the rotation angle (α) and the scale factor (μ) from two control points, denoted $\mathbf{c}$ and $\mathbf{d}$. The first step, the a and b can be calculated as an complex equation (see (10)):

$$a - bi = \frac{\left(d_1^B - c_1^B\right) + \left(d_2^B - c_2^B\right) i}{\left(d_1^A - c_1^A\right) + \left(d_2^A - c_2^A\right) i} \tag{11}$$

The rotation angle (α) and the scale factor (μ) can be calculated by the equation (5). If we have n control points, we can make $\frac{n(n-1)}{2}$ point pairs, and each point pair provides an α and a μ parameters. The final parameters are calculated by weighted median [9]; the weights are the distance between the two control points, if this is less than the median of these distances, or the median distance, if the distance is larger that the median distance.

After we have the final α and μ values (and the $\mathbf{L}$ matrix with a and b), we can calculate the element of the translation vector, by this equation derived from (2):

$$\mathbf{t}_{AB} = \mathbf{x}_B - \mathbf{L}_{AB}\mathbf{x}_A \tag{12}$$

The translation vector can be calculated with each control point. The final values of the elements of the translation vector will be the median of these values.

III. Numerical tests

A. Implementation

I have written a Python module called **helmert** for calculating the parameters of a Helmert transformation from control points by the least square method and the recommended robust method. The source code of this module is available in the Appendix A.

This Python module can be used simply from other Python modules:

```
from helmert import helmert_lsq,helmert_rob
from helmert import trp_helmert

#Helmert transformation
#from A to B coordinate system

#Points in A and B coordsys
ptA=[(1.0,1.0),(2.0,1.0),(1.0,2.0),(2.0,2.0)]
ptB=[(2.5,3.7),(3.3,3.1),(3.1,4.5),(3.9,3.9)]

#Test points in A coordsys
pt2A=[(1.2,1.8),(1.5,1.7)]

#Calculate the parameters of the transformation
tr_lsq=helmert_lsq(ptA, ptB)
tr_rob=helmert_rob(ptA, ptB)
```

```
#Calculate the transformated coordinates
#by different transformation parameters
pt2B_lsq=trp_helmert(pt2A, tr_lsq)
pt2B_rob=trp_helmert(pt2A, tr_rob)
```

The **trp_helmert** function returns the transformed coordinates by an Helmert transformation. The input is in the first (**ptl**) argument, the output is the return value of the function. The coordinates are in (x, y) tuples, and the parameters are in a (dx, dy, a, b) tuple in the second (**trp**) argument.

The **helmert_lsq** and **helmert_rob** functions calculate the parameters of an Helmert transformation from points which are known in both coordinate system. The points are in lists of tuples, the first argument is the points of the initial (A), and the second argument is the points of the aim (B) coordinate system. The return values are the parameters of the Helmert transformation in a tuple (same as in the **trp_helmert** function). The **helmert_lsq** function uses the classical less squares method, and the **helmert_rob** function uses the recommended robust method.

In the following studies I use this module for testing the recommended method.

B. Studying the result

I study the result of the method with random datasets. The initial position is a random place with coordinate between 0 and 1000. The initial transformation is also random with a translation vector coordinates between -1000 and 1000, the μ between 0.999 and 1.001 and the α between $-\pi$ and π. The coordinates of the second system are calculated by the initial transformation, and add a random normal distribution error with 0.1 standard deviation by the **normalvariate** function of the **random** Python module. The outliers has an additional error between 10 and 50 by the **uniform** function of the **random** module.

A program create one million datasets in each studied cases. The number of control points are between 5 and 19, and the number of outliers are between 0 and the third of the control points. The random datasets are generated by the function of **testtool.py** module, whose source code is in the Appendix B.

The transformation was calculated in each dataset, and the next step the test program calculates the mean squared error between the original (without any error) and the transformed positions:

$$err = \sqrt{\frac{\sum \left(\left(x_1^{TR} - x_1^{OR}\right)^2 + \left(x_2^{TR} - x_2^{OR}\right)^2 \right)}{n}} \qquad (13)$$

The outliers are excluded from the sum. The points which are in the calculation of (13) contains only normal distribution error with $\sigma = 0.1$.

The test program counts the number of the datasets where the mean squared error is less than the single, the double, the triple, etc. of the standard deviation of the random error of the points ($\sigma = 0.1$).

Table I
THE RATES OF MEAN ERRORS IN n POINT DATASET WITH o OUTLIERS

n	o	$< \sigma$	$< 2\sigma$	$< 3\sigma$	$< 4\sigma$	$< 5\sigma$
5	0	78.601%	99.997%	>99.999%	>99.999%	>99.999%
5	1	51.596%	85.872%	89.016%	89.662%	89.884%
6	0	78.674%	99.999%	>99.999%	>99.999%	>99.999%
6	1	61.809%	98.072%	99.118%	99.296%	99.353%
6	2	27.657%	50.261%	53.211%	54.048%	54.488%
7	0	74.721%	>99.999%	>99.999%	>99.999%	>99.999%
7	1	62.930%	99.850%	99.989%	99.994%	99.996%
7	2	37.275%	73.938%	77.385%	78.223%	78.603%
8	0	74.949%	>99.999%	>99.999%	>99.999%	>99.999%
8	1	66.425%	99.993%	>99.999%	>99.999%	>99.999%
8	2	46.827%	90.325%	92.605%	93.078%	93.274%
9	0	72.142%	>99.999%	>99.999%	>99.999%	>99.999%
9	1	65.211%	99.998%	>99.999%	>99.999%	>99.999%
9	2	50.987%	97.472%	98.489%	98.651%	98.705%
9	3	31.845%	72.453%	76.124%	77.033%	77.482%
10	0	72.281%	>99.999%	>99.999%	>99.999%	>99.999%
10	1	66.854%	>99.999%	>99.999%	>99.999%	>99.999%
10	2	55.880%	99.569%	99.813%	99.845%	99.856%
10	3	39.433%	86.594%	89.224%	89.827%	90.095%
11	0	70.083%	>99.999%	>99.999%	>99.999%	>99.999%
11	1	65.536%	>99.999%	>99.999%	>99.999%	>99.999%
11	2	57.036%	99.945%	99.989%	99.993%	99.994%
11	3	43.625%	94.583%	96.189%	96.510%	96.642%
12	0	70.360%	>99.999%	>99.999%	>99.999%	>99.999%
12	1	66.396%	>99.999%	>99.999%	>99.999%	>99.999%
12	2	59.761%	99.997%	>99.999%	>99.999%	>99.999%
12	3	48.747%	98.332%	99.016%	99.131%	99.172%
12	4	35.243%	85.370%	88.190%	88.852%	89.169%
13	0	68.657%	>99.999%	>99.999%	>99.999%	>99.999%
13	1	65.120%	>99.999%	>99.999%	>99.999%	>99.999%
13	2	59.501%	>99.999%	>99.999%	>99.999%	>99.999%
13	3	50.667%	99.591%	99.814%	99.843%	99.855%
13	4	39.291%	92.986%	94.886%	95.276%	95.454%
14	0	68.887%	>99.999%	>99.999%	>99.999%	>99.999%
14	1	65.877%	>99.999%	>99.999%	>99.999%	>99.999%
14	2	61.234%	>99.999%	>99.999%	>99.999%	>99.999%
14	3	53.887%	99.931%	99.975%	99.980%	99.982%
14	4	43.993%	97.179%	98.145%	98.325%	98.404%
15	0	67.458%	>99.999%	>99.999%	>99.999%	>99.999%
15	1	64.673%	>99.999%	>99.999%	>99.999%	>99.999%
15	2	60.578%	>99.999%	>99.999%	>99.999%	>99.999%
15	3	54.478%	99.990%	99.998%	99.999%	99.999%
15	4	46.049%	99.025%	99.461%	99.529%	99.556%
15	5	36.144%	92.205%	94.195%	94.622%	94.832%
16	0	67.623%	>99.999%	>99.999%	>99.999%	>99.999%
16	1	65.188%	>99.999%	>99.999%	>99.999%	>99.999%
16	2	61.762%	>99.999%	>99.999%	>99.999%	>99.999%
16	3	56.673%	99.999%	>99.999%	>99.999%	>99.999%
16	4	49.402%	99.731%	99.878%	99.897%	99.903%
16	5	40.485%	96.404%	97.549%	97.766%	97.868%
17	0	66.395%	>99.999%	>99.999%	>99.999%	>99.999%
17	1	64.145%	>99.999%	>99.999%	>99.999%	>99.999%
17	2	61.030%	>99.999%	>99.999%	>99.999%	>99.999%
17	3	56.704%	>99.999%	>99.999%	>99.999%	>99.999%
17	4	50.474%	99.937%	99.975%	99.979%	99.981%
17	5	42.684%	98.490%	99.084%	99.187%	99.234%
18	0	66.570%	>99.999%	>99.999%	>99.999%	>99.999%
18	1	64.655%	>99.999%	>99.999%	>99.999%	>99.999%
18	2	61.858%	>99.999%	>99.999%	>99.999%	>99.999%
18	3	58.086%	>99.999%	>99.999%	>99.999%	>99.999%
18	4	52.755%	99.990%	99.998%	99.999%	99.999%
18	5	46.004%	99.491%	99.728%	99.765%	99.778%
18	6	37.977%	95.931%	97.170%	97.425%	97.543%
19	0	65.575%	>99.999%	>99.999%	>99.999%	>99.999%
19	1	63.774%	>99.999%	>99.999%	>99.999%	>99.999%
19	2	61.242%	>99.999%	>99.999%	>99.999%	>99.999%
19	3	57.785%	>99.999%	>99.999%	>99.999%	>99.999%
19	4	53.222%	99.998%	>99.999%	>99.999%	>99.999%
19	5	47.222%	99.840%	99.932%	99.942%	99.946%
19	6	40.282%	98.110%	98.806%	98.938%	98.998%

The result of the test program is in the I. The first column is the count of all points of the dataset (denoted n in the header), and the second column is the count of the outliers (denoted o in the header). The table contains all cases where $5 \leq n \leq 19$ and $o \leq \frac{n}{3}$.

The test program has created 70 million test cases in total. The process need around one day in the personal computer of the author.

IV. Conclusion

The recommended method can calculate correct result even in that case the dataset of the control points contains outliers. (If the count of outliers is not too many.) The widely used less squares method produces wrong result, if the measurements contains wrong data.

This method may be useful in every cases, when we would like to create a Helmert transformation from a control point set, which may contain outlier points.

An Python 3 module has been created, which contains the implementation of the recommended method (and the less squares method as a reference). The appendixes contains the source code of this module and an helper module of the test programs. The source code can be downloaded from https://github.com/ngabor/helmert2d under GPL3 license. This code may be use in Python 3 programs, or use as a reference with another applications.

References

[1] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*. John Wiley & Sons, 2005, vol. 589.

[2] G. Balkema and P. Embrechts, "Linear regression for heavy tails," *Risks*, vol. 6, no. 3, p. 93, 2018.

[3] S. Choi, T. Kim, and W. Yu, "Performance evaluation of ransac family," *Journal of Computer Vision*, vol. 24, no. 3, pp. 271–300, 1997.

[4] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[5] A. Hast, J. Nysjö, and A. Marchetti, "Optimal ransac-towards a repeatable algorithm for finding the optimal set," 2013.

[6] H. Theil, "A rank-invariant method of linear and polynomial regression analysis," in *Henri Theils Contributions to Economics and Econometrics*. Springer, 1992, pp. 345–381.

[7] G. Nagy, "Sector based linear regression, a new robust method for the multiple linear regression," *ACTA CYBERNETICA*, vol. 23, no. 4, pp. 1017–1038, 2018.

[8] G. Nagy, T. Jancsó, and C. Chen, "The fitting disc method, a new robust algorithm of the point cloud processing," *ACTA POLYTECHNICA HUNGARICA*, vol. 14, no. 6, pp. 59–73, 2017.

[9] F. Edgeworth, "On a new method of reducing observationsrelating to several quantities," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 25, no. 154, pp. 184–191, 1888.

Appendix A

The source code of `helmert.py`:

```
"""Calculate of parametres of Helmert transformation from point pairs"""

from math import sqrt, atan2, sin, cos
from statistics import median
import numpy as np
from numpy.linalg import inv

def trp_helmert(ptl, trp):
    """Calculate Helmert transformation of ptl [(x,y,z), ...]
       by parameters trp (dx, dy, a, b)"""
    return [(trp[0]+pt[0]*trp[2]+pt[1]*trp[3],
             trp[1]-pt[0]*trp[3]+pt[1]*trp[2]) for pt in ptl]

def helmert_lsq(ptl1, ptl2):
    """Calculates parameters from 1 to 2 coordsys, by point lists
       (ptl1 and ptl2 [(x,y,z), ...]), least squeres method"""
    npt=len(ptl1)
    if len(ptl2)!=npt:
        raise ValueError('Different list length (ptl1 and ptl2)')
    prex=(0.0, 0.0, 1.0, 0.0)
    a=np.zeros([npt*2,4], dtype=np.double)
    p=np.identity(npt*2, dtype=np.double)
    l=np.zeros([npt*2,1], dtype=np.double)
    for i in range(npt):
        a[2*i  ][0] = 1.0
        a[2*i  ][1] = 0.0
        a[2*i  ][2] = ptl1[i][0]
        a[2*i  ][3] = ptl1[i][1]
        a[2*i+1][0] = 0.0
        a[2*i+1][1] = 1.0
        a[2*i+1][2] = ptl1[i][1]
        a[2*i+1][3] =-ptl1[i][0]
        l[2*i  ][0] =(prex[0]+prex[2]*ptl1[i][0]+prex[3]*ptl1[i][1]-ptl2[i][0])
        l[2*i+1][0] =(prex[1]-prex[3]*ptl1[i][0]+prex[2]*ptl1[i][1]-ptl2[i][1])
    x=-inv(a.transpose() @ p @ a) @ (a.transpose() @ p @ l)
    return (prex[0]+x[0][0], prex[1]+x[1][0], prex[2]+x[2][0], prex[3]+x[3][0])

def wmed(wml):
    """Wighted median from wml [(weight, vaule), ...]"""
    wmls=sorted(wml, key=lambda element:element[1])
    wmsh=sum(map(lambda element:element[0], wmls))/2.0
    wmsi=0.0
```

```
    for wme in wmls:
        wmsi+=wme[0]
        if wmsi>wmsh:
            return wme[1]

def helmert_rob(ptl1, ptl2):
    """Calculates parameters from 1 to 2 coordsys, by point lists
       (ptl1 and ptl2 [(x,y,z), ...]), robust method"""
    npt=len(ptl1)
    if len(ptl2)!=npt:
        raise ValueError('Different list length (ptl1 and ptl2)')
    dml=[]
    al=[]
    bl=[]
    for ipt in range(npt):
        for jpt in range(ipt):
            c1=complex(ptl1[jpt][0]-ptl1[ipt][0], ptl1[jpt][1]-ptl1[ipt][1])
            c2=complex(ptl2[jpt][0]-ptl2[ipt][0], ptl2[jpt][1]-ptl2[ipt][1])
            cr=(c2/c1).conjugate()
            dml.append( 0.5*(sqrt(c1.real**2+c1.imag**2)+sqrt(c2.real**2+c2.imag**2)) )
            al.append(cr.real)
            bl.append(cr.imag)
    dmed=median(dml)
    wl=[min(dm, 1.5*dmed) for dm in dml]
    a=wmed(zip(wl,al))
    b=wmed(zip(wl,bl))
    dx=median([pp[1][0]-a*pp[0][0]-b*pp[0][1] for pp in zip(ptl1, ptl2)])
    dy=median([pp[1][1]+b*pp[0][0]-a*pp[0][1] for pp in zip(ptl1, ptl2)])
    return (dx, dy, a, b)
```

Appendix B

The source code of `testtool.py`:

```
"""Helper functions for testing transformation methods"""

import random
from math import pi, sin, cos, sqrt

def trcoordlist(ptl1, ptl2, trfunc, olpt=()):
    """Compare the points of ptl1 transformed by trfunc,
and the points of ptl2. Calculate the squared mean with
and without outliers, whose indices are in the olpt.
The result is printes to stdout"""
    trp=trfunc(ptl1, ptl2)
    print('The parameters of the transformation: ',trp)
    sumn=0.0
    sumnall=0.0
    for pp in zip(ptl1, ptl2, [(i in olpt) for i in range(len(ptl1))]):
        x1=pp[0][0]
        y1=pp[0][1]
        x2=pp[1][0]
        y2=pp[1][1]
        x2r=trp[0]+trp[2]*x1+trp[3]*y1
        y2r=trp[1]-trp[3]*x1+trp[2]*y1
        dx=x2-x2r
        dy=y2-y2r
        dr=sqrt(dx**2+dy**2)
        if not pp[2]:
            sumn+=dr**2
        sumnall+=dr**2
        print(f'{x1:9.2f}   {y1:9.2f}   {x2:9.2f}   {y2:9.2f}   {x2r:9.2f}   \
{y2r:9.2f}   {dx:9.2f}   {dy:9.2f}   {dr:9.2f}')
    print(' squared mean: ',sqrt(sumn/(2*(len(ptl1)-len(olpt)))))
    print('with outliers: ',sqrt(sumnall/(2*len(ptl1))))

def randtrp(offs=(-1000.0,-1000.0,1000.0,1000.0),
            scaleiv=(0.999, 1.001), rotiv=(-pi, pi)):
    """Create a random Helmert transformation"""
    scale=random.uniform(scaleiv[0], scaleiv[1])
    rot=random.uniform(rotiv[0], rotiv[1])
    a=scale*cos(rot)
    b=scale*sin(rot)
    dx=random.uniform(offs[0], offs[2])
    dy=random.uniform(offs[1], offs[3])
    return (dx, dy, a, b)
```

```
def randptp(ptn, trp, stderr=0.07071, pt1win=(0.0,0.0,1000.0,1000.0),
            oln=0, oldst=(1.0,10.0)):
    """Create a list of ptn random point in pt1win area (ptl1),
the transformed point by a Helmert transformation of trp (ptl2),
and the random modified points by stderr normal variate error (ptl2e).
oln points will be outlier, whose indices are in olpti.
The function returns (ptl1, ptl2, ptl2e, olpti).
"""
    ptl1=[(random.uniform(pt1win[0], pt1win[2]),
           random.uniform(pt1win[1], pt1win[3])) for i in range(ptn)]
    ptl2e=[(trp[0]+trp[2]*pt[0]+trp[3]*pt[1],
            trp[0]-trp[3]*pt[0]+trp[2]*pt[1]) for pt in ptl1]
    ptl2=[(pt[0]+random.normalvariate(0,stderr),
           pt[1]+random.normalvariate(0,stderr)) for pt in ptl2e]
    olpti=random.sample(range(ptn),oln)
    for oi in olpti:
        ptx,pty=ptl2[oi]
        ptx+=random.uniform(10.0,50.0)*random.choice([-1,1])
        pty+=random.uniform(10.0,50.0)*random.choice([-1,1])
        ptl2[oi]=(ptx, pty)
    return (ptl1, ptl2, ptl2e, olpti)

def trpstat(ptl1, ptl2, olpt=[]):
    """Calculate the squared mean and the mean of the absolute values
of the difference between the points of ptl1 and ptl2 lists.
The calculation skipts points, whose indices are in the olpt.
The function returns (squared mean, absolute mean) in a tuple."""
    nn=len(ptl1)
    n=nn-len(olpt)
    sumr2=0.0
    sumra=0.0
    for ptp in zip(ptl1, ptl2, range(nn)):
        if not ptp[2] in olpt:
            dx=ptp[1][0]-ptp[0][0]
            dy=ptp[1][1]-ptp[0][1]
            dr2=dx**2+dy**2
            sumr2+=dr2
            sumra+=sqrt(dr2)
    return (sqrt(sumr2/n), sumra/n)
```

Automatic Processes in Photogrammetry

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— **Digital Photogrammetric Workstations (DPWs) and especially some software packages serving the processing of UAV images offer automatic algorithms to accelerate the gaining of end products like digital terrain or surface models or orthophotos. To achieve the final steps in the image evaluation, the orientation process, particularly the exterior orientation requires manual measurements of control points. The measurements of tie points necessary for building models from stereo image pairs or finding tie points for strips of images needed for aerial triangulation require automatic image matching. These algorithms are based mainly on finding the same terrain points, lines or other objects on different images using the area-based image matching, semi-global image matching or other object-based image matching techniques. The paper summarizes the recent status and the perspectives of these image processing solutions.**

Keywords— **DPW, UAV, photogrammetry, image matching**

I. Introduction

In digital photogrammetry the orientation and the evaluation process are supported by automatic processes. Fig. 1 summarizes the possible steps for automation.

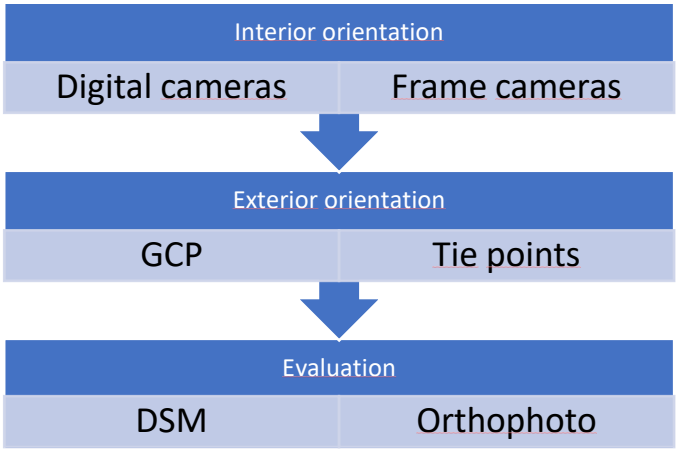


Figure 1. Possible steps for automation

During the interior orientation we build the geometric model of the camera including the distortion model of the projected image. For digital cameras, the interior orientation is an easy step since we need to input only the sensor data including the calibrating focal length, the coordinates of the principle point, the pixel size, the image size (width and height n pixels) and the parameters describing the image distortion.

For frame cameras, this step is more complicated since the photo coordinate system should be determined by the fiducial marks. The measurement of fiducial marks can be automated based on an area-based image matching technique.

The goal of the exterior orientation is the determination of the coordinates of the exposure centre and the calculation of the rotation angles of the camera relatively to the ground coordinate system for each photo. If the IMU/INS (Inertial Navigation System) together with the GNSS equipment is capable to measure the exterior orientation elements with high accuracy during the flight then the exterior orientation doesn't need ground control points (GCPs), which means the direct georeferencing.

The image matching algorithms are developing very fast and they become so sophisticated that the DSM (Digital Surface Model) and orthophoto production can be fully automatic.

All these aspects helped the spread of UAV (Unmanned Aerial Vehicle) technology as well.

II. Flight Logs and Metadata

The IMU together with the GNSS equipment logs the coordinates of the exposure centre and the yaw, pitch, roll rotation angles. The GNSS data can be used directly in the exterior orientation process, but the yaw, pitch, roll angles should be converted to phi, omega and kappa angles used in photogrammetry. The main difference between the two sets of angles is:

- Yaw, pitch, roll angles define the rotation of the camera body with respect to its navigation coordinate system.
- Omega, phi, kappa angles define the rotation between the image coordinate system and a ground coordinate system used in photogrammetry.

The conversion is more less straightforward, but we should consider some aspects [1]:

- the meridian convergence,
- the interpretations of angular range and sign of rotations,
- misalignments ex, ey, ez between INS and camera.

Each image contains several metadata including the time of exposure, the focal length, the aperture value, the GPS coordinates, the yaw, pitch, roll angles. These data can be used as approximate values for the precise interior and exterior orientation. Fig 2 shows a fragment of image metadata captured by an UAV camera.

GPS information:	
GPSVersionID	3.2.0.0
GPSLatitudeRef	N
GPSLatitude	47 13 25.2638 (47.223684)
GPSLongitudeRef	E
GPSLongitude	18 16 10.7072 (18.269641)
GPSAltitudeRef	Sea level
GPSAltitude	282.13 m

Figure 2. GPS information block in image metadata

Based on the calculated S_{avg} and pixel size Δ_p, we can get the GSD value, in other words the ground resolution by the following equation:

III. ORIENTATION

The whole orientation process is divided into two steps: interior orientation and exterior orientation.

The task of the interior orientation is well defined and can be automatized. In case of frame cameras., according to the method of least squares, we try to solve the task of interior orientation with adjustment, like all orientation tasks, by introducing more measurements than minimally necessary. In case of affine transformation, we need to measure at least four fiducial marks. The system of equations based on equations (1) is linear for the unknowns and can be solved directly [3], [4].

$$\begin{aligned} X &= a_0 + a_1 x + a_2 y \\ Y &= b_0 + b_1 x + b_2 y \end{aligned} \quad (1)$$

Notation in (1):
X, Y : photo coordinates
x, y : image (pixel) coordinates
$a_0, a_1, a_2, b_0, b_1, b_2$: transformation parameters

The measurement of fiducial marks can be automatic with the help of the cross correlation between the stored image of the fiducial mark as a pattern and the image segment containing the fiducial mark. This area-based image matching is useful not only for the automatic measurement of fiducial marks, but it can be used also for the measurement of common points (tie points, ground control points for the exterior orientation and DSM points for orthophoto production).

A. *Area-based image matching*

Most programs apply the raster correlation pattern matrix and use the cross-correlation formula (2) [3].

$$\rho = \frac{\sum_{r=1}^{R}\sum_{c=1}^{C}(g_1(r,c)-\mu_1)(g_2(r,c)-\mu_2)}{\sqrt{\sum_{r=1}^{R}\sum_{c=1}^{C}(g_1(r,c)-\mu_1)^2 \sum_{r=1}^{R}\sum_{c=1}^{C}(g_2(r,c)-\mu_2)^2}} \quad (2)$$

Where:
g_1 - the greyscale value of the pixel in the target area
g_2 - The greyscale value of the pixel in the search area
r, c - row, column index
μ_1, μ_2 - Average grey values in the target and search areas
R, C - number of rows and columns in the sample area

The value of the correlation coefficient ρ can vary from -1 to 1, where 1 represents the maximum fitting. If the value of the correlation coefficient reaches a predetermined threshold (typically 0.7), then the fitting is accepted.

Point measurement with autocorrelation does not give a 100% reliable result, so if it is possible and supported by the evaluation program. It is worthwhile to create a correlation image, where the weakly correlated places can be seen, and we know the areas where more ex-post control may be needed. The other main source of problems is the automated evaluation of homogeneous areas (water, arable land, forest, sandy areas, large roads, airports). The correlation procedure is ambiguous at these locations, and it can result errors in elevation data. The masking procedure is commonly used to prevent errors of this nature, i.e., these areas are excluded from the correlation procedure. This method is effective when there are contiguous areas and their height can be specified after a small number of height values. Another way to reduce this type of error is to incorporate subsequent filtering and smoothing procedures.

In general, it can be stated that the subsequent manual verification of the measurements made in this way and the performance of additional measurements are inevitable. When producing terrain models, the result can be dramatically improved if we already have altitude values to describe important terrain details (e.g. mountain peak, road network, embankments) before performing the correlation process.

The method outlined here can be used

- for automated measurement of fiducial marks,
- for measurement of Gruber points for relative orientation,
- for measurement of any field points,
- for automated measurement of DSM points.

B. *Feature-based image matching*

In the area-based image matching we search for the correspondence of common points. We can extend this approach if we are looking for the matching of features like points, lines, curves, or polygons where each feature is described also by a feature descriptor. In this case the matching procedure is done only after the identification of features and feature descriptors in each image. The separation of features and feature descriptors is done by different image processing techniques. Commonly used feature detectors are [4]:

- Förstner operator,
- Sobel operator,
- Laplace operator,
- SUSAN (smallest univalue segment assimilating nucleus) operator,
- FAST (features from accelerated segment test) operator,

- SIFT (scale invariant feature transform) operator,
- Affine invariant SIFT operator,
- SURF (Speed-Up Robust Features) operator.

C. *Semi-global image matching*

The autocorrelation method described in Section III.*A* finds the greatest fit between image details by comparing square or rectangular areas when searching for related points. Such a rectangular area behaves like a low-pass filter, blurring the sharp boundaries of depth differences, resulting in an edge-free elevation model. The smaller the image matrix used for pattern matching, the smaller the blurring of the edges. Ideally, the search matrix would be the size of one pixel, but this is not a feasible solution due to ambiguity. A pixel alone carries too little information to decide a clear fit. The resulting altitude model will not be consistent, so additional information would be needed, e.g. to constraint of "smoothness". Another problem with cross-correlation based on pattern fitting is that we assume there are no height differences within the correlation pattern matrix, or at least only to a negligible extent. In this case, the pattern matching will fail or leading to the determination of an incorrect height value in the worst case.

In general, when we try to fit the left image $I_1$to the right image I_2 with maximum reliability, we are actually looking for the minimum of the energy function (3) [2]:

$$E(D) = \sum_p \left(C(p, d_p) + \sum_{q \in N_p} P\left[|d_p - d_q| \geq 1 \right] \right) \qquad (3)$$

The first term of the function (3) sums the cost C of fitting for the given pixel p, where d_p represents the parallax between the images I_1 and I_2. Because the fitting is done pixel by pixel, only the intensity values $I_1(p)$ and $I_2(q)$ are compared to calculate the cost. In a simpler solution, the absolute difference in intensity values is considered as a cost. In the formula, on the other hand, the second term represents a kind of penalty P for those pixels q where, taking into account the neighborhood N_p of the pixel p, the parallax difference is greater than one pixel. That is, the term of penalty P can result in mismatch even if the difference in intensity values between $I_1(p)$ and $I_2(q)$ is zero, but the parallax is too large relative to the surrounding pixels. This also means that this concept may give preference to point pairs where the difference in intensity values may be larger, but the parallax difference does not exceed 1 pixel. The resulting d_p parallaxes are stored separately in a matrix D with equal extent to the extent of the I_1 image, which can be displayed as a depth image with appropriate coding. In the depth image, different shades represent different parallaxes or depth ranges derived from them. The member P, which penalizes a sudden change in height, is bound to the pixel p in the formula, which in turn is related to the neighbouring pixels q, which are also related to their own neighbors. As a result, each pixel is related to each pixel while looking for the minimum of the function $E(D)$. This method is called global image matching, which is difficult to implement in practice. The main reason is that the computation time increases exponentially as the image size increases and above a certain image size, the practical application of the method does not pay off.

The semi-global image matching (SGM) method is not based on the whole image to reduce the computation time, but only on a path taken in 4 or 8 directions around the pixel p. Paths are chosen radially and symmetrically around the pixel p (Fig. 3).

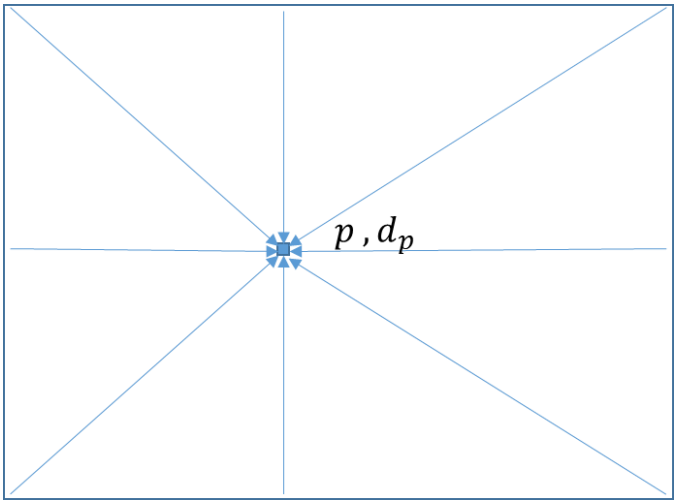


Figure 3. 8-way path to select the optimal d_p parallax

The cost calculation is performed separately for each path according to the energy function (3) and the parallax d_p belonging to the lowest cost is stored, from which the complete depth image can be generated (Fig. 4).

Figure 4. Applying SGM on a pair of images (source: http://lunokhod.org)

With this method, we can only process stereo image pairs. Depth maps from multiple stereo image pairs, on the other hand, can be integrated into a common model.

IV. DSM AND ORTHOPHOTOS

DSM data acquisition from a stereo image pair can be further automated during digital evaluation by various stereo-correlation methods, which automate stereoscopic orientation and thus allow the extraction of ground point coordinates.

A common feature of correlation methods is that the search for homologous point pairs is done using image patterns (correlation image matrices). The image patterns can be compared in the original image coordinate system or - after generating the normal stereogram from the original image pair - in the normal-case images, and finally the comparison can be realized in the raster space of the orthophoto, in which case the orthophoto is generated. In addition to the correlation image matrix, the comparison can be done independently or in addition to shape and topology recognition. Then the alignment

of the related points in the image (line, polyline) is preceded by the selection according to the interest operators, which further increases the robustness of the correlation.

V. Conclusions

As a summary we can say that a fully automatic evaluation of stereoscopic images pairs or block of images is not possible in every case. The image matching techniques are working well only in those cases if the following conditions are fulfilled [4]:

- GCPs are marked with same signs (size, colour, form is unique and same for every GCP).
- There are no occlusions in the corresponding images and the common points are visible.
- There are no ambiguous object structures or transparent surfaces, which can result several candidates for common points or features.
- For regions with poor texture, the solution is stable or non-sensitive with respect to minor disturbances in the image (noise).
- Intensities in all images cover the same spectral regions.
- Constant illumination, atmospheric effects, and media interfaces for the period of image acquisition.
- Stable object surface over the period of image acquisition.
- Macroscopically smooth object surface.
- Opaque object surface.
- Largely diffuse reflection off the surface.
- Known approximate values for orientation data (image overlap) and object data (geometric and radiometric parameters).

We can meet some conditions in the above list but not all ones. Therefore, the fully automatic orientation and evaluation of images remains a task. On the other hand, we can see a very progressive development in this area thanks to the development of UAV software solutions where the demand and the requirement for urge solutions for fully automation are coming from the end users.

References

[1] M. Bäumker and F.J. Heimes: New Calibration and Computing Method for Direct Georeferencing of Image and Scanner Data Using the Position and Angular Data of an Hybrid Inertial Navigation System, Proceedings of OEEPE Workshop on Integrated Sensor Orientation, 2002

[2] Hirschmüller H.: Semi-Global Matching – Motivation, Developments and Applications, in: Dieter Fritsch (Ed.): Photogrammetric Week'11, Stuttgart, Wichmann Verlag, Berlin, ISBN 978-3-87907-507-2, pp. 173.183., 2011

[3] Kraus K.: Photogrammetry, Geometry from Images and Laser Scans, Walter de Gruyter GmbH & Co. KG, Berlin, 2. kiadás, ISBN 978-3-11-019007-6, 2007

[4] T. Luhmann, S. Robson, S. Kyle and I. Harley: Close Range Photogrammetry, Whittles Publishing, ISBN 1-870325-50-8, pp. 284., 2006

The role of GIS in spatial planning

János Katona
Óbuda University Alba Regia
Technical Faculty
Székesfehérvár, Hungary
katona.janos@amk.uni-obuda.hu

***Abstract*— GIS makes it possible to link large amount of spatial data to each other. Data integration results in synergy, also known as combined power. The application of GIS provides such new consequences that we could not have previously thought of. The databases used in spatial development are also suitable for registration, query and analysis. However, perhaps the biggest role of GIS is in decision support. The application of GIS contributes to the planning of the location of different objects, to the determination of the spatial distribution of different activities or to the determination of the areas that need to be developed. This article aims to present the role of GIS in spatial development through the example of the beneficiary settlements.**

Keywords — spatial analyses, spatial planning, settlement development

I. Introduction

According to the National Spatial Development Concept, the spatial system is the spatial joint appearance of the systems of various organizations of different natural, social and economic objects (spatial elements), their connection related to a given area, their complex system. The spatial system is formed by a spatially different set of society, economy, geographical location, natural factors and human resources. So within a given territorial unit, these factors form a system, influence each other and interact with each other. The combination of these creates a very diverse territorial capital. Therefore, modern societies intervene into the functioning of the territorial system of specific territorial units for various reasons. Since the 1950s modern states have begun to consciously influence the territorial system development In order to correct large disparities as well as market mechanisms. [1] [2]

Regional development is nothing more than "Monitoring and evaluating the social, economic, infrastructural and environmental territorial processes covering the country and the regions, determining the necessary intervention directions; definition of short- and long-term comprehensive development strategies, concepts and plans, coordination of development goals and programs, implementation of plans". [3]

Regional development and regional policies form a complex, diversified system at different territorial levels and define several strategic and development concepts. These documents directly and indirectly determine the development of each area. Areas can be delimited at the regional, county, district and municipal levels. The present study examines the delimitation of the beneficiary areas at the municipal level. [4] [5]

II. Levels of spatial planning in Hungary

The tasks of spatial development can be divided into urban and rural areas based on the different functions of the areas, based on which we distinguish between urban development and rural development. However, it is not practical to separate the two development categories, because many economic and social problems can only be solved through the cooperation of villages and cities. Contrary to the Hungarian approach, small towns, and their surroundings, where industry has already appeared, but even agriculture plays an important role, are considered rural areas in the European Union. The city and its surroundings can be considered as a living unit, and its development can only be imagined by integrating the territorial and settlement approach. The aim of rural development is the continuous self-sustaining and sustainable development of the capacities of rural areas, so it has a very close relationship with the agricultural economy. Rural development policy helps to create sustainable agriculture, diversify farms, expand employment opportunities, and thus the competitiveness of rural areas.

At the same time, spatial planning is a hierarchically structured institutional system, the national, regional, county, micro-regional and settlement levels of which are distinguished. Higher-level territorial units necessarily include lower-level ones, and conversely, higher levels are aggregates of smaller spatial units. According to the target areas at different levels, we can differentiate between settlement development, micro-regional development, regional development, etc. According to some views, settlement development does not belong to the concept of regional development, because the scope of the Regional Development Act does not cover settlement development, but in the spatial science sense, the totality of levels forms regional development. The development strategy and the management plan must be based on a unified situation analysis and concept, both must be considered when planning the various programs.

The spatial plan is a plan document defining the long-term technical-physical structure of the country and certain regions, which ensures the long-term utilization and protection of spatial features and resources, enforcement of ecological principles, coordinated development of technical-infrastructure networks and optimal land use structure. At the national and regional level, the main goal is to create a geographically dispersed, small volume but mutually reinforcing networking, in which the two most important elements are the agricultural economy and the natural environment. The other levels of planning also have specific, unique functions, but the physical, socio-economic environment, social community, values, and interests have a common denominator. Certain territorial-environmental value choices may restrict owners (possibly other citizens) from exercising their rights. These should be enacted and included in the settlement plan. The development plan aims to promote conflicts between national spatial planning interests and national, territorial and local land use intentions.

At the regional level, a regional development concept is being prepared for several counties to be managed together from a social, economic or environmental point of view. The concept identifies the situation of the different territorial units and makes suggestions for the most effective developments, thus helping regional policy decisions. The spatial planning task of the counties is primarily the elaboration of infrastructural and environmental protection developments affecting several settlements, as well as the coordination of national and local zoning plans. The settlements are mostly connected to a well-defined central settlement, which means the micro-regional system, for which a planning document is also prepared. In order to limit their settlement development goals and the enforcement of individual interests, local governments regulate the formation of the natural and artificial environment in a development plan. Whereas previously land use was primarily regulated, today land use is increasingly regulated. Contrary to higher-level plans, the settlement level is not dominated by comprehensive regulation of land use, but by administrative and official elements. Once the settlement plan is adopted, a regulatory plan is drawn up that directly delimits possible improvements.

III. National Spatial Development and Spatial Planning Information System in Hungary

The National Spatial Development and Spatial Planning Information System (TeIR) aims to provide central, regional and local public administrations, other legal entities, unincorporated companies and natural persons with:

- provide an opportunity to get to know the population, economy, condition of the built, landscape and natural environment of the country, its territorial characteristics, to monitor its changes, to compare them with the European Union,
- service information by displaying data and indicators and analyses obtained during their processing, presenting text and map documents of regional development concepts and programs, spatial planning plans, settlement development concepts, integrated settlement development strategies and settlement planning tools,
- assist governmental, regional, regional, county, district, municipal development and planning, other regional and sectoral planning, development and monitoring organizations in preparing and making spatial development and spatial planning decisions, social, economic and environmental by continuously monitoring the changes in its territorial characteristics, to analyse the effects of decisions and to prepare regional development concepts and programs, settlement development concepts, integrated settlement development strategies, and spatial planning plans and settlement planning tools,
- offer information to local and regional authorities and regional development councils for planning, program management, tender evaluation and monitoring activities.

The data content of TeIR is very extensive. Spatial planning and specialization (Protected natural areas of Hungary, Location of landfills in the country, Hungary's transport network, Operating and abandoned mines in Hungary, The natural environment of Hungary, Hungary's electricity system, Contact information, Changes in the administrative borders of Hungary, The administrative system of Hungary, CORINE land cover, Drinking water quality in Hungary) also includes the following:
Remediation data
Nature and landscape protection
Public employment
Contact times and distances
Institutional provision
Aggregate data for industrial parks
Hazardous waste generated
Infocommunication data
Unemployment
General Agricultural Census
Regional Statistical Data System
Budapest Statistical Data System
County-Regional Statistical Data System
Census, microcensus data
Employees by occupational groups
Average rate of local taxes
Municipal management data
Sick pay data
Energy data
Local road data
Road accidents
Simplified Business Tax (EVA)
Personal income tax (PIT)
Corporate tax (TAO)
PIT civilian 1% offerings
Patient number, prescription data
Forest areas
Inpatient care
Drinking water quality
Outpatient treatment
Screening data
Data on disadvantaged primary school students
Higher education enrollment data
Students of foreign nationality
Flood and inland water damage
Dangerous plants
Domestic patent applications of Hungarian applicants. [6]

IV. The method of classifying beneficiary settlements

The criteria for the classification of the beneficiary settlements are set out in Decree 105/2015. (IV. 23.). When classifying settlements on the basis of territorial development, it is necessary to take into account a complex indicator formed from social and demographic, housing and living conditions, local economy and labour market, as well as infrastructure and environmental indicators (four groups of indicators). [7]

A. Indicator group 1: Social and demographic situation

- Mortality rate (number of deaths per 1000 population) is the average of the last five years, ‰
- Migration difference per thousand inhabitants (average of the last five years), main
- Number of beds providing nursery care and day care services per ten thousand permanent residents aged 0-2, pcs
- Proportion of recipients of regular child protection benefits from the permanent population aged 0-24,%

- Number of people receiving active care (regular social assistance and employment replacement support) per thousand permanent residents, persons.

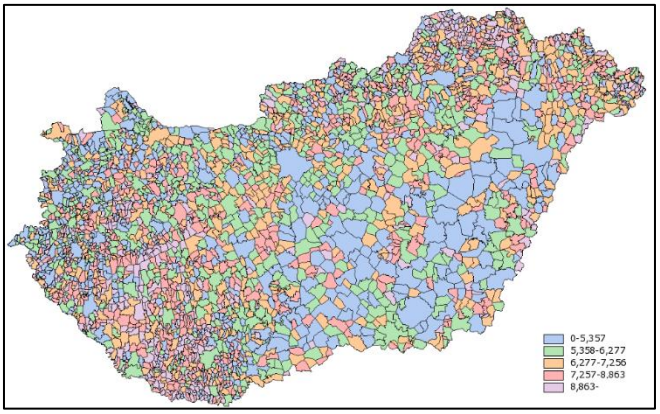


Fig.1. Internal emigration [%] (Edited in TeIR)

B. Indicator group 2: Housing and living conditions

- Average price of used apartments, HUF
- Proportion of dwellings built during the last five years out of the housing stock at the end of the period,%
- Dwellings without comfort (inhabited) as a percentage of inhabited dwellings
- The income forming the PIT base per permanent resident, thousand HUF
- Number of age-weighted passenger cars operated by natural persons per thousand inhabitants, pcs
- Social-economically and infrastructurally beneficiary settlements
- Significant unemployment affected municipalities
- Beneficiary settlement in both aspects

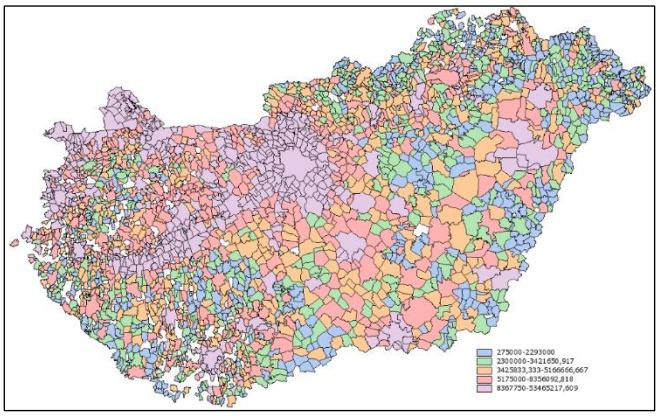


Fig.2. The average price of second-hand flats sold in 2016 [HUF] (Edited in TeIR)

C. Indicator group 3: Local economy and labour market

- Proportion of those aged 18 and over with at least a high school diploma,%
- Proportion of registered jobseekers in the working age permanent population,%
- Proportion of permanently registered jobseekers for at least 12 months out of the permanent working age population,%
- Proportion of registered jobseekers with no more than primary school,%
- Number of operating enterprises per thousand inhabitants, pcs
- Number of retail stores per thousand inhabitants, pcs
- Proportion of local tax revenues of local governments from current year revenues,%

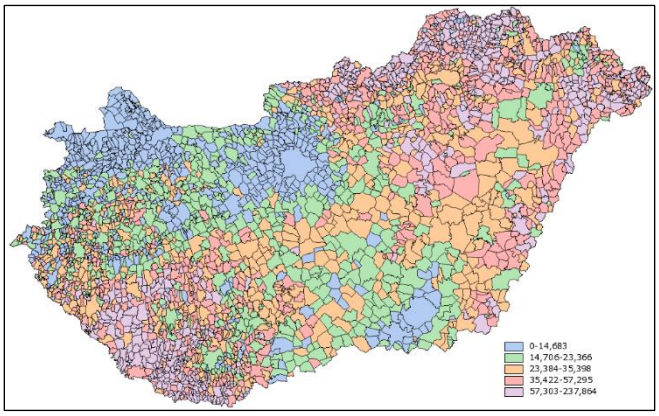


Fig.3. Proportion of registered jobseekers in the working age permanent population [per 1000 people] (Edited in TeIR)

D. Indicator group 4: Infrastructure and environment

- Proportion of dwellings connected to the public sewerage network,%
- Proportion of dwellings involved in regular waste collection,%
- Number of broadband internet subscribers per thousand inhabitants, pcs
- Proportion of roads built out of all municipal roads,%
- Indicator of the achievement of the county seat, minutes
- Indicator of access rate for controlled-access highway, minutes

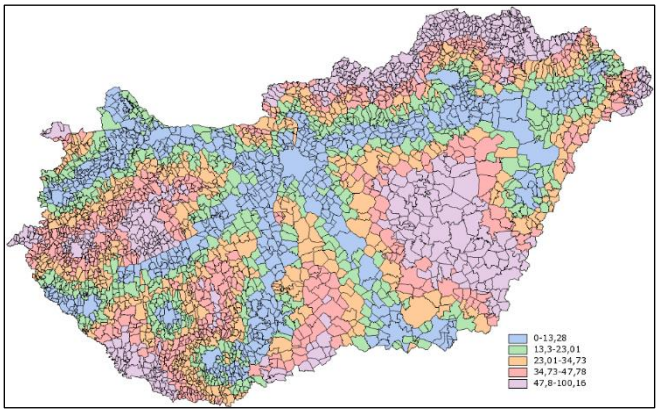


Fig.4. Access rate for expressways [min] (Edited in TeIR)

V. AGGREGATION OF INDICATORS

The method of calculation is laid down in the cited government decree. The transformation of the basic indicators to the same scale is performed by a normalization procedure according to the equation 1.

$$fa_{i,j,norm} = \frac{fa_{i,j} - \min(fa_{i,j})}{\max(fa_{i,j}) - \min(fa_{i,j})} \cdot 100 \qquad (1)$$

$fa_{i,j,norm}$: normalized basic indicator

$fa_{i,j}$: basic indicator

min ($fa_{i,j}$): the smallest value of the basic indicator

max ($fa_{i,j}$): maximum value of the basic indicator

Calculation of group indicators: the average value of the basic indicators within a group gives the value of the group indicator of the given group according to the equation 2.

$$fa_i = \frac{1}{n}\sum_{j=1}^{n} fa_{i,j,norm} \quad (2)$$

fa_i: group indicator

$fa_{i,j,norm}$: normalized basic indicator

n: number of indicators in group

Calculation of a complex indicator: the average value of the four group indicators gives the value of the development complex indicator according to the equation 3.

$$f_i = \frac{1}{m}\sum_{i=1}^{m} fa_i \quad (3)$$

fa_i: group indicator

f_i: complex indicator

m: number of group indicators

As a result of the analysis, 1053 disadvantaged or beneficiary settlements are demarcated. (Fig.5.) The range of beneficiary settlements should be reviewed in accordance with the planning cycles of the European Union. The review may also be initiated by the Government on a separate schedule. [8]

Almost half (45.9%) of the settlements in Hungary have never been included in the socio-economic and infrastructurally disadvantaged settlements, and 16.4% only in one or two ways have been added to the list. At the same time, almost 10% of the Hungarian settlement population are more than six times between the demarcated settlements, so their peripherality was well established when calculating a complex indicator. More than 80% of the settlement population of Pest, Komárom-Esztergom, Győr-Moson-Sopron and Fejér counties was never included in the list of underdeveloped settlements. At the same time, more than 10% of the settlements of Borsod-Abaúj-Zemplén, Szabolcs-Szatmár-Bereg, Baranya and Somogy counties were included in the lagging settlements during the larger delimitation. The former is extra large concentrates the backward settlements in proportion, as Borsod-Abaúj-Zemplén more than a third of the county settlements are almost all the demarcations considered peripheral. [9]

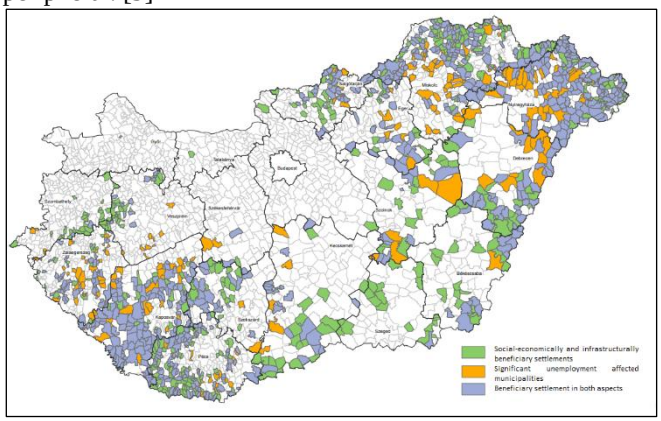


Fig.5. Beneficiary settlements in Hungary (TeIR)

VI. Conclusions

Overall, it can be stated that many settlements in Hungary have an absolute advantage over the others, while others have an absolute advantage they are at a disadvantage. However, the research also shows that this case may change over time, as settlements that have changed their beneficiary status depending on when and which indicators are used, they were delineated. Breaking out of disadvantage is a goal that is everything it floats there in front of a troubled settlement, and which you usually find very difficult alone inability to achieve. Territorial development and territorial policies at different levels are responsible to help these settlements to recover from their disadvantages so that so that they could eliminate large territorial disparities in the country, or at least moderate. This can be aided by a GIS database that aggregates complex data.

Acknowledgment

The maps appearing in the article were made in the National Spatial Development and Spatial Planning Information System (TeIR). I should like to thank the Lechner Knowledge Centre for allowing you to use TeIR free of charge.

References

[1] J. Rechnitzer, M. Smahó, Területi politika. Akadémiai Kiadó, Budapest. In: Székely A. (szerk.) Közép-európai közlemények, 2011, pp. 245-249

[2] M. Gottdiener , R. Hutchison, The New Urban Sociology, 4th Edition, 2020, 456 p.

[3] Act XXI of 1996 on Regional development and Regional Planning, Hungary

[4] Government Decree 105/2015. (IV.23.) about The classification of the beneficiary settlements and the classification system, Hungary

[5] G BOR, Z. TÓTH, G. LÁSZLÓ, Low-cost distance measuring methods and their investigation for BIM modelling In: Orosz, Gábor Tamás (szerk.) AIS 2019 : 14th International Symposium on Applied Informatics and Related Areas organized in the frame of Hungarian Science Festival 2019 by Óbuda University Székesfehérvár, Magyarország : Óbudai Egyetem, (2019) pp. 54-60. , 7 p.

[6] Á. Varga (szerk.): Földrajzi Információs Rendszerek gyakorlati alkalmazása, Gazdaságföldrajz és Jövőkutatás Központ, Budapesti Corvinus Egyetem, Budapest (2016) 124 p.

[7] O. Alhusain, Z. Tóth, Á. Rakusz, L. Almási, B. Farkas, Vision-based System for Quality Control of Some Food Products ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING XXXV (2004) pp. 477-482., 6 p.

[8] J. Katona, M. Horoszne Gulyas, Determination of factors modifying land value based on spatial data In: Drótos, Dániel; Vásárhelyi, József; Czap, László; Ivo, Petráš (szerk.), Proceedings of the 19th International Carpathian Control Conference (ICCC 2018) Piscataway (NJ), Amerikai Egyesült Államok : IEEE, (2018) pp. 625-628. , 4 p.

[9] J. Pénzes, Periferikus térségek lehatárolása Magyarországon – módszertani és területisajátosságok. In: Nagy E, Nagy G (szerk.): Polarizáció - függőség - krízis: Eltérő térbeliválaszok. 251 p. Békéscsaba: MTA KRTK RKI Alföldi Tudományos Osztály, 2014, pp. 163–175.

An Implementation of Exploratory OLAP System Based on Prasad's Approach

Géza Molnár
Department of Computer Science
Corvinus University of Budapest
geza_molnar@freemail.hu

***Abstract*—Unstructured and semi-structured data stored in data warehouses (e.g. comments, descriptions) are often not processed or ignored. As a result, valuable information may be lost, although there are several ways to handle these mostly textual data. Once this type of information is extracted, the result can be further investigated; amongst others, it can be incorporated into existing OLAP analysis. Technologies, more specifically frameworks, that allow the creation of a multidimensional schema from unstructured data of data warehouse are called exploratory OLAP. This article discusses a possible implementation of exploratory OLAP system based on Prasad's model [1].**

Keywords—exploratory OLAP, text analysis, star schema

I. INTRODUCTION

. Business reports are often based on a special data structure, so-called OLAP (Online Analytical Processing) cube. OLAP is a computational method that allows users to efficiently query and analyze data from a variety of perspectives. In another approach, it is a multidimensional (MD) schema. It can be queried with MDX (Multidimensional Expressions) statements [2].

OLAP technologies work well for adequately structured data. However, it can be a challenge to work with unstructured or semi-structured data. For example, creating a report on a product campaign, it needs to combine structured information (e.g., product sales data) and unstructured information (e.g., user reviews).

Exploratory OLAP systems provide an answer to this challenge. They work on different principles, but they have in common that they all create a semantic layer in the data warehouse / BI system. This layer has several advantages; amongst others, it allows interpreting non-structured data, as well.

II. RELATED WORKS

The literature is not uniform in the conceptual framework and implementation principles of exploratory OLAP. The best-known approaches are from Abello [3], Prasad [1] and Ibragimov [4].

Abello's exploratory OLAP concept uses semantic web technologies, namely, ontologies. They are frequently used for knowledge representation [5].

In Abello's system, there is a mapping between the source data and an ontology, so-called reference ontology. The mapping makes it possible to explore Functional Dependencies (FD) and multidimensional (MD) identifiers. The purpose of the latter is to identify the facts, and so the MD scheme can be created (Fig. *1*).

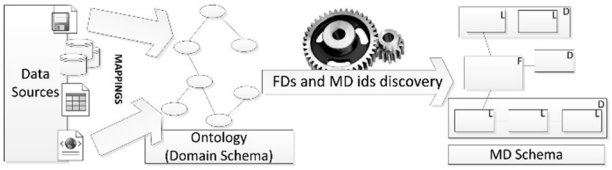


Fig. 1. Abello's model

Prasad suggested another approach. His solution is based on text mining and analysis techniques (bag-of-words approach [6]) that can be used to integrate structured and unstructured data into the data warehouse. The solution consists of text analysis (statistical and semantic analysis), identification of key terms, and creation of text tags (text tagging). Its system's main components are shown in the following figure (Fig. *2*).

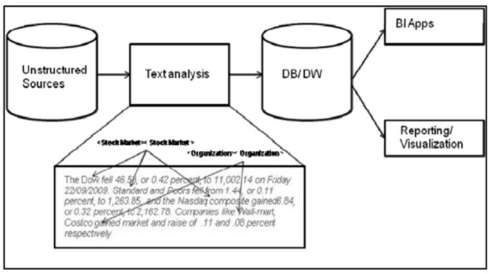


Fig. 2. Prasad's model

First, text analysis/preprocessing procedures are applied for the text extracted from unstructured sources. Then the results data will be loaded into specific tables of the database/data warehouse. The data model is star-schema (Fig. 3), so it is suitable for creating OLAP cubes and reports.

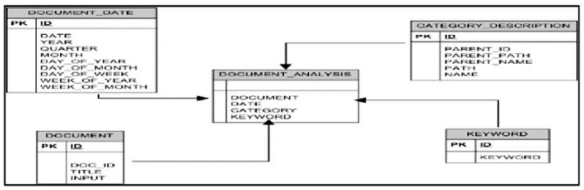


Fig. 3. Star schema

The architecture proposed by Ibragimov works with web data stored in a so-called RDF (Resource Description Framework) format. To describe things, RDF uses triplets in the form of an object-subject-statement [7]. Since it is a computer-readable and XML-based language, it has become a standard model for web data exchange. The RDF structure can be queried using special query languages such as SPARQL. Ibragimov's basic idea is to create a mapping between

SPARQL querying an RDF schema and MDX statements querying an OLAP schema. The proposed system consists of four modules (Fig. *4*):

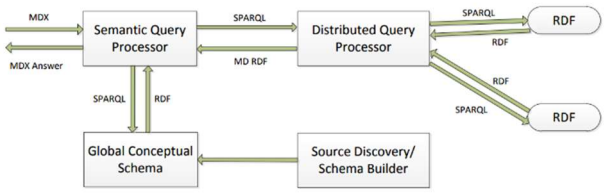


Fig. 4. Ibragimov's model

- Global Conceptual Schema - stores information about the OLAP cube
- Semantic Query Processor - Generates SPARQL queries from MDX queries
- Distributed Query Processor - queries endpoints, collects data
- Source Discovery Schema Builder - keeps in touch with users as the schema is created.

Each of the exploratory OLAP framework proposals presented above handles text data processing differently. The following table compares the three approaches.

TABLE 1 Comparison of Exploratory OLAP systems

Point of comparison	Abello	Ibragimov	Prasad
Data sources	Any	Linked Open Data	Any
Tools, technologies	Ontologies	RDF, MDX, SPARQL	Text analysis, XML
Biggest challenge	Ontologie →MD	MDX → SPARQL	Text analysis
Is there a prototype plan?	No	Yes	Yes
Output location	OLAP cube	OLAP cube	Data Warehouse
Output format	OLAP schema	OLAP schema	Star schema

The next part presents the creation of a prototype based on Prasad's idea. As can be seen from the previous table, the output of the prototype is not an OLAP cube but a star schema. It is not a problem in practice, since the OLAP cube can be built easily on the star schema.

III. Implementation of Exploratory OLAP System based on Prasad's Model

Creating the prototype consists of four steps. The first one is the data preparation, which, on the one hand, means extracting the necessary data from the source system and, on the other hand, creating an initial list of keywords related to the topic. It is followed by a text processing, which results in the data needed to create the structure proposed by Prasad. The third step is the star schema creation, as shown in Fig. 3. The last step is the creation of the OLAP cube.

The following software was used during the implementation:

- Service Desk (ticketing) program to extract data from source systems [8]
- Python 3.7 (Jupyter, Anaconda) for text processing
- MS SQL Server 2017 Express, SQL Server Management Studio and SQL Server Import and Export Wizard for star schema creation
- Power BI Desktop for emulating the OLAP-cube.

The prototype uses a ticketing (helpdesk) system's textual data, which arise when users report various problems ("incidents").

A. Data Preparation

The data source contains data of incidents from the ticketing system covering a six months interval. Each report has two textual data, namely error summary and error description. Besides, the ticket creation date (Open_Date) and the ticket identification number (Ticket Number) are also essential. For privacy reasons, the Ticket Number has been replaced with a counter-type field.

The first step of the data preparation is the data export (Ticket Number, Open_Date, Summary, Description) from the ticketing system using the filter and transformation as mentioned above. For the sake of simplicity, during the export CSV format was applied. It is appropriate either for the text analyses, or the SQL-based database/data warehouse systems as an input format. It is different from the XML format suggested by Prasad, but this is irrelevant for the later steps. Since the ticketing system contains only consistent data, so minimal data cleaning (e.g. data deduplication) was needed. The latter can occur if, e.g. multiple users report the same problem in the same way at the same time.

The second step is to define the keywords. These are words that are typically associated with error reports and can be used to categorize the incidents. The easiest way to create a list of keywords manually based on the common error descriptions. A better way is to look for the most common words (possibly word combinations) in the text data source and then select the relevant ones (e.g. server, network, printer). It is easily accomplished with a Python-script, which

- opens the source data
- eliminates unnecessary spaces
- converts everything to lowercase
- breaks the text into words
- omits stopwords and words not found in the English dictionary
- keeps only nouns
- finally, filter for words that occur at least 100 times

The resulting list has been filtered for words that can be associated with the ticketing system. It is not necessary to maintain a list of keywords for each data export. It is enough to update it a few times a year.

In case of a keyword list, also the CSV file format is preferred. Consequently, after the data preparation, two CSV files have been created with the following structure: INCIDENTS.CSV (ID, OPEN_DATE, SUMMARY, DESCRIPTION) and KEYWORDS.CSV (ID, KEYWORD). Currently, the first CSV file contains about 3000, and the second one has 50 records. Later, the amount of data will be increased.

B. *Text Analysis*

The text analysis has been performed with another Python-script. It implements the TextRank algorithm [9]. Its basic idea comes from the PageRank algorithm [10], which was initially designed to rank websites. According to this, a website's ranking position depends mostly on how many other "important" websites link to it. The TextRank algorithm was implemented by using the description of Xu Liang [11].

The PageRank algorithm assigns a directed graph to web pages. The graph's nodes are web pages, and the edges describe the links between them. Weights can also be given to each node according to the following formula:

$$S(V_i) = (1-d) + d * \sum_{j \in In(v_i)} \frac{1}{|Out(V_j)|} S(V_j) \qquad (1)$$

where

- S(Vi) – the value of each weight,
- d – dumping factor
- In(Vi) – set of connected nodes (incoming edges)
- Out(Vi) – set of connected nodes (starting edges)

The graph mathematically can be represented by an n * n matrix, where n is the number of nodes, and the jth element of the ith row equal to 1 there is an edge from the ith node to the jth node. Otherwise, the given element is 0.

The initial values of PageRank are obtained by normalizing the matrix's columns and multiplied by the vector containing the weights of each node. It is followed by an iteration in which the PageRank values are recursively modified at each step. It results in a series of random variables (irreducible, aperiodic Markov chain). Due to the damping factor d, it has a boundary distribution.

The PageRank algorithm is one of the essential elements of the Google search engine. PageRank values can also be considered as the probability of a page being clicked, and the graph contains pages that can be affected during a browse.

The TextRank algorithm works with sentences instead of web pages. Here, the jump from one website to another corresponds to the similarity of two sentences. The similarity data are stored in a matrix. The main steps of the algorithm are illustrated in the following figure:

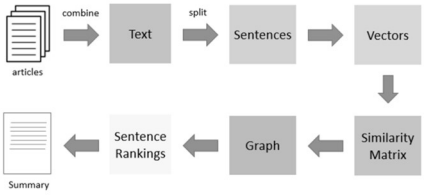


Fig. 5. TextRank algorithm

The TextRank algorithm wraps the sentences into units (words) and assigns to them grammatically appropriate categories as a label. The method is called POS tagging (Part of speech tagging), which can result in tags assigned to words, such as NOUN (noun), PROPN (proper name), and VERB (verb) [12].

It is enough to consider only nouns and proper names when searching for keywords. The keywords will be the ones with the highest weight.

The script that implements the algorithm begins by importing the necessary packages, then sets the values of initial data (number of iteration steps, attenuation factor, convergence threshold). The next step is the data processing, during which the information extracted from the text data will be brought into the following structure:

DOCUMENT (DOC_ID, DATE, TITLE, INPUT, KEYWORD, CAT)

where

- DOC_ID - the document identifier
- DATE - the date when the ticket was opened
- TITLE - the title of the document (the original summary field are included here)
- INPUT - the line that contains the keyword detected by the algorithm
- KEYWORD - the keyword discovered by the algorithm
- CAT - the keyword category (label), which can be NN: noun, NNP: proper name, VBN: verb, XX: other

The data processing algorithm's pseudocode is the following:

```
document = empty list
keyword_list = open(keywords.csv)
text = open(text.csv)
loop for incidents
    loop for sentences
        analyze
        extract keywords
        compare keywords with keyword_list
        if found common words then
            loop for common words
                create a text tag for the word
                insert a new record into the document.csv
            end of loop
        end of selection
    end of loop
end of loop
OUT: document.csv
```

It can be seen that the algorithm takes into account only the sentences containing keywords.

C. *Star Schema*

For the star schema creation, the first step is to load the results of the text analysis to the database/data warehouse. The easiest way to do this is to use the SQL Server Import and Export Wizard, which allows loading a CSV file contents into a database.

The second step is to implement the structure proposed by Prasad. It can be easily solved by creating SQL views. The relationships between the tables can be defined either here or at the Power BI data model creation. Interestingly, the fact table (DOCUMENT_ANALYSIS) does not contain measures, only date and dimension identifiers. It is important to note that both the data processing and the data loading can be automated and scheduled.

Automation will be implemented later as follows:

- The data processing script can be run using the Windows built-in task scheduler (Windows Task Scheduler)

- CSV files can be loaded into the database with the Microsoft's best-known ETL tool (SQL Server Integration Services)
- Data loading can be scheduled by the SQL Server Job Agent.

D. OLAP Cube

The OLAP cube is usually created with a server-side tool, such as SQL Server Analysis Services. However, the task can also be solved more straightforwardly - from the client side - because the data model generated using the Power BI Desktop program is compatible with it. It means that the model can be tested in Power BI and then later opened in Analysis Services without any change.

With the help of the SQL views, a star schema-based data model can be created. Power BI detects and automatically creates connections between tables, if necessary. The model ("OLAP cube") can then be tested by switching to a visual view and making, e.g. the following dashboard:

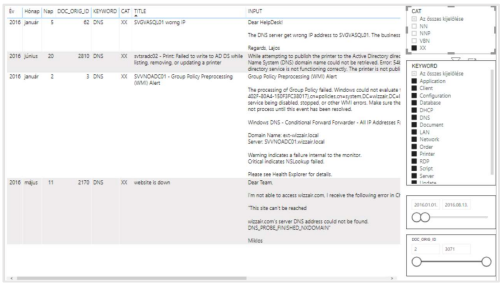


Fig. 6. Power BI Dashboard

The dashboard shows the text of each ticket description, its keywords and categories, and the date the ticket was created in a tabular layout. The list can be filtered by category, keyword, date and document ID. The results show that the process still needs to be optimized with running it many more times using different settings to achieve the best possible outcome.

IV. CONCLUSION

Textual data in databases and data warehouses can hide much valuable information. The role of exploratory OLAP systems is to uncover this hidden information. This article describes how to create a prototype of such a system. The basic idea comes from Prasad, but the implementation is slightly different. The applied algorithm creates the necessary star scheme, but its efficiency could still be improved. On the one hand, during the execution of the algorithm, it would be more effective to look at the semantic similarity of keywords instead of matching them. On the other hand, in addition to keywords, multi-word key expressions could be considered, as well.

Future work can include text mining-based taxonomy development integration [13].

ACKNOWLEDGEMENT

Funding Project No. NKFIH-869-10/2019 and its continuation in 2020 have been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the Tématerületi Kiválósági Program funding scheme.

REFERENCES

[1] K. S. N. Prasad, "Text Analytics to Data Warehousing," International Journal on Computer Science and Engineering,, 2010.

[2] E. F. C. S. B. a. S. C. T. Codd, "Providing OLAP to User-Analysts: An IT Mandate," *Computer Science,* pp. 3-5, 1993.

[3] A. Abelló, "Using Semantic Web Technologies for Exploratory OLAP: A Survey," IEEE, 2015.

[4] K. H. T. B. P. E. Z. Dilshod Ibragimov, "Towards Exploratory OLAP Over Linked Open Data – A Case Study," International Workshop on Business Intelligence for the Real-Time Enterprise, 2015.

[5] T. Gruber, "A Translation Approach to Portable Ontology Specification," Knowledge acquisition, Vol. 5 No. 2, 1993, pp. 199-200.

[6] B. K. a. M. T. Matthew Gentzkow, "Text as Data," *Journal of Economic Literature,* vol. 57(3, pp. 537-539, 2019.

[7] H. L. a. R. S. K. Selc ¸uk Candan, "Resource Description Framework: Metadata and Its Applications," *ACM SIGKDD Explorations Newsletter,* vol. 3, pp. 8-12, 2001.

[8] "CA Service Desk Manager," [Online]. Available: https://www.ca.com/us/products/ca-service-desk-manager.html.

[9] P. T. Rada Mihalcea, "TextRank: Bringing Order into Text," in *Association for Computational Linguistics*, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, 2004.

[10] P. P. Anuj Joshi, "Google Page Rank Algorithm and It's Updates," *International Journal of Management, Technology And Engineerin,* pp. 1108-1110, 2018.

[11] X. Liang, február 2018. [Online]. Available: https://towardsdatascience.com/textrank-for-keyword-extraction-by-python-c0bae21bcec0.

[12] F. M. H. U. Khan, "Comparison of different POS Tagging Techniques (n-gram, HMM and Brill's tagger) for Bangla," in *Springer, Dordrecht*, Advances and Innovations in Systems, Computing Sciences and Software Engineering, 2007.

[13] A. &. G. S. Ko, "A Research Review and Taxonomy Development for Decision Support and Business Analytics Using Semantic Text Mining.," *International Journal of Information Technology & Decision Making (IJITDM),* no. 19(01), pp. 97-126, 2020.

[14] P. Joshi, 2018. [Online]. Available: https://www.analyticsvidhya.com/blog/2018/11/introduction-text-summarization-textrank-python/..

Open-source Robot Arms in Research and Education

Balázs Máté
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
balazs98528@gmail.com

Dániel Valicsek
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
vb.valicsekdaniel@gmail.com

Márta Seebauer
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
seebauer.marta@amk.uni-obuda.hu
ORCID: 0000-0001-8459-8975

Károly Széll
Alba Regia Technical Faculty
Óbuda University
Székesfehérvár, Hungary
szell.karoly@amk.uni-obuda.hu
ORCID: 0000-0001-7499-5643

***Abstract*— There is a growing industrial demand on engineers with up-to-date robotics knowledge which requires the improvement of the theoretical and practical education of robotics on university level. This paper proposes an environment for teaching the planning, assembly and programming of robotic arms. Two open source platforms, Nyrio One and BCN3D Moveo, are reviewed from research and educational point of view.**

Keywords—robotic arm, Arduino, Raspberry, educational robots, industry 4.0, digital education

I. Introduction

As digital education is increasingly gaining ground [1]–[3], state-of-the-art topics like robust engineering solutions [4]–[6], predictive maintenance [7]–[10], industry 4.0 [11], [12] , robotics [13], [14], control theory [15]–[18] or sensor technology [19]–[21] need new concepts for the structure of the university curriculum. Thus, a further challenge of this aspect is the modification of the laboratories to meet these new requirements.

In today's world, robots play an extremely important role, without which it would be almost impossible to meet the needs of the world. For instance, the production of cars alone, or the mass production of any product where 6-axis industrial robots help, or may completely replace, human labor. This is not surprising, as they are able to move loads of up to hundreds of kilograms, and they can also operate around the clock. Thus, it is worth automating most of the manufacturing processes, which can be solved in accordance with the economical requirements. With the application of automation, almost any process can be performed quickly, precisely, efficiently, from welding, painting and packaging, to servicing CNC machining centers.

II. Problem Description

The implementation is based in the Robotic Center of the Alba Regia Technical Faculty of Obuda University. The main purpose of this paper is the introduction of a 6-axis desktop robot arm, on which students can get acquainted with the basics of robotics. One of these would be available to each student, or one per group of two people. Thus, one of the primary considerations would be to minimize the cost of making such a robotic arm. Using 3D printing technology, the production of more complex elements can be done extremely cheap. For the introduced prototype a Wanhao Duplicator i3 printer were used. The axes of the robot arm are NEMA 17 stepper motors and MG995 servomotors. As control unit Raspberry, Arduino Mega 2560 microcontroller and Ramps 1.4 control panels are used.

III. BCN3D Moveo

The first investigated open source robot arm is called BCN3D Moveo [22] (see Fig. 1). The aim is to assemble the complete unit from the lowest possible budget using commercially available elements. This requirement made a few modifications necessary on the original open source model. The last shaft of the original model is driven by a NEMA14 standard size stepper motor. As it was not commercially available in Hungary, it had to be redesigned to accommodate a NEMA17 motor. As the motor is larger than the original, the model had to be strengthened at the relevant geometries.

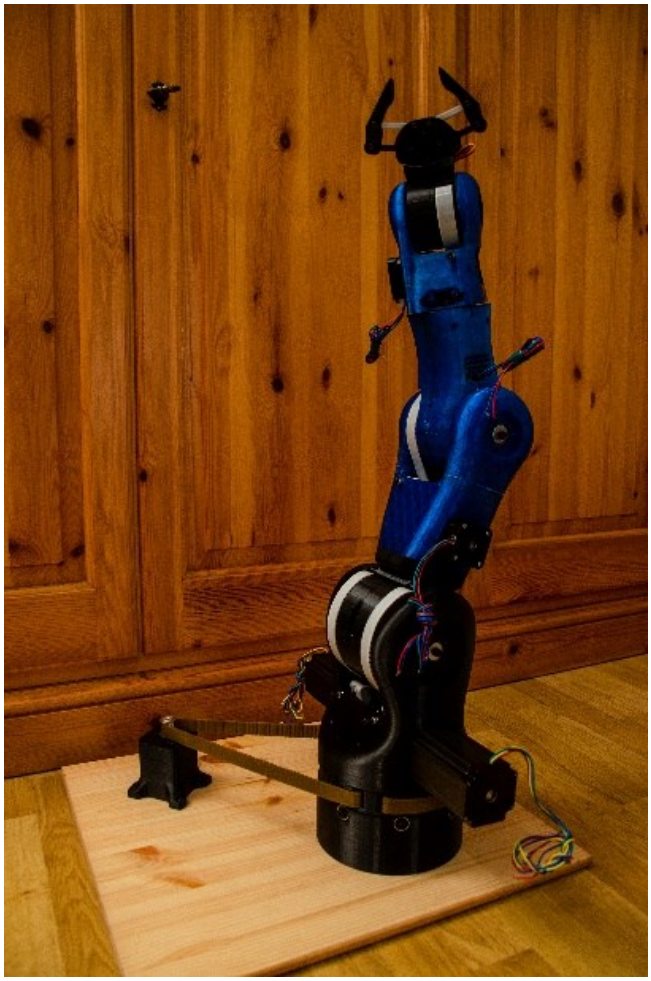


Fig. 1: BCN3D Moveo

PETG was chosen as the raw material for 3D printing. The finished workpieces are a bit brittle, but this durable material with a higher fill value is extremely solid. It can withstand multiple crashes and still perform its tasks. Furthermore, since all its elements are made in 3D printing, they can be replaced with a commercial printer in case of more serious damage. As several mistakes occurred during the assembly, elements had to be replaced. Once the right printing technology was found, replacing them was easy and fast. Some of the commercial material required a little finishing, e.g. the toothed pulley was delivered raw. Thus, the center holes and fixing threads on them had to be made. Also, 3D printing technology includes a bit of finishing work in all cases, because the "supports" must be removed and the contours have to be deburred. A university workshop, fablab provides all the tools to perform these basic tasks. In case all the mechanical elements arrived, and everything is at our disposal, assembling is not a difficult task, which can be done in 3-4 days even by an inexperienced university student (see Fig. 2).

Fig. 2: Assembly of BCN3D Moveo

The setup offers a wide range of possibilities for the education of engineers from different fields (see Fig. 3). As control unit amongst others Arduino or Raspberry might be used. As software Node-RED, ROS or MATLAB might be applied. Since it contains only simple stepper motors, the robot can even be operated with a PLC. Talented students can even add modifications with their own improvements to improve the mechanical printed element of the robotic arm, like reducing the weight of the arm, developing a more appropriate gripper or mapping current weak points with a load test.

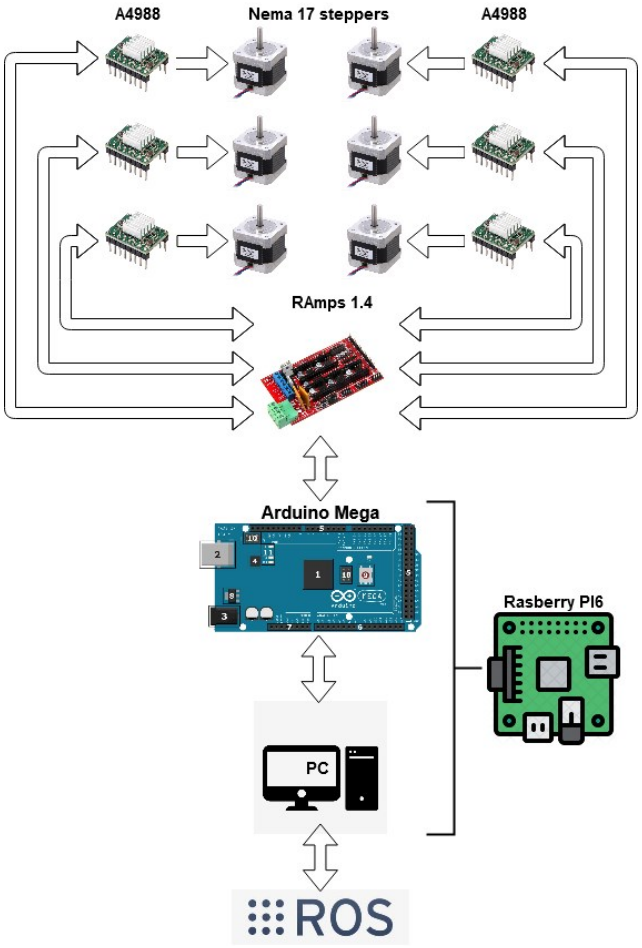


Fig. 3: Control architecture of BCN3D Moveo

Purchasing all mechanical parts, the price of the robotic arm can be kept below 350 EUR facilitating that each workstation in the training room can have its own robotic arm. Students can create their own robot arm for the educational material, gain experience, how to make an automated machine. They have insight into the whole process. Considering the safety aspects, industrial robots are dangerous and require a lot of care. This way, students can acquire the necessary programming basics in a safe environment. With the knowledge they have already acquired, they will be ready to start with the industrial robots of the university like the Fanuc, Universal Robots, KUKA and ABB education cells with more confidence and humility.

IV. NYRIO ONE

The second investigated open source robot arm is called Nyrio One [23] (see Fig. 4). The axes of the Nyrio One consist of 4 NEMA 17 stepper motors and 2 MG995 servomotors, and the gripper is also operated by such a servo. For programming an Arduino Mega 2560 microcontroller and a Ramps 1.4 control panel were applied, which is primarily designed to control the axes of 3D printers but is also perfect for this purpose. Several A4988 motor control circuits are also required to move the stepper motors.

Fig. 4: Nyrio One

The robot can be purchased from Nyrio's official website for a few thousand dollars, which is relatively expensive, but everyone has access to the models, source codes, and lists of mechanical and electronic parts needed to purchase. In view of the relatively inexpensive implementation, the robot has been modified in several respects.

Considering the easy printability, both the PLA and PETG correspond to the raw material of the robotic arm. The smaller elements were printed from PLA, while the larger ones were printed from PETG. The parts printed from PLA were made using a proprietary Wanhao Duplicator i3 Mini printer, the slicing software used is Wanhao Cura, Settings for 3D printing:

- Heating block temperature: 210 ° C
- Print speed: 60-70 mm / s
- Nozzle diameter: 0.4 mm
- Wall thickness: 0.8 mm
- Layer height: 0.2 mm
- Fill factor: 20%
- The use of a support is necessary in most cases

Most of the parts were printed based on the original STL files without modification, but the elements related to the last 2 axes needed redesign due to the MG995 servomotors. In addition, I wanted to mount the gripper of of the BCN3D MOVEO, on the Nyrio, thus the design of an adapter was necessary. The robot would include two more 100 mm long aluminum tubes, which were also made of PLA material, which made the robot a bit lighter. With this element, however, the correct orientation of the form must be taken into account during printing, since the torques acting on the finished workpiece allow the layers to separate from each other, so it is worth printing this tubular piece horizontally, so this phenomenon can be eliminated.

The assembly videos on Nyrio's official website are useful, but there are some issues and problems that arise during construction. It is advisable to prepare the already printed elements before assembling, first remove all the supporting materials from the pieces. After that, any surface defects in the printed elements should also be removed by sandpaper and a hand-held mini sander.

The first 4 shafts are driven by a NEMA17 stepper motor, the first 3 motor shafts have a GT2 type ribbed wheel, the torque of the robot's shafts is driven by a GT2 ribbed belt, the belt tension can be adjusted with M3 bolts and self-locking nuts. An endless strap was placed on the first axis of the robot, I solved the endlessness of the strap with gluing and a thread with a thin line. The stepper motor responsible for driving the fourth shaft is equipped with a planetary gear with a ratio of approximately 1: 5, the torque of which is transmitted by a small clutch, which allows a slight distortion in both the axial and radial directions due to possible angular errors. On the 5th and 6th axes, MG995 servomotors provide rotational movement, which can be easily installed with a few screws. Due to the low load, it is not necessary to use a gear unit on these shafts.

With the exception of the last, 6th axis, all axes are equipped with bearings, and due to their construction, the first and fourth axes, in addition to the radial deep groove ball bearings, also have axial ball bearings to ensure the smooth movement of the robot. It is also worth mentioning the torsion steel spring located in the second shaft (see Fig. 5), which takes over part of the torque from the loads from the stepping motor responsible for driving the shaft, thus providing the motor with the correct value of the holding torque so that it will not lose step.

Fig. 5: Assembly of Nyrio One

The created robot can be moved in a simulation environment using a robot operating system (ROS) running on Ubuntu installed on a virtual machine, followed by the movement of the real, physical environment (see Fig. 6).

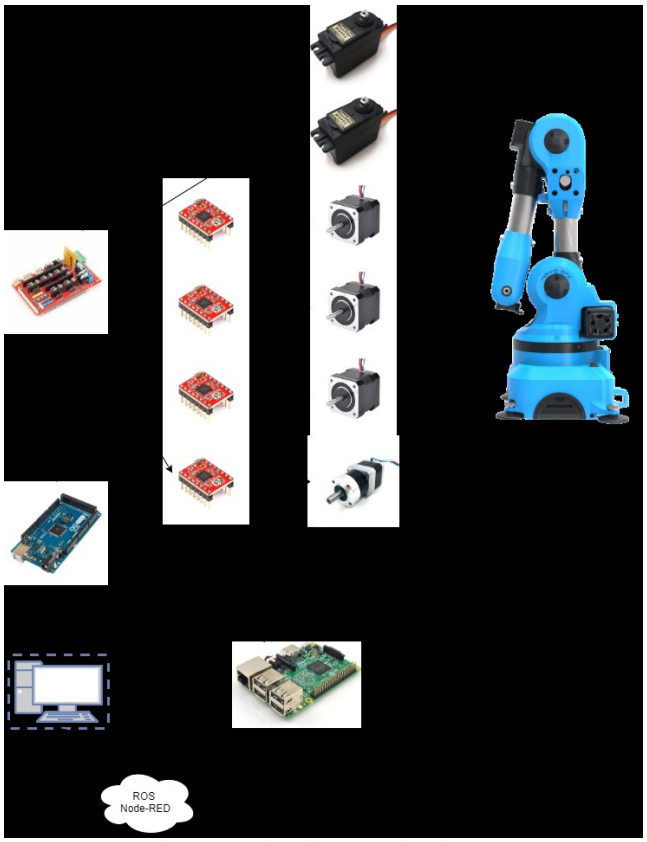


Fig. 6: Control architecture of Nyrio One

V. Further Possibilities

There is a lot in common between a robot arm and a 3D printer: stepper motors, motor controllers, etc. This facilitates the design of various devices applying the same elements. Students first create a custom 3D printer as a low-budget project. Its size is such that it can easily make the largest element of the robotic arm outlined above. Students learn about the technology with all its pros and cons. They acquire the necessary knowledge to control the motors with the help of ARDUINO. They can develop their programming or mechanical assembly knowledge. Later, they can print the above-mentioned robot arm on this printer. As a next step, by disassembling the 3D printer, the purchased and printed elements can be transformed into a robotic arm using almost all the components of the printer, such as bearings, stepper motors and the associated motor controls, belts etc. Only a few new parts are necessary. With this transformation option, university education can be taken to a new, higher and more practical level.

VI. Conclusion

The paper introduced the application of two different open source robotic arm models for the use of university studies and research purposes. The purpose of the experimental setups is to introduce the mechanical, electrical and software structure of a robotic arm on university level. The proposed environment facilitates the education of several engineering fields with relatively low costs.

Acknowledgment

The authors thankfully acknowledge the financial support of this work by the Howmet Aerospace Foundation, the Hungarian State and the European Union under the EFOP-3.6.1-16-2016-00010 and 2019-1-3-1-KK-2019-00007 projects.

The research has been supported by the NTP-HHTDK-19-022 grant. The authors declare no conflict of interest. The authors would like to express their gratitude to those who have provided their help in the various experiments.

References

[1] M. Pogatsnik, “Measuring Problem Solving Skills of Informatics and Engineering Students,” in *IEEE Joint 19th International Symposium on Computational Intelligence and Informatics and 7th International Conference on Recent Achievements in Mechatronics, Automation, Computer Sciences and Robotics, CINTI-MACRo 2019 - Proceedings*, Nov. 2019, pp. 93–98, doi: 10.1109/CINTI-MACRo49179.2019.9105277.

[2] M. Pogátsnik, “Életpálya-építés serdülő- és ifjúkorban - kutatás közben,” *Egyéni Kül. Szerepe Tanulásban És Pályaválasztásban*, pp. 149–162, 2015.

[3] M. Pogatsnik and R. B. Kendrovics, “Communication and Reading Comprehension among Informatics and Engineering Students,” in *SAMI 2020 - IEEE 18th World Symposium on Applied Machine Intelligence and Informatics, Proceedings*, Jan. 2020, pp. 235–240, doi: 10.1109/SAMI48414.2020.9108764.

[4] G. Györök and B. Beszédes, “Fault tolerant power supply systems,” in *11th International Symposium on Applied Informatics and Related Areas (AIS 2016)*, 2018, pp. 68–73.

[5] G. Györök and B. Beszédes, “Highly reliable data logging in embedded systems,” in *2018 IEEE 16th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, Feb. 2018, pp. 49–54, doi: 10.1109/SAMI.2018.8323985.

[6] G. Györök and B. Beszédes, “Adaptive Optocoupler Degradation Compensation in Isolated Feedback Loops,” *2018 IEEE 12th Int. Symp. Appl. Comput. Intell. Inform. SACI*, pp. 167–172, 2018.

[7] G. Manhertz and Á. Antal, “The effect of air-fuel equivalence ratio change on the vibration components of an internal-combustion engine,” *RECENT Innov. Mechatron. 2 1-2*, pp. 1–6, 2015, doi: 10.17667/riim.2015.1-2/13.

[8] G. Manhertz and Á. Bereczky, “Development of a vibration expert system to analyze and predict malfunctions in internal-combustion engine,” in *Proceedings of ARES’14 : Workshop on Application of Robotics for Enhanced Security*, 2014, pp. 24–29.

[9] G. Manhertz, D. Modok, and Á. Bereczky, “Evaluation of short-time fourier-transformation spectrograms derived from the vibration measurement of internal-combustion engines,” in *2016 IEEE International Power Electronics and Motion Control Conference (PEMC)*, Sep. 2016, pp. 812–817, doi: 10.1109/EPEPEMC.2016.7752098.

[10] G. Manhertz, G. Gardonyi, and G. Por, “Managing measured vibration data for malfunction detection of an assembled mechanical coupling,” *Int. J. Adv. Manuf. Technol.*, vol. 75, no. 5, pp. 693–703, Nov. 2014, doi: 10.1007/s00170-014-6138-3.

[11] É. Hajnal, "Big Data Overview and Connected Research at Óbuda University Alba Regia Technical Faculty," in *AIS 2018 - 13th International Symposium on Applied Informatics and Related Area*, 2018, pp. 1–4.

[12] A. Selmeci and T. Orosz, "Usage of SOA and BPM changes the roles and the way of thinking in development," in *2012 IEEE 10th Jubilee International Symposium on Intelligent Systems and Informatics*, Sep. 2012, pp. 265–271, doi: 10.1109/SISY.2012.6339526.

[13] P. Galambos, "Cloud, Fog, and Mist Computing: Advanced Robot Applications," *IEEE Syst. Man Cybern. Mag.*, vol. 6, no. 1, pp. 41–45, Jan. 2020, doi: 10.1109/msmc.2018.2881233.

[14] A. I. Karoly, P. Galambos, J. Kuti, and I. J. Rudas, "Deep Learning in Robotics: Survey on Model Structures and Training Strategies," *IEEE Trans. Syst. Man Cybern. Syst.*, pp. 1–14, Sep. 2020, doi: 10.1109/tsmc.2020.3018325.

[15] C. Budai and L. L. Kovács, "On the Stability of Digital Position Control with Viscous Damping and Coulomb Friction," *Period. Polytech. Mech. Eng.*, vol. 61, no. 4, pp. 266–271, Sep. 2017, doi: 10.3311/PPme.10537.

[16] C. Budai and L. L. Kovács, "Friction Effects on Stability of a Digitally Controlled Pendulum," *Period. Polytech. Mech. Eng.*, vol. 59, no. 4, pp. 176–181, Oct. 2015, doi: 10.3311/PPme.8298.

[17] C. Budai, L. L. Kovács, J. Kövecses, and G. Stépán, "Combined effects of sampling and dry friction on position control," *Nonlinear Dyn.*, Jul. 2019, doi: 10.1007/s11071-019-05101-7.

[18] C. Budai, L. L. Kovács, and J. Kövecses, "Combined Effect of Sampling and Coulomb Friction on Haptic Systems Dynamics," *J. Comput. Nonlinear Dyn.*, vol. 13, no. 6, Jun. 2018, doi: 10.1115/1.4039962.

[19] Z. Pentek, T. Hiller, and A. Czmerk, "Algorithmic Enhancement of Automotive MEMS Gyroscopes with Consumer-Type Redundancy," *IEEE Sens. J.*, pp. 1–1, Aug. 2020, doi: 10.1109/jsen.2020.3017094.

[20] Z. Pentek, T. Hiller, T. Liewald, B. Kuhlmann, and A. Czmerk, "IMU-based mounting parameter estimation on construction vehicles," in *2017 DGON Inertial Sensors and Systems, ISS 2017 - Proceedings*, Dec. 2017, vol. 2017-December, pp. 1–14, doi: 10.1109/InertialSensors.2017.8171504.

[21] T. Hiller, Z. Pentek, J. T. Liewald, A. Buhmann, and H. Roth, "Origins and Mechanisms of Bias Instability Noise in a Three-Axis Mode-Matched MEMS Gyroscope," *J. Microelectromechanical Syst.*, vol. 28, no. 4, pp. 586–596, Aug. 2019, doi: 10.1109/JMEMS.2019.2921607.

[22] "BCN3D MOVEO - A fully Open Source 3D printed robot arm," *BCN3D Technologies*, Jul. 28, 2016. https://www.bcn3d.com/bcn3d-moveo-the-future-of-learning/ (accessed Nov. 08, 2020).

[23] "Nyrio One," *Niryo*. https://niryo.com/ (accessed Nov. 08, 2020).

Aspects of direct georeferencing in photogrammetry

Attila Varga
Doctoral School of Applied Informatics and Applied Mathematics,
Óbuda University, Budapest, Hungary
vargaa@stud.uni-obuda.hu

***Abstract* – Reliable and high-precision GNSS systems and inertial measurement units are playing increasingly important role in aerial surveying. As a result, direct georeferencing is receiving increasing attention in photogrammetry. In this case, exterior orientation elements of the images are available without measuring ground control points. Using this method, fast and near real-time photogrammetric processing can be implemented. Using exterior orientation parameters provided by inertial navigation systems, the direct georeferencing task can be fulfilled by solving an overdetermined system of equations, therefore we get an approximate solution. The system of equations can be solved by analytical method, but also by numerical method using iteration. This paper reviews the characteristics of direct georeferencing, typical measurement errors, and compares solution methods, analyzing their accuracy and computational needs using MATLAB.**

***Keyword* – photogrammetry, direct georeferencing, MATLAB**

I. Introduction

Direct georeferencing in photogrammetry is the process where exterior orientation parameters are measured directly. It means that image projection centers (X, Y and Z coordinates) and the rotation angles are measured at moment of taking photo. Position of projection centers can be measure by GNSS and rotation angles produces by inertial measurement unit (IMU). As a result of direct georeferencing the measurement of GCPs is not required.

Using this method photogrammetric workflow is getting simplified. One of advantages of direct georeferencing is the more economical workflow. In addition, there is no need for expensive and resource-consuming surveying of GCPs. Direct georeferencing enables fast and real-time or near real-time photogrammetric processing. Sensitive point of direct georeferencing is the accuracy of measurement of exterior parameters of images.

The realization of direct georeferencing was made possible by development of inertial navigation, including gyroscopes. Nowadays light weight satellite navigation systems with correction services (e.g. RTK) allow geodesy level accuracy and the same is true for gyroscopes too. Gyroscopes used in the early inertial measurement units were large, expensive and complex equipment, but above all they were inaccurate that made direct georeferencing possible only with large errors. Development of inertial navigation over the last years has resulted the accuracy of inertial instrument and significant reduction of their price and size.

Introduction of microelectromechanical systems (MEMS) in gyro technology resulted low-cost, light gyroscopes, while for professional application laser gyroscopes are available at an affordable price and in compact form. This meant that reliable and accurate inertial navigation systems can be installed also on the board of UAVs. A contemporary inertial navigation system consists of GNSS with correction system and inertial measurement unit, including gyroscope and magnetometer. However, it should be taken in account, that even today the errors of gyroscope (mainly its drift) provide more inaccuracy in direct georeferencing than accuracy of triangulation method using GCPs.

II. Theoretical model

In Fig. 1. the theoretical model of direct orientation can be seen. The task is to determine the position of the P point in the geographical coordinate system by the way of measuring the position of the projection center, the camera orientation and measuring points in images.

The GNSS instrument installed on the aircraft provides position in geographical coordinate system ($\overline{r}^{GNSS}$). The IMU provides the orientation of the aerial vehicle (indicated by $\mathbf{R}^{IMU}$ rotation matrix). It should take in account the fact that the GNSS instrument and the IMU are usually located in different part of the aircraft.

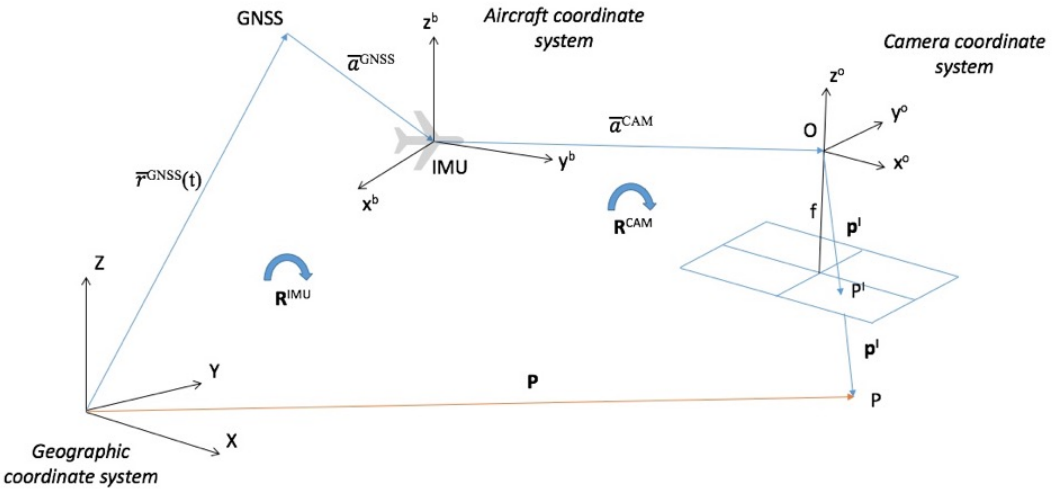


Figure 1. The model of direct georeferencing

The orientation of the camera relative to the aircraft ($\mathbf{R}^{CAM}$ rotation matrix) may also be vary. Such a case if the camera installed on gimbal. For simplification, these two rotation matrices are handled in one rotation matrix and all exterior orientation parameters transformed to the projection center.

III. Errors in direct georeferencing

Resulting errors of direct georeferencing come from interior and exterior orientation of images. In conventional photogrammetric processing using aerial triangulation errors occurred in measuring GCPs transmitted to the exterior orientation parameters by error propagation during the calculation process.

In direct georeferencing, besides the errors of interior orientation, errors of sensors can be detected directly in exterior orientation parameters. These errors coma from errors of GNSS unit and errors of IMU, mainly from gyroscope. In terms of accuracy the gyroscope is the most critical part of the inertial system.

Errors of interior orientation can also be different:

- *focal length error,* caused by distortion of the lens due to changes in pressure and temperature,
- *error of principal point,* that can be resulted by geometry error of lens caused by defect in manufacturing the lens,
- *distortion,* caused by geometry and material inhomogeneity of lens,
- *pixel error*, caused by measurement error of image point.

Fortunately, majority of interior errors can be eliminated or reduced by calibration.

IV. Solutions

In direct georeferencing determination of external orientation parameters is realized at the same time as exposing the photo by measuring data from external sensors (GNSS and IMU). In this way the photogrammetric process is considerable simplified, essentially the main task is the restoration of spatial data of terrain points (or point cloud), which is usually the last step in the conventional photogrammetric process using triangulation and measurement of GCPs.

Spatial data for point is restored by processing stereo image pairs. However, one terrain point can be represented on multiple images, so multiple stereo image pairs can be processed. It means, that for a terrestrial object point multiple coordinates are calculated. The result should be calculated by performing least square adjustment. In the following it the paper deals with only the calculation of stereo image pairs.

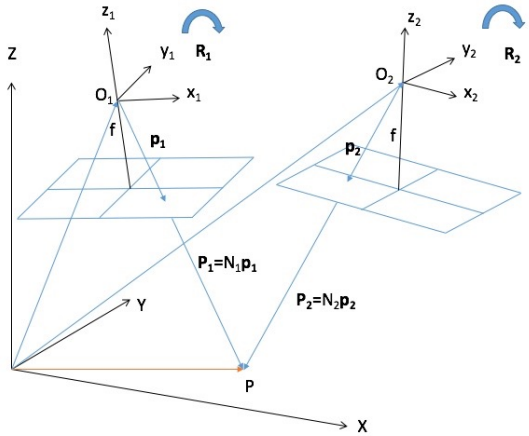


Figure 2. Theoretical model of direct georeferencing

A. Solving by geometrical method

Knowing exterior orientation data position of terrestrial point can be calculated by using spatial intersection from two projection centers. Due to the collinear criterion, the P position vector to the P terrestrial point can be calculated by the following (1) formula:

$$\boldsymbol{P} = \boldsymbol{P_o} + N\boldsymbol{p} \tag{1}$$

Where $\boldsymbol{P}$ *is a position vector,* $\boldsymbol{P_o}$ ais a position vector of projection center, *N is the scale factor,* $\boldsymbol{p}$ is the position vector of the terrestrial point represented on the image, transformed into the geographic coordinate system.

The transformation between the geographic coordinate system and the image coordinate system is provided by the rotation matrix. Image coordinates of the object (ξ, η) can be measured from images, and the focal length (f) is also known. Exterior orientation parameters measured by sensors. The above formula (1) in matrix form is the following:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} X_{01} \\ Y_{01} \\ Z_{01} \end{bmatrix} + N_1 \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \eta_1 \\ -f \end{bmatrix}$$
$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} X_{02} \\ Y_{02} \\ Z_{02} \end{bmatrix} + N_2 \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \begin{bmatrix} \xi_2 \\ \eta_2 \\ -f \end{bmatrix} \tag{2}$$

In the equation system consisting of six equations there are five variables: X, Y, and Z represent the coordinates of P terrestrial point, N_1 and N_2 represent scale factors for images. It means that the system of equation is overdetermined and having solved the result is an approximation. The r_{mn} coefficients are known and calculated from rotation angles consisting of trigonometric expressions.

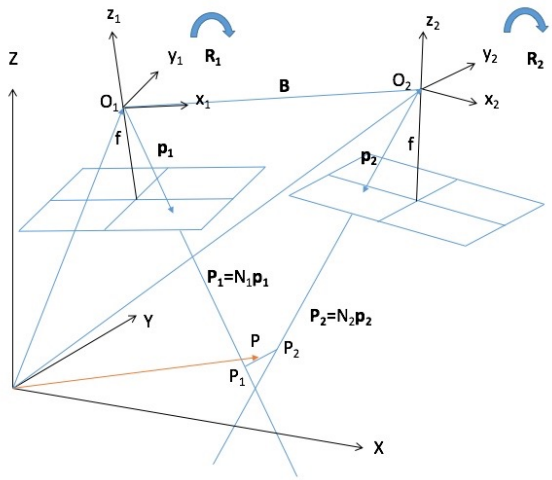


Figure 3. The real model in case of non-intersecting projection rays

It should also take in account that due to measurement errors (in exterior and interior orientation) the sights of projection do not intersect is reality (see Fig. 3.). The aim is to find solution where the point is as close as possible to the two projection lines.

In this paper there are two possible way of solution is presented. The firs one is the most commons method known from professional sources [1]. It is based on geometric approach.

According to the collinearity criterion, vector of the image point and vector of the terrestrial point are aligned is a straight line (projection ray). It means that cross product of these two vectors is zero.

Having converted the formula, the N scale factor can be expressed by the following way:

$$\begin{gathered} \boldsymbol{p_2} \times \boldsymbol{P_2} = 0 \\ \boldsymbol{p_2} \times (N_1\boldsymbol{p_1} - \boldsymbol{B}) = 0 \\ \boldsymbol{p_2} \times N_1\boldsymbol{p_1} - \boldsymbol{p_2} \times \boldsymbol{B} = 0 \\ N_1 = \frac{|\boldsymbol{p_2} \times \boldsymbol{B}|}{|\boldsymbol{p_2} \times \boldsymbol{p_1}|} \end{gathered} \tag{3}$$

Three N_1–N_2 pairs can be calculated by this method. In the case of $X=X_1=X_2$, $Z=Z_1=Z_2$ (in the XZ plain) N scale factors resulted by expressing cross product by coordinates:

$$N_1 = \frac{B_x z_1 - B_z x_2}{x_1 z_2 - z_1 x_2},\ N_2 = \frac{B_x z_1 - B_z x_1}{x_1 z_2 - z_1 x_2} \tag{4}$$

Having resulted N_1 and N_2 scale factors coordinates of object point (X, Y, Z) are calculated:

$$\begin{gathered} X = X_1 = X_2 = X_{01} + N_1 x_1 = X_{02} + N_2 x_2 \\ Z = Z_1 = Z_2 = Y_{01} + N_1 z_1 = Y_{02} + N_2 z_2 \\ Y_1 = Y_{01} + N_1 y_1,\ Y_2 = Y_{02} + N_2 y_2 \end{gathered} \tag{5}$$

In the real practice only one N1–N2 pair is usually calculated. In this recent case the Y coordinate is resulted by averaging the calculated Y_1 and Y_2 coordinates:

$$Y = \frac{Y_1 + Y_2}{2} \tag{6}$$

Accuracy can be improved by calculating multiple N-N pairs and averaging the coordinate values. But for computational economy reasons, this step is mostly ignored.

In visual approach the above means that task is solved in the XY coordinate plain where projected images of spatial projection rays intersect each other. In this case the problem has been simplified to a two-dimensional problem.

Suppose that the best solution using geometric approach if transverse between two projection rays is calculated. The ideal theoretical intersection of

non-intersecting spatial lines lies at half of the transverse.

B. Solving by numerical method

Another solution is to define an equation system with the introduction of a residual as an unknown then using iteration while minimizing the residual.

To fulfill the numerical solution a similar method was chosen, taking advantages of MATLAB numerical computing environment. Adjusting slightly equations from collinear condition a system of linear equations is resulted. This is a more formal solution and well-managed by MATLAB. It also was a test to use MATLAB for solution photogrammetric tasks.

MATLAB as a first step to solve the system of linear equations is trying to solve it as a definite equation system. If neither the Cholesky nor the Gaussian method (decomposition of coefficient matrix) has been successful, MATLAB will solve equation system by using QR decomposition and iteration, minimizing residual vector. It is possible that MATLAB provides a solution even is case if it cannot be solved, therefore it is recommended to check the magnitude of residual vector.

Multiplying rotation matrix and image point position vector (see Eq. 2.) a new coefficient vector has been introduced containing c_{mn} coefficient, the following formula is resulted:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - N_1 \begin{bmatrix} c_{11} \\ c_{12} \\ c_{13} \end{bmatrix} = \begin{bmatrix} X_{O1} \\ Y_{O1} \\ Z_{O1} \end{bmatrix} \qquad \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} - N_2 \begin{bmatrix} c_{21} \\ c_{21} \\ c_{23} \end{bmatrix} = \begin{bmatrix} X_{O2} \\ Y_{O2} \\ Z_{O2} \end{bmatrix} \tag{7}$$

Equations are written separately:

$$\begin{aligned} X - c_{11}N_1 &= X_{O1} \\ Y - c_{12}N_1 &= Y_{O1} \\ Z - c_{13}N_1 &= Z_{O1} \\ X - c_{21}N_2 &= X_{O2} \\ Y - c_{21}N_2 &= X_{O2} \\ Z - c_{21}N_2 &= X_{O2} \end{aligned} \tag{8}$$

Equation system is expressed in general $Ax = b$ form:

$$\begin{bmatrix} 1 & 0 & 0 & -c_{11} & 0 \\ 0 & 1 & 0 & -c_{12} & 0 \\ 0 & 0 & 1 & -c_{13} & 0 \\ 1 & 0 & 0 & 0 & -c_{21} \\ 0 & 1 & 0 & 0 & -c_{22} \\ 0 & 0 & 1 & 0 & -c_{23} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ N_1 \\ N_2 \end{bmatrix} = \begin{bmatrix} X_{O1} \\ Y_{O1} \\ Z_{O1} \\ X_{O2} \\ Y_{O2} \\ Z_{O2} \end{bmatrix} \tag{9}$$

The solution of $Ax = b$ system of linear equations in MATLAB is performed by using the command $x = A \setminus b$. The residual error can also be calculated: $r = Ax - b$

V. Calculation example

To compare the two methods, sample date was used taken from paper published on the internet [2]. Parameters of images of pair:

Projection center (UTM coordinates in meter)

$$\mathbf{O_1} = \begin{bmatrix} 432588.64254 \\ 4921230.83708 \\ 1550.10375 \end{bmatrix}$$

$$\mathbf{O_2} = \begin{bmatrix} 433038.78510 \\ 4921222.37288 \\ 1550.44516 \end{bmatrix}$$

Rotation angles in radians

Ω_1 = -0.000716
Φ_1 = -0.000678
K_1 = -0.027806

Ω_2 = -0.001074
Φ_2 = -0.000923
K_2 = -0.013588

Focal length

f = 100 mm

Image coordinates

$$\mathbf{p_1} = \begin{bmatrix} 4018.444\text{e}-006 \\ 4018.444\text{e}-006 \end{bmatrix}$$

$$\mathbf{p_2} = \begin{bmatrix} -31650.000\text{e}-006 \\ 76907.020\text{e}-006 \end{bmatrix}$$

Coordinates of point P_1, P_2 and P (location vectors) calculated by analytical method:

$$\mathbf{P_1} = \begin{bmatrix} 432664.847020694 \\ 4922168.468291757 \\ 323.811866059 \end{bmatrix}$$

$$\mathbf{P_2} = \begin{bmatrix} 432664.847020694 \\ 4.922168.836508479 \\ 323.811866059 \end{bmatrix}$$

$$\mathrm{P} = \begin{bmatrix} 432664.847020694 \\ 4922168.652400119 \\ 323.811866059 \end{bmatrix}$$

Difference vector between point P_1 and P_2:

$$P_1—P_{2=} = \begin{bmatrix} 0 \\ -0.368216722272336 \\ 0 \end{bmatrix}$$

It can be seen, that there is a difference only in the Y coordinates, since in the calculation we started from the conditions of $X=X_1= X_2$ and $Z=Z_1=Z_2$. This means a difference of 36,8 cm in the Y coordinate.

The coordinate values of the P point calculated by the numerical solution of system linear equations:

$$P_{num} = \begin{bmatrix} 432664.857654643 \\ 4922168.687976455 \\ 323.765144622 \end{bmatrix}$$

The residual error vector is also acceptable:

$$r = \begin{bmatrix} 0.002209184225649 \\ 0.116025147959590 \\ 0.088850910886549 \\ -0.002209184400272 \\ -0.116025149822235 \\ -0.088850910887686 \end{bmatrix}$$

The minimum distance between the two projection lines was also calculated using coordinate geometry method. It is a transverse between two lines, perpendicular to both spatial lines. The best solution for spatial intersection is the transverse halfway point (P_{TR}). Calculating the coordinates, the location vector of the P_{TR} point is the following:

$$\mathbf{P_{TR}} = \begin{bmatrix} 432664.857654643 \\ 4922168.687976455 \\ 323.765144622 \end{bmatrix}$$

In the case of analytical solution, the difference of point P from PTR:

$$\mathbf{P—P_{TR}} = \begin{bmatrix} -0.010633948782925 \\ -0.035576337017119 \\ 0.046721436621510 \end{bmatrix},$$

the magnitude of this vector is 0,05967 m.

In the case of numerical solution, the difference of point P_{num} from point P_{TR}:

$$\mathbf{P_{num}—P_{TR}} = \begin{bmatrix} -0.058207660913467 \\ -0.931322574615478 \\ -0.012732925824821 \end{bmatrix} \cdot 10^{-9}$$

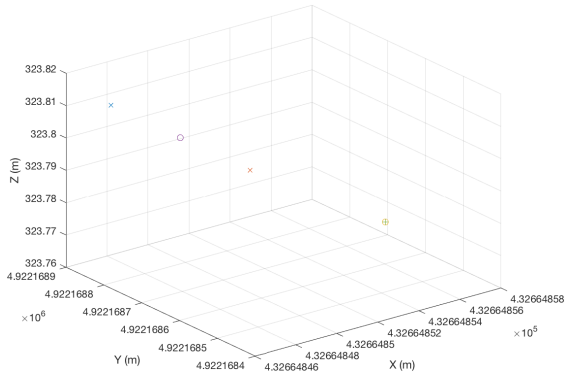


Figure 4. Graphical representation of calculated object point. From left to right: P_1, P, P_2, P_{TR} and P_{num} shown together.

VI. Conclusions

Direct georeferencing is a cost-effective and fast method in photogrammetry that is increasingly used in aerial survey. Its accuracy nowadays already is acceptable for aerial mapping or 3D point cloud generation. Solving the problem results in an overdetermined system of equations, so we only get an approximate solution to the point of intersection of the non-intersecting projection rays.

There are two methods were presented. One is an analytical solution based on geometric similarity. The other is to solve the system of equations as a system of liner equations by numerical method. The midpoint of the transverse of the two projecting lines is considered the solution. Deviation of the analytical solution from this optimal point is only 6 cm. Given that the solution of the linear system of equations also involves error minimization at the iteration, as expected, the result is practically the same as the best theoretical solution. This means that both solutions provide acceptable accuracy in practice. According to the preliminary assumption, the linear system of equations can be suitable for the simultaneous calculation of several images, but the solution of the system of equations requires further investigation.

To solve this problem, MATLAB was used, whose basic data format is the matrix. Experience shows that MATLAB is a very flexible and efficient tool for solving photogrammetric problems.

References

[1] Kraus, Karl: Photogrammetry – Geometry from Images and Laser Scan, Walter de Gruyte, Berlin, 2000, ISBN 978-3-11-019007-6, pp. 181—183.

[2] Constantin Razvan Cretescu: *Direct georeferencing using unmanned Aerial Vehicles,* in: Journal of Young Scientist, Vol. V, 2017, ISSN 2344-1283

[3] Xiuxiao Yuan, Xueping Zhang: Theoretical accuracy of direct georeferencing with position and orientation system in aerial photogrammetry, in: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B1, Bejing 2008

Mapping within-field soil water condition using Sentinel 2

M. Verőné Wojtaszek

Institute of Geoinformatics Alba Regia Technical Faculty Óbuda University Székesfehérvár, Hungary
wojtaszek.malgorzata@amk.uni obuda.hu

***Abstract*—Soil and crop water condition monitoring is a fundamental step in agricultural production. There are different methods to measure water stress, some of them are based on soil moisture measurements while others are based on calculations of vegetation indices, evapotranspiration or soil water balance. Currently, the use of remote sensing technologies for the analysis of plant water status comprises a wide range of available methods such as infrared thermometry for canopy temperature measures, microwave radiation for soil water content assessment, and spectral vegetation indices for the study of the reflectance responses of canopies to different environmental conditions. The aim of the presented work is to investigate the applicability of the optical remote sensing in mapping the moisture content within agricultural field.**

Keywords—Sentinel-2, NDVI, NDWI, OPTRAM, water stress monitoring

I. INTRODUCTION

Growth and reproduction of crops strongly depends on their environment, and any change in environmental conditions that determines a shift from the crop's optimal state can be considered as stressful. Water is a very important factor in agriculture and plays an essential role in crop production throughout the growing season. Water is required for the germination of seeds and as soon as growth starts water serves as a carrier in the distribution of mineral nutrients. Biomass production is inseparably connected with water demand. Many of the biochemical reactions that are part of growth occur in water or water itself participates in the reactions, e.g: respiration, photosynthesis.

Water is a vital component in the functioning of plants and soil moisture is the dominant factor controlling its supply [1]. So that information on water stress conditions and their analysis is necessary to achieve optimal outputs. Crop water stress monitoring represents a fundamental step in agricultural production, measures of plant water status are required to better understand the mechanisms of plant response and adaptation to water stress, and for the optimisation of crop production [2], through precision irrigation. In order to increase water savings and enhance agricultural sustainability, implementation of suitable irrigation scheduling methods is essential [3], and requires early detection of water stress in crops, before it causes irreversible damage and yield loss.

There are different methods to measure water stress, some of them are based on soil moisture measurements while others are based on calculations of vegetation indices, evapotranspiration or soil water balance. Conventional methods for monitoring crop water stress rely on in situ soil moisture measurements and meteorological variables to estimate the amount of water lost from the plant-soil system during a given period [4]. Other methods of detecting plant water status involve soil water balance calculations, direct and indirect measurement of plant water status, via stomatal conductance and leaf water potential. The disadvantage of measurements based on soil sampling is that these methods are time consuming and produce point information.

Currently, the use of remote sensing technologies for the analysis of plant water status comprises a wide range of available methods such as infrared thermometry for canopy temperature measures, microwave radiation for soil water content assessment, and spectral vegetation indices for the study of the reflectance responses of canopies to different environmental conditions. Soil optical reflection [5, 6], thermal emission [7, 8] and microwave backscatter [9, 10] are highly correlated with soil moisture content, numerous methods for optical, thermal and microwave RS of soil moisture have been developed.

Numerous studies are based on the calculation and analysis of spectral indices and have shown that there is relationship between reflectance values and canopy changes due to water stress. Measures reflectance indices within the VIS and NIR spectral range (e.g.: NDVI, RDVI, OSAVI) to indicate canopy changes due to water stress. Some indices (e.g.: PRI) are sensitive to the photosynthetic pigment changes due to water stress and others are used to measure the reflectance trough in the NIR and SWIR region (WI, SRWI, and NDWI) to represent canopy moisture content.

The so-called "trapezoid" or "triangle" model is one of the most widely applied approaches to RS of soil moisture utilizing both optical and thermal data. The model, hereinafter termed Thermal-Optical TRAapezoid Model (TOTRAM), is based on the interpretation of the pixel distribution within the land surface temperature-vegetation index space [11, 12]. Sadeghi et al. [13] proposed a novel physically-based trapezoid model, so called OPtical TRApezoid Model (OPTRAM), which is based on a relationship between soil moisture and shortwave infrared transformed reflectance. The theoretical basis of OPTRAM, evaluate the predictive capabilities of the universally parameterized OPTRAM with Sentinel-2 and Landsat OLI observations was presented in 2017.

The research presented in the article focus on the exploitation of Visible, Near Infrared and Short Wave Infrared

(SWIR) reflectance properties for the construction of vegetation and water indices and use of the optical trapezoid model targeted (OPTRAM) to the estimation of soil water content within field. The aim of the work is to investigate the applicability of the model in the examination and mapping of the moisture content of smaller areas (agricultural field). The model ability to provide plant physiology, vegetation characteristics, and crop water status at the canopy scale can improve the site-specific decision-making process in a precision agriculture.

II. MATERIALS AND METHODS

The model ability to provide crop water status at the field level was tested on a relatively small agricultural area. From a pedological point of view, the study area has average properties. Its humus content does not exceed 1-2%. The area has varied topographic features, the west part is heavily eroded. During the study period (2017-2018) the area was covered by winter wheat.

A. Satellite images and data analysis

Multispectral ESA Sentinel-2 satellite images acquired from the ESA Sentinel Scientific Data Hub were used in this study. A total of 17 cloud-free images were available, but after preliminary interpretation, 7 of them were selected for further processing. Sentinel-2 is a high spatial (10 to 60-m) resolution multispectral satellite having 13 spectral bands covering the visible, NIR and SWIR electromagnetic frequency domains and temporal resolution of ~5-day. The characteristics of the data used in the research include the Table 1.

TABLE I. THE SPECIFICATION OF REMOTE SENSING DATA

Datasets	The specification of RS data	
	Spectral resolution	*Acquisition date*
Sentinel-2	VIS, NIR, SWIR (bands: 3, 4, 8, 12)	2017 (Nov.19, Dec. 19.) 2018 (Jan. 18, Mar. 24, Apr. 13, May 13, Jun. 12)

The data extraction includes the following steps: data pre-processing, multi-level image segmentation, delimitation of indices and the application of the optical trapezoid model OPTRAM. It is important to note that the Sentinel-2 Level-1C processing includes radiometric and geometric corrections including ortho-rectification and spatial registration on a global reference system with sub-pixel accuracy. Calculation of the TOA (the top-of-atmosphere) reflectances also occurs in this process. Hierarchical framework was used to identify the boundary of study area at super-object level and determine the unit of investigation (20 m) at the second level for mapping surface moisture. After pre-processing vegetation index (NDVI) and (STR) were determined. Reflectance at the red band (B4: 665 nm) and the near infrared (B8: 842 nm) were used to calculate the NDVI. Reflectance at the SWIR band (B12: 2190 nm) was used for calculation of the STR following the Sadeghi et al. [14].

By measuring the reflectance of the plants or soil at various wavelengths, it is possible to collect a lot of information about the status of the plants and soil properties. The reflectance of light spectra depends on the land cover type (plant, soil types), water content within tissues, and other intrinsic factors. The reflectance of vegetation is low in the blue and red regions of the visible spectrum, due to absorption by chlorophyll for photosynthesis. It has a peak at the green region which gives rise to the green colour of vegetation. In the near infrared (NIR) region, the reflectance is much higher than that in the visible band due to the cellular structure in the leaves. In the mid infrared (SWIR) there are more water absorption regions (Fig. 1). Those regions are used to examine correlation between root zone soil moisture and the vegetation status.

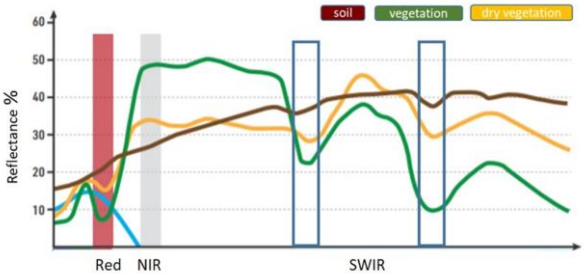


Fig. 1. Spectral reflectance of vegetation and soil with different levels of water content (yellow line: dry vegetation). The spectral bands used to calculate the indices are highlighted.

Based on RED, NIR and SWIR bands different indices are calculated to quantify plant vigour and relate it to root zone soil moisture. The soil moisture status influences the vegetation water status and thereby changes the spectral characteristics of the vegetation.

The most common vegetation index used in the water stress models (e.g.TOTRAM) is the Normalized Difference Vegetation Index (NDVI).

The water-absorption bands in the 1300–2500 nm region show the highest sensitivity to leaf water concentration in most crops. Short infrared is very sensitive to the water contained in the object whether it is soil or vegetation. NDWI (Normalized Difference Water Index) is one of the popular indicators of this kind of investigations.

B. The vegetation and water indices

During the quantitative interpretation of remote sensing information from vegetation can be created by extracting vegetation information using individual light spectra bands or a group of single bands for data analysis. Vegetation indices use various combinations of multispectral satellite data to produce single images representing the amount of vegetation present, or vegetation vigour. They can reduce the data volume for analysis and provide combined information that is more strongly related to changes than any single band. The construction of VI algorithms are effective tools to measure vegetation status. Vegetation information from remote sensed images is mainly interpreted by differences and changes of the green leaves from plants and canopy spectral characteristics. The data from near infrared (0.7–1.1 m) and red (0.6–0.7 m) or other bands are combined in different ways according to their specific objectives. As the population of plants increases, the amount of biomass causes an increase in the overall near-infrared reflectance and a reduction in the red reflectance. From previous research, there are known relationships between the indices using those two regions of the spectrum and the amount of vegetation. By calculating the Normalized Difference Vegetation Index (NDVI we can get information on the crops' vigour (Fig. 2). Low index values usually indicate little healthy vegetation while high values indicate much healthy vegetation.

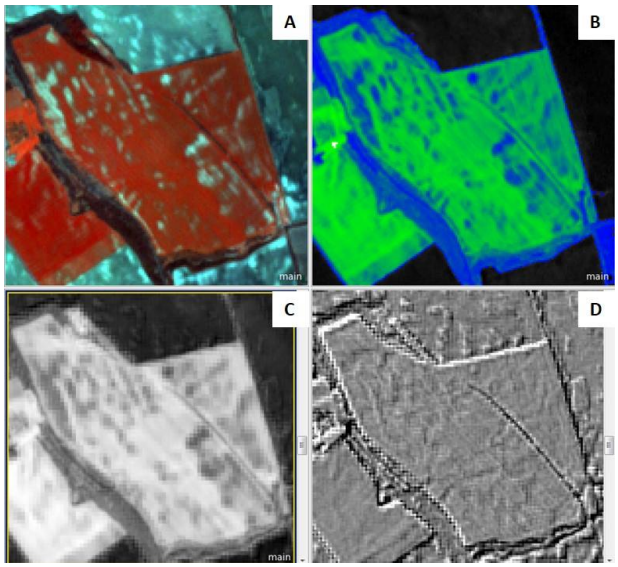


Fig. 2. A: Satellite image (Sentinel 2: fals coloure composition), B: NDVI, C: NDMI, D: NDRE (Normalised Difference Red Edge Index).

The NDWI is used to monitor changes related to water content in water bodies. As water bodies strongly absorb light in visible to infrared electromagnetic spectrum, NDWI uses green and near infrared bands to highlight water bodies. Index values greater than 0.5 usually correspond to water bodies. Vegetation usually corresponds to much smaller values.

NDWI index is often used synonymously with the NDMI index, often using NIR-SWIR combination as one of the two options. NDMI seems to be consistently described using NIR-SWIR combination (Fig. 2). As the indices with these two combinations work very differently, with NIR-SWIR highlighting differences in water content of leaves, and GREEN-NIR highlighting differences in water content of water bodies.

NDRE (Normalized Difference Red Edge) this index is more appropriate than NDVI index for intensive management applications throughout the crop growing season. It is a modification of the NDVI index, however works as a better measure of vegetation health than NDVI especially for mid-late season crops that have elevated levels of chlorophyll because VRE bands are more translucent to leaves than red light and hence is seldom completely absorbed by a canopy.

C. The optical trapezoid model (OPTRAM)

The traditional trapezoid model, TOTRAM, is based on the pixel distribution within the Land Surface Temperature-Vegetation Index space (LST-VI). An inverse linear relationship between surface soil moisture and LST is then assumed (Fig. 3).

The OPtical TRApezoid Model (OPTRAM) was developed for Soil Water Content (SWC) estimation making use of optical satellite data, and is based on the linear physical relationship between soil moisture and Shortwave Infrared Transformed Reflectance (STR) [14].

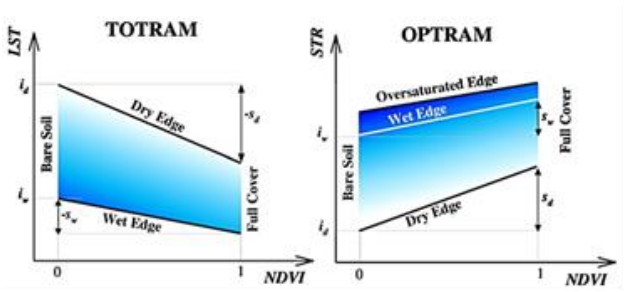


Fig. 3. Sketch illustrating parameters of the traditional thermal-optical trapezoid model (TOTRAM) and the new optical trapezoid model (OPTRAM) [14].

OPTRAM requires a parametrization at a given location based on the pixel distribution within STR-NDVI space (Fig. 3), where STR is the Shortwave Infrared Transformed Reflectance defined as follow:

$$STR = (1 - R_{SWIR})^2 / (2\,R_{SWIR}) \quad (1)$$

The model parameters can be obtained for a specific location from the dry and wet edges of the optical trapezoid depicted in (Fig. 3)

$$STR_d = i_d + sd_{NDVI} \quad (2)$$

$$STR_w = i_w + swNDVI \quad (3)$$

where STR_d and STR_w are the STR at θ_d and θ_w, respectively.

The soil moisture for each pixel can be estimated as a function of STR and NDVI:

$$W = \frac{i_d + s_d NDVI - STR}{i_d - i_w + (s_d - s_w)NDVI} \quad (4)$$

where i_d, s_d, and i_w, s_w are dry and wet edges parameters.

III. RESULTS AND DISCUSSION

The OPTRAM model Eq. (4)] was parameterized based on the pixel distribution within the STR-NDVI space. The model was run separately for each available images, also one integrated trapezoid incorporating pixel distributions from all selected images was used to try universally parameterized.

During the work dry (id and sd) and wet (iw and sw) edges were determined by visual inspection of the STR-NDVI spaces so that the trapezoids surrounded the majority of the pixels (Table 2). From id and sd (dry edge parameters) and iw and siw (wet edge parameters), the normalized moisture content, W, was estimated for each pixel with Eqs. (1,2,3,4). The results are illustrated in the Fig. 4.

TABLE II. OPTRAM PARAMETERS OBTAINED FOR THE STUDY AREA (AGRICULTURAL FIELD) BASED ON SENTINEL-2 (2018)

Data	Dry edge i_d	Dry edge s_d	Wet edge i_w	Wet edge s_w
2018.01.18	3.8	1.2	4.6	0.9
2018.04.08	0	0.5	1.8	2.9
2018.06.12	3.5	1	4.6	2.4

Pixel distributions within the STR-NDVI space for 2 images (as example) are depicted in Fig. 4. Corresponding model parameters are listed in Table II. According to the study results a trapezoidal shapes were formed by the pixels in the

STR-NDVI space in all cases. The normalized moisture content (W), was estimated for each images.

Soil moisture variability within the field are clearly detected by using this model. However, far-reaching conclusions cannot be drawn. Further research is needed. Currently, additional agricultural areas have been included in the research, and devices suitable for meteorological measurements have been placed in the area. Field measurements are absolutely necessary to continue the study and validate the results.

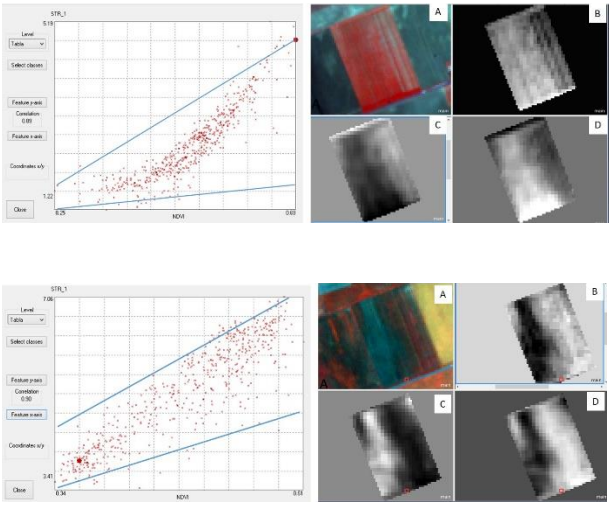


Fig. 4. Pixel distributions within the STR-NDVI space for images (2018 Apr. and Jul.) A: false colour composition, B: NDVI, C: B12 values, D: W values (OPTRAM)

REFERENCES

[1] A. Ghulam, Q. Qin, Z. Zhan, Designing of the perpendicular drought index. Environ. Geol. 52 (6), 1045–1052 (2007)

[2] Y.1. Osakabe, K. Osakabe, K.1. Shinozaki, P. Lam-Son, Response of plants to water stress Front Plant Sci., 5 p. 86 (2014)

[3] Y. Osroosh, R.T. Peters, C.S. Campbell, Q. Zhang, Automatic irrigation scheduling of apple trees using theoretical crop water stress index with an innovative dynamic threshold Comp. Electron. Agric., 118, pp. 193-203 (2015)

[4] M. González-Dugo, M. Moran, L. Mateos, R. Bryant, Canopy temperature variability as an indicator of crop water stress severity Irrigation Sci., 24 pp. 233-240 (2006)

[5] M.L. Whiting, L. Li, S. L. Ustin, Predicting water content using Gaussian model on soil spectra. Remote Sens. Environ. 89 (4), 535–552. (2004)

[6] H. Zhang, M. Han, J.L. Chavez, Y. Lan Improvement in estimation of soil water deficit by integrating airborne imagery data into a soil water balance model Int. J. Agric. Biol. Eng., 10 (3) pp. 37-46 (2017)

[7] W.W. Verstraeten, F. Veroustraete, C. J. van der Sande, I. Grootaers, J. Feyen, Soil moisture retrieval using thermal inertia, determined with visible and thermal spaceborne data, validated for European forests. Remote Sens. Environ. 101 (3), 299–314 (2006)

[8] L. Hassan-Esfahani, A. Torres-Rua, A. Jensen, M. McKee, Assessment of surface soil moisture using high-resolution multi-spectral imagery and artificial neural networks. Remote Sens. 7 (3), 2627–2646 (2015)

[9] N.N. Das, B. P. Mohanty, E. G. Njoku, Characterization of backscatter by surface features in L-band active microwave remote sensing of soil moisture. IGARSS 2008IEEE International Geoscience and Remote Sensing Symposium. 2 IEE E, pp. II–817 (2008)

[10] I.E. Mladenova, T.J. Jackson, E. Njoku, R. Bindlish, S. Chan, M.H. Cosh, T.R.H. Holmes, R.A.M. De Jeu, L. Jones, J. Kimball, S.Paloscia, Remote monitoring of soil moisture using passive microwave-based techniques—theoretical basis and overview of selected algorithms for AMSR-E. Remote Sens. Environ. 144, 197–213 (2014)

[11] R. Nemani, L. Pierce, S. Running, S. Goward, Developing satellite-derived estimates of surface moisture status. J. Appl. Meteorol. 32 (3), 548–557 (1993)

[12] T.N. Carlson, R.R. Gillies, E.M. Perry, A method to make use of thermal infrared temperature and NDVI measurements to infer surface soil water content and fractional vegetation cover. Remote Sens. Rev. 9 (1–2), 161–173 (1994)

[13] H. Feng, C. Chen, H. Dong, J. Wang, Q. Meng, Modified shortwave infrared perpendicular water stress index: a farmland water stress monitoring method. J. Appl. Meteorol. Climatol. 52 (9), 2024–2032 (2013)

[14] M. Sadeghi, S.B. Jones, W.D. Philpot, A linear physically-based model for remote sensing of soil moisture using short wave infrared bands. Remote Sens. Environ. 164, pp. 66–76 (2015)

Information Technology Safety Awareness – a review of regularly used terms and methods

Anikó Szarvák
Applied Mathematics and Applied Informatics Doctoral school
University of Óbuda
Budapest, Hungary
aniko.szarvak@nik.uni-obuda.hu

Valéria Póser
Applied Mathematics and Applied Informatics Doctoral school
University of Óbuda
Budapest, Hungary
poser.valeria@nik.uni-obuda.hu

***Abstract*— Over the last twenty years there have been enormous advances in the field of Information Technology. There is a quite well-defined need on information security and of course awareness. Problems with human behavior in an information security context are assumed to be caused by a lack of facts available to the audience. Awareness therefore is largely treated as the broadcast of facts to an audience in the hope that behavior improves.**

This study targets three different but connecting topics. First, reviews, the definition of Awareness and based on it creates an overview of different definitions. Second, reviews the information security awareness training, pitfalls and outcomes. And last but not least gives a non-technical way to reach people (end-users) and gives possible ways to change their behavior according to information security related topics. (*Abstract*)

Keywords—awareness, awareness training, information security awareness, information security, computer systems (*key words****)***

I. INTRODUCTION

There are a lot of available material about information security awareness, but these materials are not easy to adopt by organizations and not really known by individuals. Due the recent changes, employees are working from home instead of office, there are lot more threat appears which are cannot handle in usual information security countermeasures.

There are studies pointing out the lack of information security education, training, the awareness, consciousness itself. "An employee's perceived fairness of the ISP has a direct impact on her attitude and in turn intention toward compliance." "ISA plays a key role in employees' compliance behavior". [1]

"The proliferation of terms to describe similar concepts is identified as a big challenge for information security research. For example, information security education (ISE), information security training (IST) and information security awareness (ISA) have been defined differently by several researchers around the world, each with a different focus and purpose." [2] which is not helps the security personnel.

Decker [3] displays the lack of correlation between ISA, ISE and IST as a result of unclear definitions for the three concepts. To find the root cause of the problem, the first thing is a definition overview. There are a lot of definition on Awareness, specially Information Security Awareness.

II. DEFINITION REVIEW

Dictionary definition of awareness: knowledge or perception of a situation or fact. Knowledge that something exists or understanding of a situation or subject at the present time based on information or experience [4]. This can help to determine the definition and add a subject of the awareness.

TABLE I. AWARENESS RELATED ABBREVATIONS

Abbreviations	Expositions, Terms
	Used in this study
IT	Information Technology
ICT	Information and Communication Technology
ISA	Information Security Awareness
GISA	General Information Security Awareness
ISP	Information Security Policy
SAT	Security Awareness Training
SETA	Security Education, Training and Awareness

[a.] Table 1.Awareness related abbreviations)

We assume there are a general common understanding on we need awareness while users are enabled to use information technology both at work and at home. Reviewing the definitions, it is not easy to point out which awareness, education or training people need.

A. Definition of awareness

From different point of view, Security Awareness is defined differently, described with additional significant content. Wolf et al says "Security awareness is the effort to impart knowledge of or about factors in information security to the degree that it influences users' behavior to conform to policy" [4].

Security awareness is defined in NIST Special Publication 800-16 [5] as follows: "Awareness is not training. The purpose of awareness presentations is simply to focus attention on security. Awareness presentations are intended to allow individuals to recognize IT security concerns and respond accordingly" Based on this definition Security Awareness is to recognize and to respond an event in the field of information security.

Siponen's definition on Security awareness: Security awareness is a state in which users in an organization are aware of and are ideally committed to the security objectives of their organizations. [6]

According to Information Security Forum (ISF) [7], "security awareness is a continual process of learning by which, trainees realize the importance of information security issues, the security level required by the organization, and individuals' security duties"

B. Information security awareness

Information Security Awareness: "ISA is an employee's general knowledge about information security and his cognizance of the ISP of his organization" [8] In this definition ISP is the Information Security Policy. In general, if we view the topic of information security awareness from a more nature way, we have to have a statement that there is no information security policy for general internet home usage.

Connecting to this topic, a more proper definition is added by Siponen: "ISP Awareness is defined as an employee's knowledge and understanding of the requirements prescribed in his organization's ISP and the aims of those requirements" [6]

Information security awareness can be defined as the combination of a person's knowledge of security concepts and the person's awareness or consciousness of the existence of any security related policies. [9]. According to the review of these publications we see this definition is the must proper.

C. General Information Security Awareness

Based on the statement of Information security awareness, General Information Security Awareness, hereinafter GISA, should be the definition of information security awareness for general users, but Bulgurcu et al.'s definition not reflects it. They define "General Information Security Awareness is an employee's overall knowledge and understanding of potential information-security-related issues and their ramifications, and what needs to be done in order to deal with security-related issues" [1].

D. Technology awareness

Additionally, we found a different approach of the awareness definition which comes from the technology perspective. Technology awareness as a "user's raised consciousness of and interest in knowing about technological issues and strategies to deal with them" [10].

E. Definition of information security awareness

"Security awareness contains two equally important pieces. The first piece is the dissemination of accurate, current and appropriate knowledge of policy to individuals. Policies explain to individuals the threats they need to be cognizant of as well as the appropriate actions to take upon encountering a threat. The second portion of awareness is the delivery of policy in a manner that convinces an individual to change his or her behavior. These two portions are equally important; without one, the other is ineffective." [4]

In Al-Daeef et al's review of these definition says the following: "Previous definitions have highlighted three key components of security awareness, they are; ongoing or continual process, knowledge delivery method, and individuals' behavior impact. Based on that, information security awareness has defined as the security knowledge that has been gradually acquired through a continuous and updated catchy training manner to influence trainees' behavior." [12]. We agree with this conclusion of their study. they defined the security awareness as a independent concept from ISP.

Information security awareness – ISA – knowledge, desired behavior is based on the Information security Policy – ISP – which is designed and shaped for an organization. It may contain general information and widely accepted and implemented desired behavior items, but it is consistently fitted for the organization. For example, if an organization not allows using for users / employees their own equipment for work, BYOD – and its security countermeasures – is not described in it.

Bulgurcu et al. come to the conclusion in theirs study that "ISP awareness can be different from general ISA" [1] which opens the possibility that there have to be a good definition which is able to define the information security awareness properly and independently.

Taken into consideration that every definition above is defines a piece of Awareness related to information security, information security policy, general information security we suggest splitting these into two separate definitions. One is related to compliance of an organizations' ISP and one is general, which is not related to organizations, policies. This is close to the dictionary version of Awareness, in general and relates to information technology safety.

An independent definition of Information Technology Safety Awareness is knowledge or perception of a situation or a fact in Information Technology field or related area. Based on this definition, awareness is independent from any information security policy but relying to information technology. In this definition information technology (IT) is the use of computers to store, retrieve, transmit, and manipulate data [13] or information. IT is a subset of information and communications technology (ICT). IT is including, but not limited to personal computers, mobile handheld devices, tablets, wearables and servers.

With this definition we are able to separate and distinguish a general knowledgebase from a highly shaped Information Security Policy and its expectations. Narrowing the knowledgebase, we can implement an education model, qualitative and quantitative measurement model and we are able to specify trainings to enhance the education.

F. Definition of security education security training

Security training in addition, considered as an important protection approach by security standards of International Organization for Standardization (ISO) [14], and National Institute of Standards and Technology (NIST) [15].

In [16] for example, A Security Education, Training and Awareness (SETA) was defined as an educational program that aim to reduce security breaches that caused because of the lack of employees' security awareness. SETA was designed to educate employees how to focus on security issues to protect themselves and their organization's data and network. SETA program integrates security in all tasks that employees do; from locking computer screens when they move away from their desks; to report unusual activities regarding to emails, files and staff.

In Anandpara et al's experiment for example did not improve participants' ability to differentiate between phishing and legitimate emails, it however has made them more suspicious (i.e. more aware) [17].

In some cases, education and training are used as synonyms, such in [12] but there is a quite well described difference between the two definition. Education is "the process of teaching or learning the knowledge, especially in a school or college". Training is "the process of learning the skills you need to do a particular job or activity". [1]

Based on this, education on information technology safety may lead a better understanding of SAT, SETA. Building on

the knowledge of the term's vocabulary can lead to better understanding of SAT, SETA projects – which are based on organizations ISA may lead better compliance to it.

Prior to the findings above Nurse et al.'s study formulates "One of the main reasons why users do not behave optimally is that security systems and policies are poorly designed" [18]. At this point we do not agree it with fully. Maybe there are another significant component causes the failure and the low effectiveness of ISP and aligning to it.

III. Technology or Human behaviour

In the study made by Bada et al. finds "Behavior change in a cyber security context could possibly be measured through risk reduction, but not through what people know, what they ignore or what they do not know. Answering questions correctly does not mean that the individual is motivated to behave according to the knowledge gained during an awareness program." [19] which can be an indicator on not only poorly designed and implemented ISP is the matter.

There are two different point of view established in past years in connection with information security awareness and its necessity. First is focusing on technology and driving there is no need on education and training of awareness in higher level. We do not agree with it at this point, because some kind of awareness, specially information technology safety is highly recommended. With this as a baseline common understanding of users or employees, based on risk appetite, it can be completed with technical controls, focusing on the security of the employer's property. But most of these controls are not effective in the employee's personal life and time. "It is recommended that when possible to use hardware or software to implement and enforce policy. The money that is saved by not repeating ineffective awareness messages can be redirected to those awareness areas that are not enforceable by hardware or software. It is also recommended that behavioral-based measurements used to determine compliance." [4]

The second point of view enhances the security awareness and pointing out that awareness is not only matter of knowledge, but behavior, motivation, emotion, interest and culture impacted too. In Al-Omari, et al.'s study the first to address the role of users' general knowledge of information security issues on their attitude to comply with ISPs. The result suggests that an employee's attitude toward compliance with ISP can be enhanced by his/her general security awareness [10] Also they found "that creating security-aware culture within the organization will shape users' attitude and behavior to be more security-conscious".

The real problem with awareness knowledge, education and training is defined in an ISF article: "People know the answer to awareness questions, but they do not act accordingly to their real life" and "Simple transfer of knowledge about good practices in security is far from enough." [11]. Based on these quotes we assume the root of the problem can be highlighted. These problems can be:

- Not properly defined goals of the trainings, materials,
- No base ground defined for the awareness,
- Too narrow or too wide curriculum,
- Not a uniform set of basic knowledge from the audience and
- Not well-defined common understanding on the topic.

Without these prerequisites there is a small chance for the effective training. Findings in support of these can be found in recent studies, like Bada et al.'s article about cybersecurity awareness: "The fact today is that security awareness as conceived is not working. Naturally, an individual that is faced with so many ambiguous warnings and complicated advice, may be tempted to abandon all efforts for protection, and not worry about any danger. Threatening or intimidating security messages are not particularly effective, especially because they increase stress to such an extent that the individual may even be repulsed or deny the existence of the need for any security decision." [19]

A. Information security vocabulary

A new Pew Research Center survey finds that Americans' understanding of technology-related issues varies greatly depending on the topic, term or concept. While a majority of U.S. adults can correctly answer questions about phishing scams or website cookies, other items are more challenging. For example, just 28% of adults can identify an example of two-factor authentication – one of the most important ways experts say people can protect their personal information on sensitive accounts. Additionally, about one-quarter of Americans (24%) know that private browsing only hides browser history from other users of that computer, while roughly half (49%) say they are unsure what private browsing does. [20]

According to the finding above the Pew Research Center pointed out in a follower article: "the typical (median) respondent answered only five of these 13 knowledge questions correctly (with a mean of 5.5 correct answers). One-in-five (20%) answered more than eight questions accurately, and just 1% received a "perfect score" by correctly answering all 13 questions." [21] This may refer there is a lack of basic knowledge of technical vocabulary. Based on this Chan et al.'s finding can be seen in a different way: "Employee awareness levels for the organization are surprisingly low, with employees lacking in both knowledge of concepts and awareness of the organization's policies. The lack of information security awareness was reflected in the admission by many employees that they had previously engaged in behaviors (such as password sharing) which were in direct violation of information security. This may be a result of the lack of policy promotion and enforcement." [9] If the vocabulary or the understanding of the topics are not given to the people – not only for employees – we can not expect consciousness on the security.

IV. A reaction based case

"The tendency for technical experts in the field of information security to tell people what they think they ought to know (and may in fact already know) needs to be recognized as a failure. A common theme from the safety approach is that to achieve effective and efficient communications it is critical to understand the relevant beliefs of the audience. It is not enough to know what behaviors exist that are causing information security risk. Communicators must understand why the behavior is occurring which requires an understanding of an audience's constraints and supporting beliefs." [22] In this study the authors highlighting the Extended Parallel Processing Model [23] which introduces reactions based on fear.

Independently from the studies and researches above, during an information security awareness campaign at an organization we sent out e-mail communication to end-users to raise their awareness on cybersecurity and information security policy. We used the SANS monthly newsletters on Awareness called OUCH! [24]. In first time, the communication of the risks, countermeasures are lay on technical facts. There was no feedback from the end-users of the organization or just a few.

In a case when we sent out a newsletter about ransomware and its capabilities of destruction flavored emotionally, we got feedbacks from our end-users, full of understand of the risk and caution – awareness – on the topic. They highlighted they are touched by the message and they are felt the situation. The message sounded like this: "You can see kidnappings in movies, where the detainees demand a ransom. Imagine that this could happen, kidnapping pictures of your loved ones, your family, from your computer."

Making connections with strong emotions – like fear, love – during communication opened a possibility of an efficient way to enhance the overall information security awareness campaign. This underlines the viewpoint: For information security awareness techniques to improve in effectiveness and efficiency it is clear that information security awareness content can't be created from what the technical experts want to tell people, the contents of a technical standard or the prevailing best practice topics but rather the audience themselves that it seeks to influence and protect. [22]

V. Conclusion

There is a need to separate ISP based ISA from the overall basic Information Technology Security Awareness. An independent definition of Information Technology Safety Awareness is knowledge or perception of a situation or a fact in Information Technology field or related area.

There is a need on education and training of information technology safety where education is "the process of teaching or learning the knowledge, especially in a school or college". Training is "the process of learning the skills you need to do a particular job or activity". The education and the training must be shaped to people and must contain emotional charge instead of dry technology facts.

Based on this, education on information technology safety may lead a better understanding of SAT, SETA. Building on the better understanding, the SAT, SETA projects – which are based on organizations ISA may lead better compliance in information security field.

With this definition we are able to separate and liberate a general knowledgebase from a highly shaped Information Security Policy and its expectations. Narrowing the knowledgebase, we can implement an education model, qualitative and quantitative measurement model and we are able to specify trainings to enhance the education.

Additionally, the awareness message can be delivered to the end-users, the individuals flavored with emotions can help to get their attention and enables them to enhance consciousness.

VI. Future work

We assume there is a need to develop a baseline for information technology safety awareness and create a reference questionnaire to measure the knowledge and emotional base. To complete this, we are going to create a vocabulary baseline and a thematic education material to raise the knowledge base. With this as an education reference there is a good chance to create or maintain ISP based SAT to enhance employee's engagement to comply the recent security awareness requirements.

VII. Acknowledgement

Authors gratefully acknowledges the support of the Applied Mathematics and Applied Informatics Doctoral School on this research.

References

[1] Bulgurcu, B., Cavusoglu, H., & Benbasat, I. (2009). Roles of Information Security Awareness and Perceived Fairness in Information Security Policy Compliance. Paper presented at the AMCIS 2009 Proceedings. Paper 419.

[2] Hussain Aldawood, Geoff Skinner: Educating and Raising Awareness on Cyber Security Social Engineering: A Literature Review 2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE) p62-68 978-1-5386-6522-0/18 ©2018 IEEE

[3] L. Decker, "Factors affecting the security awareness of end-users: A survey analysis within institutions of higher learning," PhD dissertation, School of Bus. & Technol., Cappella Univ., Minneapolis, MN, 2008.Cambirdge Dictionary Online: Awareness, Education and Training, https://dictionary.cambridge.org/dictionary/english/ Accessed: 2020. 08. 21.

[4] Wolf, M., D. Haworth, and L. Pietron, Measuring an information security awareness program. Review of Business Information Systems (RBIS), 2011. 15(3): p. 9-22.

[5] National Institute of Standards and Technology - NIST: Building an Information Technology Security Awareness and Training Program. Wilson, M. and Hash, J. Computer Security Division Information Technology Laboratory. October 2003. http://csrc.nist.gov/publications/nistpubs/800-50/NIST-SP800-50.pdf

[6] Siponen, M. 2000. 'A Conceptual Foundation for Organizational Information Security Awareness'. Information Management and Computer Security, 8(1): 31-41.

[7] Information Security Forum (ISF).: The Standard of Good Practice for Information Security, Security Standard. 2007.

[8] Cavusoglu H., Cavusoglu H., Son J.-Y. and Benbasat I. 2008. 'Information Security Control Resources in Organizations: A Multidimensional View and Their Key Drivers'. UBC Working Paper.

[9] Hong Chan, Sameera Mubarak Significance of Information Security Awareness in the Higher Education Sector International Journal of Computer Applications (0975 – 8887) Volume 60– No.10, December 2012 [6] Al-Omari, Ahmad; El-Gayar, Omar; and Deokar, Amit, "Information Security Policy Compliance: The Role of Information Security Awareness" (2012). AMCIS 2012 Proceedings. 16. http://aisel.aisnet.org/amcis2012/proceedings/ISSecurity/16

[10] Dinev, T., & Hu, Q. (2007). The centrality of awareness in the formation of user behavioral intention toward protective information technologies. Journal of the Association for Information Systems, 8(7), 23.

[11] Information Security Forum (ISF): From Promoting Awareness to Embedding Behaviours, Secure by choice not by chance, February 2014. https://www.securityforum.org/shop/p-71-170

[12] Melad Mohamed Al-Daeef, Nurlida Basir, Madihah Mohd Saudi Security Awareness Training: A Review Proceedings of the World Congress on Engineering 2017 Vol I ISBN: 978-988-14047-4-9 pp446-451

[13] Daintith, John, ed. (2009), "IT", A Dictionary of Physics, Oxford University Press, ISBN 9780199233991, retrieved 1 August 2012

[14] ISO/IEC . ISO/IEC 27001:2005 - . Tech. rep., International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). 2005.

[15] NIST, Nist special publication 800-12, An introduction to computer security: the NIST handbook. 1995: DIANE Publishing.

[16] Hight, S.D., The importance of a security, education, training and awareness program, November 2005.

[17] Anandpara, V., et al., Phishing IQ tests measure fear, not ability, in Financial Cryptography and Data Security. 2007, Springer. p. 362-366.

[18] Nurse, J.R.C., Creese, S., Goldsmith, M., Lamberts, K.: Guidelines for usable cybersecurity: Past and present, in The 3rd International Workshop on Cyberspace Safety and Security (CSS 2011) at The 5th International Conference on Network and System Security (NSS 2011), Milan, Italy, 6-8 September.

[19] Maria Bada, Angela M.Sasse and Jason R. C. Nurse: Cyber Security Awareness Campaigns: Why do they fail to change behaviour?

[20] M. Anderson and E. Vogels, ''Americans and digital knowledge,'' Pew Res. Center, Washington, DC, USA, Tech. Rep., 2019. [Online]. https://www.pewresearch.org/internet/2019/10/09/americans-and-digital-knowledge/

[21] K. Olmstead and A. Smith, ''Americans and cybersecurity,'' Pew Res. Center, Washington, DC, USA, Tech. Rep., 2017. [Online]. https://www.pewresearch.org/internet/2017/03/22/what-the-public-knows-about-cybersecurity/

[22] Stewart, Geordie; Lacey, David: Death by a thousand facts: Criticising the technocratic approach to information security awareness, Information Management & Computer Security, Volume 20, Number 1, 2012, pp. 29-38(10) ISSN: 0968-5227

[23] Delaney, A, Lough, B, Whelan, M and Cameron, M: A Review of Mass Media Campaigns in Road Safety Monash University Accident Research Centre 2004

[24] https://www.sans.org/security-awareness-training/ouch-newsletter

Methods of functional measurement of software

Katalin Nátz*, Tamás Orosz**, Zsigmond Gábor Szalay*
* Szent István University, Budapest, Hungary
** Óbuda University/AMK, Székesfehérvár, Hungary
* natz.kata@gmail.com
**orosz.tamas@amk.uni-obuda.hu
szalay.zsigmond.gabor@szie.hu

***Abstract* — Nowadays software development is the most expensive element of a projects in the information technology sector. Many software projects fail due to budget overruns, late delivery or incorrectly planned forecasts.**
Software products could be manufactured if they have such frameworks that the budget can be constrained. Before developing a software product, it is important to estimate the cost of software development.
The estimation of software costs - the so-called software cost estimation model should predict the effort that is indispensable for building a software system. In the last 40 years, many estimation models have been composed for the following aspects: budgeting, risk analysis, project planning and control, and investment analysis for software improvements. Software provides the IT landscape for business processes and enables the reusability of the software.
There are numerous ways of measuring the functional size of the software: COCOMO (Constructive Cost Model) and ESTIMACS based on the parameters of implementation ability, flexibility and traceability as well as techniques for software cost estimation. Index terms - Function points (FP), Cosmic, SLOC (Lines of code), Use Case Points (UCP) etc.

***Keywords*: Function point, Story point, Use Case point, Software measurement, Business processes, Financial planning**

I. INTRODUCTION

Applications and large integrated systems such as enterprise information and cloud-based systems are evolving very rapidly, and the number of users associated with them is also growing. These applications work with so much data and are so complex that it is difficult for the average user to understand how they work. New coding tools are already enabling software developers to meet ever-increasing and increasingly complex customer needs more efficiently and quickly. An understandable, comparable and clear method is needed to interpret, estimate and monitor projects [12].

The purpose of FSM (Functional Size Measurement) methods is to determine the size of software by quantifying the totality of functions provided to users. Functional point analysis (FPA) was the first FSM method [1][3][9]; it was originally introduced in the mid-1970s. Today, businesses around the world use this methodology. Allan Albrecht, of IBM, was the first to develop a method to assess the limit and extent of measurement. Its name is forever intertwined with the issue of software functionality. However, one of the main results is that it made the productivity of software projects measurable using the method [4]. Since its founding in 1986, the International Function Point Users Group (IFPUG) has continuously developed and updated the original Albrecht methodology for software (IFPUG Counting Practices Manual) [3][4]. In 2009, a new version of the IFPUG Function Point (FP) standard, CPM 4.3.1, was released and has been in use since early 2010 [4].

The appearance of the FPA technique has enabled the ICT community to facilitate the practice of measuring software functionality in relation to the "line of code" (LOC) approach [5]. However, all this requires clear, complete and detailed descriptions, application and design documents [2].

In one of their research, Meli and Santillo mention two cases where it would be very important to set up an alternative estimation method that is compatible with FP standard rules. One case came to the fore when a development project was at such an early stage when it was simply not yet possible to calculate FP according to IFPUG standards [18].

The second case was encountered when the detailed documentation for the measurement of the existing software, which should have been prepared during the development of the project, was not available [2].

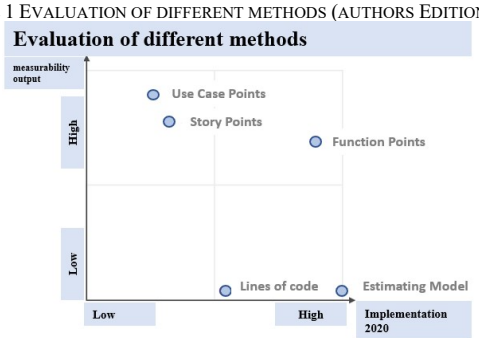
The demand for methods to estimate function points has been increased for organizations involved in the software business. This need is becoming increasingly indispensable in agile software development, which need usable and cost-effective software [15].

II. METHODS FOR MEASUREMENT

During software development, it is not easy to decide which method would be applicable to both traditional and agile and mixed-development software.

There are several methods for measuring emissions (Table 1), which must meet the following conditions: the method of measurement used in the future must be independent of the type of software or the specific technology [16].

TABLE 1 EVALUATION OF DIFFERENT METHODS (AUTHORS EDITION)



The functional size of the developed software can be determined, for example, by the amount of source code. Today, however, thanks to modern programming languages and the development environment, more and more functions can be developed with less and less source code, so the Lines of Code method did not meet the conditions. Within the company, several methods have been used for years, which are as follows:

- Story Point
- Use Case Points
- Lines of code
- COCOMO
- Function Point

III. STORY POINT

In agile software development, the extent of performance is usually determined by user stories, so-called it is described with a User Story and the magnitude of the related development is estimated in the "Story Point" unit.

The Story Point, which also determines the complexity of the User Story, is estimated within a given development team (and not always by a score, but by, for example, T-shirt size - S, M, L, XL, XXL). Of course, comparability of different teams is not possible, but it is not a goal. Within the company group, the projects are approx. 40-55% do not use the agile method.

IV. USE CASE POINTS

The Use Case Points method is an estimation method that starts from the function point method and forms a bridge to the object-oriented world; however, Use Case Points and function points cannot be mapped directly from each other. In the case of the UCP method, the Use Cases and the so-called actors are measured, and the associated transactions are classified. For each Use Case, the complexity of the associated transactions is also determined.

To be able to use the UCP method in general in all areas of the company, it would require UML modeling and comprehensive implementation of Use Cases, however, this is impossible due to the current corporate structures and their operation. However, if it did, it would be the most cost-effective method in the long run - UML can be used to automate the process.

V. LINE OF CODE

Lines of codes (SLOC) are software metrics that can be used to measure the functional size of an application by counting the number of lines that can be found in the text of the program's source code. SLOC is one of the most widely used sizing metrics in telecommunication sector and otherwise used to forecast the effort required to develop an application and to estimate productivity or maintainability after the software has been designed. There are some disadvantages to Line of Codes metrics that effectively operates the quality of the software because the output of SLOC metrics is used as input to other software estimation methods such as the COCOMO model [14][15].

VI. COCOMO

The COCOMO model (Constructive Cost Model) offers direct estimates of the effort. This is a simple model based on inputs related to the size of the system and several cost drivers that affect productivity.

COCOMO is a summary of three models: a base model that is applied at the beginning of the project, an intermediate model that is applied after the requirements are specified, and an advanced model that is applied after the design is completed [16]

VII. FUNCTION POINT ANALYSIS (FPA)

The function point, as a unit of measure, determines the functional size of the applications or software that are created or modified or deleted during the prepared development, including the functions supported by the IT system beyond the measurement boundaries. However, of the functions developed, only those that are ordered by the client can be measured.

The FPA method is suitable for measuring entire applications, projects or releases. The method is independent of the type of application or development. It can also be called a structured problem-solving algorithm, which breaks down systems into smaller components in order to make the technical description and implementation of a given project more understandable and analyzable.

A. Basics of FPA

The user view is decisive for the determination of the functional size, if one wants to determine the function points. There are special rules in the standard for assigning the point value to transactions and databases. Regarding to the number of elementary functions and the standardized values, the assigned FPs is accumulated.

In the analysis, the elementary processes are the smallest unit that has a meaning for user and constitutes a complete transaction. It is counted as the same compared with another one if it requires 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][13].

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, platform 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 [11][13].

C. *Transactional functions*

Transactions are the function that the user provides to process application data. There are three transactions - external input (EI), external output (EO) and external inquire (EQ) (Table 2)

External Inputs: The main purpose of an input is to maintain one of several internal databases or to change system behavior. To process technical data or control information, an input contains processing logic that goes into the application beyond the limit value [4][9].

External Outputs: The output provides 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.

D. *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].

E. *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 OF EI, EO, EQ (AUTHORS EDITION)

EI		DET		
		1-4	5-15	>=16
FTR	0-1	3	3	4
	2	3	4	6
	>=3	4	6	6

EO		DET		
		1-5	6-19	>=20
FTR	0-1	4	4	5
	2-3	4	5	7
	>=4	5	7	7

EQ		DET		
		1-5	6-19	>=20
FTR	0-1	3	3	4
	2-3	3	4	6
	>=4	4	6	6

Both files include Data Element Type and / or Record Element Type. Both files contain the data element type and / or the data record element type. A data element type (DET) is a unique, user-recognizable, non-repeated attribute or field of an ILF or EIF (Table 3). A record element type (RET) is a user-definable subset 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 ILF, EIF (AUTHORS EDITION)

ILF		DET		
		1-19	20-50	>=51
RET	1	7	7	10
	2-5	7	10	15
	>=6	10	15	15

EIF		DET		
		1-19	20-50	>=51
RET	1	5	5	7
	2-5	5	7	10
	>=6	7	10	10

VIII. CONCLUSION

There are several methods for determining the functional size of applications. It is always the situation that decides which method the company chooses. If UML models are available, it is almost clear that the Use Case Point method is worthwhile, as it can automate the process. If the complete documentation is available, the function point analysis can be used to obtain results.

Nonetheless, 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.

Regarding the multinational companies, FPA method is perhaps the easiest way to measure applications and there is the possibility to figure out automation. Unlike other methods, is a more tangible methodology for many companies [16].

To conduct a Function Point Analysis, it will be necessary to use the functional requirements provided by the project team. We differentiate between transactional functions and data functions. After identifying the transactional functions and data functions we must determine the complexity and the functional size for each. Regardless of the implementation 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].

According to the measurement results, it can be summarized that the FPA method is not 100% accurate, the subjective factors cannot be disregarded. The smaller a project is, the greater the deviation and variability of the FPA calculation is due to the characteristics of the method. The reason can be the lack of weighting and granularity. Therefore, a modified model should be built where the KPI -s are refined after weighting. According to the new trends, the agile method is to be enriched, since the agile methodology was introduced in the near past.

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] Mahir Kaya, Onur Demirörs: E-Cosmic: A Business Process Model Based Functional Size Estimation Approach, 2011 37th EUROMICRO Conference on Software Engineering and Advanced Applications, DOI:10.1109/SEAA.2011.60

[3] The Object Management Group: Automated Function Points, Publisher OMG, 2014

[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, Software Development Magazine, 2005

[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. School of Computing and Mathematical Sciences Liverpool John Moores University, UK, 1998.

[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; MTA Információtechnológiai Alapítvány (2003), DOI: 10.13140/RG.2.2.12580.68484

[12] Ömüral, Demirörs: Effort estimation methods for ERP projects based on function points; DOI: 10.1145/3143434.3143464

[13] K.Nátz, T. Orosz: Function point analysis by an SAP application, in AIS 2018: 13th International Symposium on Applied Informatics and Related Areas

[14] K. Bhatt1, V.Tarey, P. Patel: Analysis Of Source Lines Of Code(SLOC) Metric, International Journal of Emerging Technology and Advanced Engineering (ISSN 2250-2459, Volume 2, Issue 5, May 2012)

[15] S. Bhatia, J. Malhotra: A survey on impact of lines of code on software complexity, published in 2014 International Conference on Advances in Engineering & Technology Research (ICAETR - 2014), DOI: 10.1109/ICAETR.2014.7012875

[16] M.Alkoffash, A. Alrabea: Which Software Cost Estimation Model to Choose in a Particular Project, Journal of Computer Science 4 (7): 606-612, 2008, DOI: 10.3844/jcssp.2008.606.612

[17] D. Hallman: Die COCOMO-Modelle im Licht der agilen Softwareentwicklung,in BAMBERGER BEITRÄGEZUR WIRTSCHAFTSINFORMATIK UND ANGEWANDTEN INFORMATIK, 2018, https://doi.org/10.20378/irbo-53211

[18] R. Meli and L. Santillo, "Function Point Estimation Methods: Comparative Overview", FESMA '99 Conference proceedings, Amsterdam, October 1999.

Efficiency Improvement in Infrastructure Migration Projects through Communication, Scheduling and Control

Ágnes Bártfai
Doctoral School of Applied Informatics and Applied Mathe*matics*
Óbuda University
Budapest, Hungary
bartfai.agnes@uni-obuda.hu

***Abstract*— The article introduces a very interesting example of improving the project efficiency via simple project management tools communication and control. In this case study the project manager takes over the project with delay. The execution of the project scope within timing and budget seems to be "mission impossible". We will see if we apply the appropriate planning, significantly improve communication and apply continuous control we can close a successful project.**

Keywords—project management, infrastructure migration, communication, planning, scheduling, control

I. INTRODUCTION

In my previous article "How to turn issues into project success" I reflected the wider definition of the project success. We will see in this article a similar example how to win from a loser position. This time I would like to highlight the importance of the appropriate application of three project management tools: communication, scheduling and control.

II. THE IRON TRIANGLE

In the modern corporate landscape, a project is typically "bound" or constrained by three elements, which may be expressed in different ways. The triple constraint theory, also called the Iron Triangle in project management, defines the three elements (and their variations) as follows:

- Scope, time, budget
- Scope, schedule, cost
- Good, fast, cheap

While the names of the three elements of the triangle may change, they all measure essentially the same thing: a fixed budget, a fixed schedule or timeline, and a fixed set of expectations or deliverables.

The triangle comes into play when something affects one of its "legs." If that happens, you may need to adjust one or both of the other elements to accommodate the change. For example, if a client suddenly shortens a time frame, then a project will likely need more resources, or perhaps a scope reduction.

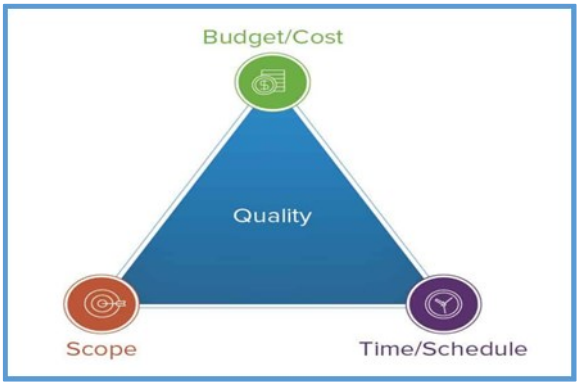


Figure 1: The project management triangle of scope, time and cost

The Iron Triangle can be considered the model of the hard skills of program and project management. **What makes these hard skills effective is the communication around the Iron Triangle topics.**

III. COMMUNICATION

A. The information transfer model

Shannon was an American mathematician, Electronic engineer and Weaver was an American scientist who worked for Bell Telephone Labs in the United States. Dr. Shannon was working on methods to decrease the noise on a telephone call while increasing the clarity of the person speaking on the phone.

In 1948 both of them join together to write an article in "Bell System Technical Journal" called "A Mathematical Theory of Communication" and also called as "Shannon-Weaver model of communication".

This communication model is the information transfer model. The information transfer model is also known as the transmission model.

Shannon and Weaver's original mathematical model contains five elements:

1. An information source, which produces a message.
2. A transmitter, which encodes the message into signals.

3. A channel, for which signals are adapted for transmission.

4. A receiver, which reconstructs the encoded message from a sequence of received signals and decodes it.

5. An information destination, which processes the message.

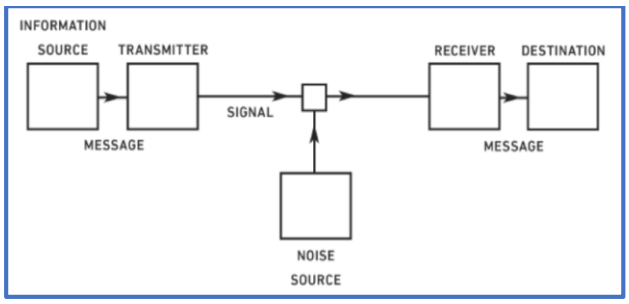


Figure 2. The information transfer model

This is a very simple and general model, also called as a mother of all communication models. Tt is effective in reducing the noise in the communication channel while intensifying the signal. The transmission model of communication describes communication as sending and receiving information.

The biggest issue with the Shannon Weaver Model of Communication is that when we transfer information, when we transmit that communication, we're assuming that understanding has occurred.

In the real project management life we really need to rethink about when we do communicate, has understanding occurred?

B. The emergence model of communication

The second communication model is the emergence model created by Dr. Nicholas Luhmann in 1934. Dr. Niklas Luhmann was a prolific writer of many books, report articles, and lectures. He attributed his creative success to the creation of his Zettelkasten, which we can almost consider an early version of the Internet. He would take notes and connect them in a specific numbering scheme. Using his numbering scheme, he could connect ideas and thoughts into new insights and theories. The Zettelkasten is the inspiration for Dr. Luhmann's emergence model of communication - meaning emerges from combining pieces of communication with each other.

The emergence model of communication is similar to the information transfer model because both models deal with transmitting information from the sender to the receiver. The difference between the two models concerns understanding.

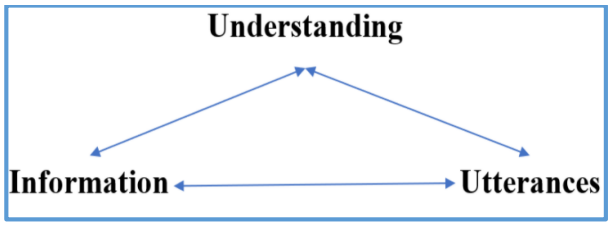


Figure 3. The emergence model of communication

On the left-hand side of the triangle is the information to be exchanged. On the right-hand side of the triangle are utterances. Utterances refer to how the information is encoded, transmitted, and decoded. At the top of the triangle between information and utterances is understanding. According to Dr. Luhmann, understanding emerges between the parties involved in the communication. Understanding is not wholly held by one communication partner and then transmitted to the other communication partner. Both partners work together to create understanding.

IV. THE IMPORTANCE OF SCHEDULING

When I do not know where to go how can I get there? The definition of the project scope gives me the goal what I have to achieve as a project manager with my team. The key element of the how is the work breakdown structure. The WBS contains all the work that must be completed in order to deliver the project scope. The basic element of the work breakdown structure are activities. The activities describe the actual tasks. They generally begin with a verb in the project plan. The group of activities create work packages. There is an important rule of creating activities: one resource should complete an activities within 5-10 days.

In the project plan we should identify Milestones. A Guide to the Project Management Body of Knowledge (PMBOK® Guide), Project Management Institute, 2017, Page 711. definition is the follwoing: Milestones are used to identify "a significant point or event in a project." They are used to identify the completion of key deliverables or the completion of key events in a project. The milestone has no cost, no resource and no duration.

The WBS contains the detailed list of the work to be completed. In many cases the tasks are not independent from each other, some activity must finish before the next can start. In this logical sequence the first task is the predecessor of the next task and should be completed first. Than the the successor task follows it.

Costs are allocated to the project at the activity level and includes all costs associated with the completion of each activity. By linking costs to the activities in the project, we can establish a time phased view of the budget for the project. The Budget plan and continuous control belong to the project manager responsibility – as shown in the Iron Triangle. The introduction of budget planning is not my intention in this article as I would like to focus on the WBS.

If we determine the order of the activities, the required resources and the duration (how long) each activity should take to complete: the outcome will be the project schedule.

Identify applicable funding agency here. If none, delete this text box.

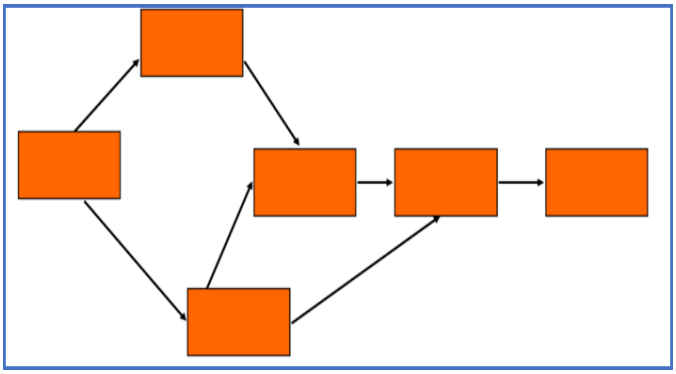


Figure 4. Network diagram in scheduling

The creation of the project schedule is usually done by a network diagram. It is a visual depiction of the flow of work (activities) in order to complete the project.

The Critical Path Method (CPM) is a common network diagramming method used to determine the project schedule. The PMBOK® Guide (Project Management Institute, 2017, Page 704.) provides us with the following definition: CPM is "a method used to estimate the minimum project duration and determine the amount of scheduling flexibility on the logical network paths within the schedule model.". It is also a common method used in scheduling software tools such as Microsoft® Project. Other scheduling tools are: CPM involves identifying which activities are critical and must be completed on schedule or it will delay the entire project.

The documentation of the project schedule is the task and responsibility of the project manager, but it does not mean that it needs to be done alone. In my opinion the best practice of creating the project schedule is a cooperation with the technical or execution team. It is usually feasible in case of small projects usually in the BAU business as usual phase of the migration project.

The project schedule is usually is a forecast but in this case it can be very punctual as the plan is based on personal commitments. The project manager as a leader has the task to help the team members to execute their job without any issues.

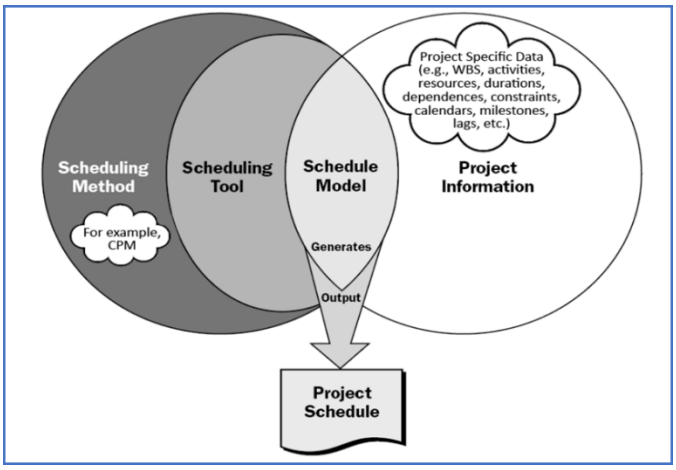


Figure 5. Scheduling Overview

The timely project execution will depend on the dedication of the team members. During our teamwork we can apply agile tools: Mural, Trello, Bluesight and a very open communication. In case of small project the schedule can be created based on collaboration. This means that the project manager puts the schedule together based on personal commitments. When the project is executed in a matrix organization, where the resources are shared the project manager has an additional task: support the team members in fixing their resource problem. For example talk with their manager.

The project manager has to consider the risk of delay and somehow a certain buffer time should be built in the project schedule as well. There are some cases when it is not possible at all – for example in the case study below. In such a case the communication is essential during the teamwork: the technical subject matter experts should explain very detailed why it is not possible to execute the tasks within the requested date, what prevent them from the timely execution. Once we put the plan together, the SMEs should inform the PM immediately in case of any problem, issue that can cause or caused delay. If the project schedule should be changed than the PM should inform the impacted stakeholders immediately and should adjust the schedule immediately. This situation can be stressful for the stakeholders – and this brings the question how can we handle the stress during the project work? We will take a look at it in the case study.

V. CONTROL DURING THE PROJECT EXECUTION

A. *What is control?*

Project controls are processes for gathering and analyzing project data to keep costs and schedules on track. The main objective of project controls is to minimize the variance in costs and schedule from what was originally planned. The project controls measure the project status.

Project controls contribute to successful project management, as it alerts project stakeholders to potential trouble areas and allows them to course correct, if needed.

Project controls activities must run through the complete project life cycle—from the initiation phase until closure—to monitor and control the various factors that impact cost and schedule. Depending on the project phase the controlling activities can be different for example:

- documentation,
- monitoring project costs,
- tracking project schedule,
- feedback and reporting.

B. *The importance of project control*

On the webpage of Ecosys (www.ecosys.net) the following information can be found. “In a 2018 survey, 88% of respondents said they perceive project controls to be important or critical to the success of enterprise projects. The

report also confirms the correlation between project controls and success: those that perceived controls as 'critical' were twice as likely to meet all project objectives. Those who perceived project controls as 'not important at all' were more than 3 times more likely to fail."
The study results emphasize the importance of project control. I share this opinion: there is no project success without strong and consistent tracking and control.
In the communication plan it must be established what will be the frequency of the project status meetings.

VI. The case study: z/OS Upgrade from 2.2 to 2.3

The case study introduces a project that was executed at a multinational IT service provider company.
The project scope was: migrate the z/OS systems from 2.2 up to 2.3. for four clients on twelve servers by an exact date (when the z/OS support of 2.2 will be terminated). The original length of the project was two years, I took over the leadership for the last five months.

A. Take over the project – the discovery phase

The first task was to understand what happened in the last time period and get familiar with the project organization. The z/OS upgrade was feasible only in case of prepared servers. The prerequisite of the z/OS upgrade was the following: all software product had to be compatible with the new z/OS level. The preparation itself required significant effort and time.

- The technical team members had to check on each server the available software products. They were responsible for providing the actual status. The software lists were collected in excel sheets: one sheet for each server, one document for each client. One excel document usually included 3 sheets – for the development, test and production environments. The length of one list was between 100 up to 300 lines.
- If the software product was not compatible with the planned z/OS upgrade date the software upgrade should be executed first. In some cases this task required cooperation between the SMEs, like downloading a PTF than execute the upgrade on the server.

It turned out that there was no project management at all. The technical team members worked individually. They put their data without coordination and cooperation into the excel sheets. Due to the lack of proper scheduling and communication most of these data were missing from the four excel sheets what I took over.

B. Change the management system with hard and power skills

That was a good approach, manage the project in smaller pieces but parallel for each customer. The biggest missing project management tasks were the time planning, communication and control.

I made a very simple action in terms of the communication: I organized planning and status meetings. First of all I emphasized the importance of the task and the delay of the project and I asked the team members for a very strong cooperation. I asked for feedback many times if their understood my messages. During the planning meeting I put the z/OS upgrade dates into the excel sheets, so it was based on personal commitments. Once the plan was ready, it was communicated to the client. Delivering the plan was our last chance to get back the trust of the customer.

The communication was changed within team, it became more frequent, clearer and focused to the understanding. Due to these changes the continuous postponing was ended, and the focus was on the software and z/OS upgrade execution.

If the project leader finds the way to the heart of the team with the right tone of communication, the success can be reached more likely.

In my opinion the continuous tracking and control of the project progress is a very important success factor in terms of the project delivery. I introduces weekly status meetings and encouraged the team member to give me information if they think the agreed delivery date is jeopardized. All the agreements and results: who does what and when were documented during the meetings.

Finally with these small changes the z/OS upgrade project was on the right path to deliver the expected results to the client.

Acknowledgment

Ágnes BÁrtfai gratefully acknowledges the financial support by the Doctoral School of Applied Informatics and Applied Mathematics.

References

[1] A Mathematical Theory of Communication - https://www.itsoc.org/about/shannon-1948

[2] Project Management Body of Knowledge - https://en.wikipedia.org/wiki/Project_Management_Body_of_Knowledge

[3] Luhmann, Habermas, and the Theory of Communication - https://www.leydesdorff.net/montreal.htm

[4] https://www.prince2.com/eur/blog/project-triangle-constraints

[5] Orosz, I., & Orosz, T. (2014). Microsoft Change Management Applying Comparison of Different Versions. Acta Technica Jaurinensis, 7(2), 183-192.

[6] Orosz, I., & Orosz, T. (2017, September). Software as a service in cloud based ERP change management. In 2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY) (pp. 000181-000186). IEEE.

[7] Orosz, I., Selmeci, A., & Orosz, T. (2019, January). Software as a Service operation model in cloud based ERP systems. In 2019 IEEE 17th World Symposium on Applied Machine Intelligence and Informatics (SAMI) (pp. 345-354). IEEE.

[8] Selmeci, A., & Orosz, T. (2015, May). Trends and followers in GUI development for business applications with implications at University Education. In 2015 IEEE 10th Jubilee International Symposium on Applied Computational Intelligence and Informatics (pp. 243-251). IEEE.

[9] Orosz, I., & Orosz, T. (2012, May). Business process reengineering project in local governments with ERP. In 2012 7th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI) (pp. 371-376). IEEE.

[10] Selmeci, A., Orosz, T., & Györök, G. (2016, January). Teaching ERP user interfaces: Adequate sequences of topics and technologies. In 2016 IEEE 14th International Symposium on Applied Machine Intelligence and Informatics (SAMI) (pp. 361-367). IEEE.

[11] Orosz, T. (2020, January). Introduction of Innovative SAP Development Solutions at University Level. In 2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI) (pp. 171-174). IEEE

Generating the optimal process structure for decision sequences by Separation Network Synthesis

László Szili
Department of Computer Science and Systems Technology
University of Pannonia
Veszprém, Hungary
szili@dcs.uni-pannon.hu

Botond Bertók
Department of Computer Science and Systems Technology
University of Pannonia
Veszprém, Hungary
bertok@dcs.uni-pannon.hu

Abstract — **In this paper it is examined whether efficiency improvements and cost reductions could be achieved by changing the order of certain predefined tasks in decision processes. A method has been developed based on log analysis and mathematical modelling. A Separation Network Synthesis (SNS) model has been constructed from the business process related historical data, and extended by further constraints for resource parameters and prerequisites. By transforming the model to a Process Network Synthesis (PNS) problem, optimal and alternative suboptimal structures, i.e., decision sequences can be generated and global optimality can be guaranteed.**

Keywords: Optimization, P-graph, Business Process, Process Synthesis, Decision Network

I. INTRODUCTION

Operation of companies highly relies on their precisely defined internal processes. Each process has a list of tasks with dependencies and required resources. The process flow is affected by the results of decisions implied in the processes, or in a more complex situation, a network of decisions. This approach has led to a stable and balanced operation, but unfortunately the processes are mostly described based on past experience without investigation on their optimality and cost effectiveness. It is also likely that these processes are not adjusted to the changes of resource cost or availability, which can lead to further inefficiencies during operation.

II. METHODOLOGY: SNS BY PNS

Separation-Network Synthesis aims at determining the optimal sequence of operations from multicomponent mixtures by separations, mixing, and divisions, leading to a series of products of predefined compositions at the end. Separators can select certain components and distinguish them from others. Mixers can merge mixtures of any compositions, while dividers can share mixtures between multiple operations without changing their compositions. To explore all the potential separation sequences, first a rigorous superstructure is constructed, which incorporates all the possible separation sequences as its parts [1]. Then an optimization model can select the optimal part of it.

Process Network Synthesis (PNS) is a complete optimization framework for designing process structures or flowsheets from a set of raw materials (resources) and candidate operating units (activities) to achieve the set of desired products (outcomes). By the help of the corresponding combinatorial algorithms it can provide not only a single optimal process structure, as most optimization software do, but a series of best optimal and suboptimal process structures [2]. As a result, process designers can compare and select the most appropriate process structure for their purposes.

In the current paper a methodology is proposed, where decision sequences in business processes are formulated as a Separation Network Synthesis problem, and then optimal and alternative best networks are computed by PNS algorithms and software. The methodology is illustrated by a real life case study of restructuring a procurement process.

III. CASE STUDY

In business processes decisions need to be made by answering a questions which can have multiple answers, and costs which cover their evaluations. A decision network is defined by a set of decisions and their dependencies. By evaluating historical data, the probabilities of occurrences can be determined for each decision sequence. The aim is to determine the cost optimal decision network, which results proper outcomes with minimal average resource utilization. For demonstration, we have analyzed a procurement decision process.

A. Questions and costs

The network contains 3 questions, each of them with 2 possible answers, Yes or No respectively. An evaluation cost is also given for every question. Data for the example can be found in Table 1.

Question	Cost	Result
Q1: Centralized procurement?	200	Yes
		No
Q2: Is it in the budget?	100	Yes
		No
Q3: Approved?	300	Yes
		No

Table 1. Given questions with parameters

B. *Structure of the problem*

Based on the current process, a network can be constructed from the questions, which describes the dependencies among the questions. The network in Figure 1. is the currently applied process for the procurement.

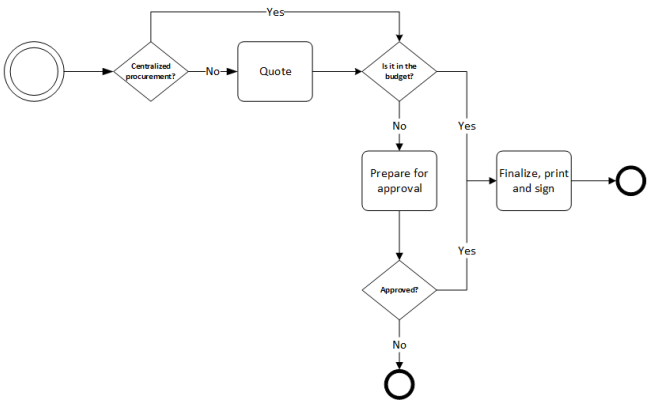


Figure 1. Structure of the original decision network

C. *Occurrences*

The 3 questions may result 8 different combinations of answers. From the analysis of historical data, the occurrence of each possible scenario can be estimated and it can be found in Table 2.

Case	Q1	Q2	Q3	#
P1	Yes	Yes	Yes	30
P2	No	Yes	Yes	20
P3	Yes	No	Yes	24
P4	No	No	Yes	16
P5	Yes	Yes	No	0
P6	No	Yes	No	0
P7	Yes	No	No	6
P8	No	No	No	4

Table 2. Potential cases with occurrences

D. *Outcomes*

It is possible that different result combinations lead to the same outcome. Each outcome in Table 3. is defined by a set of result combinations.

Outcome	Cases	
Fail	P7	P8
Procurement	P2	P4
Price list	P1	P3
Impossible	P5	P6

Table 3. Potential outcomes formed from cases

IV. Modell Transformation

In previous studies it was shown that SNS problems with multiple separation techniques [3] can be handled and solved efficiently as PNS problems [4]. Hereby I introduce our method to transform decision networks into SNS problem.

A. *Analogy between Decision Networks and SNS*

By examining the problem classes, similarities can be found. Each decision is equivalent with a 1 input multiple output sharp separator. Both has its own input which get processed into previously defined outcomes based on an examined property. The property in this case is the answer for a given question. By translating these terms to PNS, every decision can be represented by an operating unit, and every possible scenario set can be handled as a material. Every operating unit has a single input representing a case set, and these cases are classified into two outputs according to the answers replied for the question represented by the operation.

B. *Steps of the model transformation*

The proposed model transformation algorithm contains 3 main steps, which are the followings:

1. Define the raw material: In the first step, the raw material needs to be constructed. Each structure will contains only 1 raw material, which contains the mixture of every case from the original problem for which the probability is more than zero.
2. Define the products: Every Outcome represents a product in the structure. In the second step these product materials needs to be constructed.
3. Add operating units and intermediate materials

Operating units and intermediate materials are added in a recursive method, starting from the raw material. For each material an operating unit need to be added every time the material represents cases with different potential outcomes. The operating units input material will be the examined material, and the output materials need to be constructed from the input by dividing it into two new set of cases by the different answers. The newly created output materials are also need to be examined recursively until new intermediate materials are created.

C. *Assign costs*

Cost only appears for operating units. For each operating unit a proportional cost assigned, which is equal with the questions cost multiplied by the sum of the occurrences of the possible cases represented by the input material.

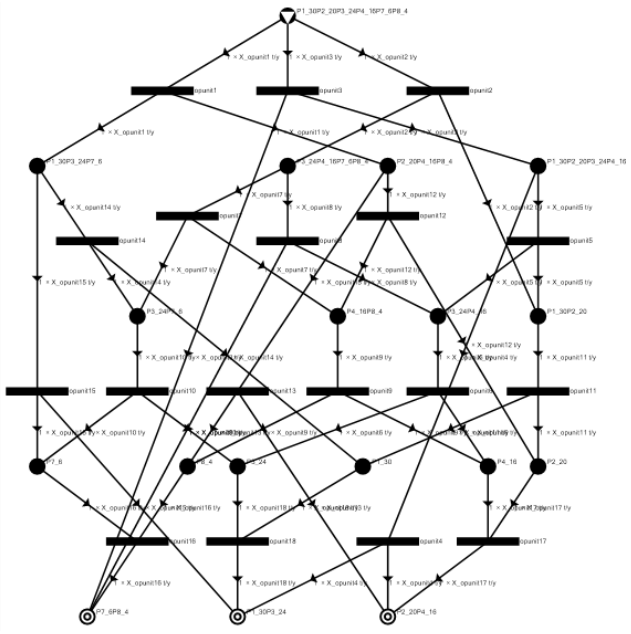


Figure 2. Generated P-graph model

D. Result of the model transformation for the example

The model transformation for the case study has led to a single raw material, 3 products, and 15 operating units with 14 intermediate materials. After connecting them, the superstructure represented by P-graph can be seen in Figure 2.

V. Results

The initial structure cost is 45,000. By Software P-graph Studio [3] the generated PNS problem can be solved efficiently. 8 structurally different solution structures had been computed by the algorithm ABB.

Number of Solution	Total Cost
1	43,000
2	45,000
3	45,000
4	47,000
5	48,000
6	48,000
7	50,000
8	57,000

Table 4. The generated solutions with total cost

A solution structure had been found with better cost than the initial structure resulting a potential of 4.4% cost saving. The cost optimal solutions' P-graph representation can be found in Figure 3.

After identifying the cost optimal structure, the P-graph representation can be transformed back to the business process of the decision network. The optimal decision network can be seen in Figure 4.

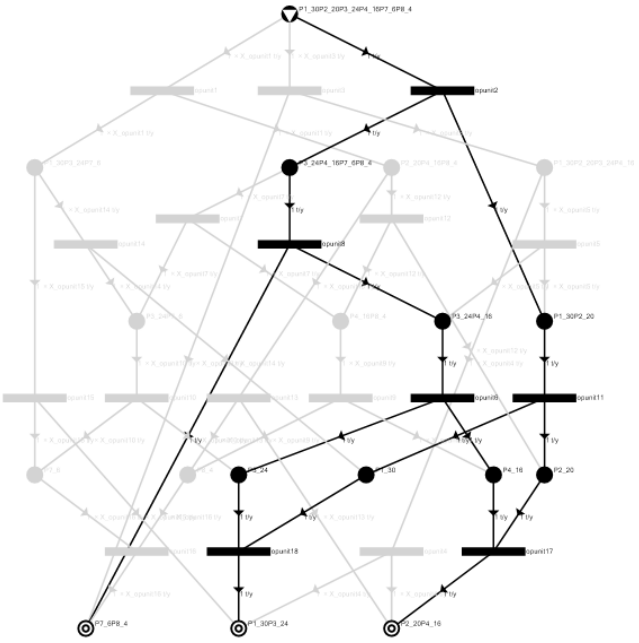


Figure 3. Cost optimal solution

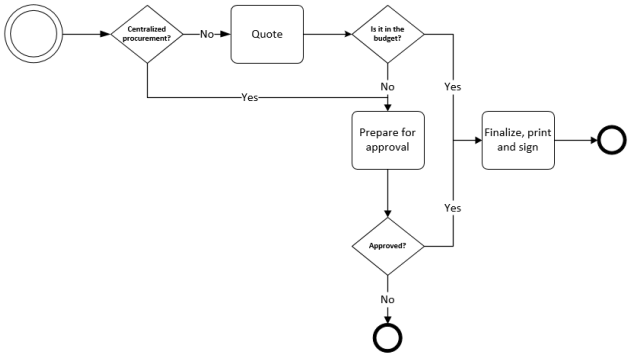


Figure 4. Optimal decision structure for cost optimal solution

VI. Conclusion

By the help of the proposed algorithmic method, the efficiency and optimality of the applied decision mechanism can be examined automatically at any time. The deviation from the optimum can be measured, and the best decision order or network in the given situation can be regenerated automatically.

References

[1] Kovacs, Z., Z. Ercsey, F. Friedler, and L. T. Fan, Separation-Network Synthesis: Global Optimum through Rigorous Super-Structure, Computers Chem. Engng, 24, 1881-1900 (2000).

[2] Friedler, F., K. Tarjan, Y. W. Huang, and L. T. Fan, Combinatorial Algorithms for Process Synthesis, Computers Chem. Engng, 16, S313-320 (1992).

[3] Heckl, I., F. Friedler, and L. T. Fan, Solution of separation network synthesis problems by the P-Graph methodology, Computers & Chemical Engineering, 34(5), 700-706 (2010).

[4] Heckl, I., F. Friedler, and L.T. Fan, Reduced Super-structure for a Separation Network Comprising Separators effected by Different Methods of Separation, Computers Chem. Engng., 33, 687-698 (2009).

[5] P-graph Studio, http://p-graph.org/pgraphstudio/, download at Oct. 1, 2020

Low-level UDP network programming performance comparison on Linux

Viktor Devecseri
Alba Regia Technical Faculty
Óbuda University
Budapest, Hungary
devi86@gmail.com

***Abstract*—This article presents some low-level UDP based network programming methods and compares them by measuring the number of transmitted messages per second on x64 and POWER9 servers. The examined methods are: a naive implementation with busy waiting variants; single- and multi-threaded synchronous and single-threaded asynchronous implementations using Boost.Asio; and an implementation using the `sendmmsg()` and `recvmmsg()` system calls for sending multiple messages at once.**

***Index Terms*—udp, user datagram protocol, socket, performance, comparison, linux, busy waiting, boost.asio, asio, sendmmsg, recvmmsg**

I. Introduction

The User Datagram Protocol (UDP) is a connectionless communication model with no guarantee of delivery, ordering of the messages, or that the messages arrive only once [1]. It is widely used, e.g. the Domain Name System (DNS) builds upon it, multi-player games and other real-time applications like streaming use it where low latency is important. Since UDP doesn't include mechanisms for handling packet loss as Transmission Control Protocol (TCP) does, this scenario must be handled by the application, if needed.

In this article different low-level network programming methods are presented for UDP sockets and their impact on the total transmitted messages per second. UDP sockets were used to avoid the additional logic and messaging built into the TCP protocol. The methods presented here were mainly inspired by [2] and [3]. The *echo protocol* was implemented in C++, where the server immediately sends back the received data to the client without any processing. After the client receives the reply from the server, it sends the next message and waits for the response. For comparison the total number of sent and received packets were measured.

Section III introduces the methods used for messaging. Section IV provides an in-depth comparison of the methods with different message sizes. There are two scenarios: (1) one client with one server, (2) four clients with one server. In the tests the client and server processes were running on different physical servers.

II. Testing environment

A. Hardware and Operating System

The measurements were done in a lab environment with multiple x64 and POWER9 based servers with CentOS 7 operating system. The same measurements were carried out on both architectures without mixing them to have a clear comparison between the architectures. The following type of servers were used:

- IBM Power System LC922 (9006-22P) – *ppc64le*: dual IBM POWER9 02CY228 (2.7 GHz, turbo: 3.8 GHz) CPU; 512 GiB DDR4 RAM; Mellanox ConnectX-5 100 Gbit/s NIC.
- Supermicro SSG-6048R-E1CR60L – *x86_64*: dual Intel Xeon E5-2683 v4 (2.1 GHz, turbo: 3.0 GHz) CPU; 512 GiB DDR4 RAM; Mellanox Connect-X 4 100 Gbit/s NIC.

Note that on the POWER9 servers all-in-all 160 threads could run in parallel (2 CPUs · 20 cores/CPU · 4 threads/core), whereas on the x64 servers 64 threads (2 CPUs · 16 cores/CPU · 2 threads/core).

B. Software

The test programs were compiled on both architectures with *gcc 7.3.1* with the *-O2* flag and *boost 1.74.0* [4] is used in some tests.

The client applications count the sent and received number of messages and report them every second for one minute duration.

III. Methods

This section introduces the different networking methods and compares them on both the x64 and POWER9 servers. In these measurements there was one client and one server, and the message size was 10 bytes.

A. Baseline

To have a comparison baseline for further measurements the textbook example was created using the BSD Socket API (`recvfrom()`, `sendto()`) [1]. The server synchronously receives from a single socket and sends back the received data to the client. The client works similarly but first it sends the data to the server and then waits for the reply. Even though this article is about UDP performance comparison, this test was carried out for TCP too.

On Fig.1 the results are presented. The POWER9 servers achieved a bit better result with 57 160.4 MPS (round trip time: 17 μs) than the x64 servers with 52 941.3 MPS (round

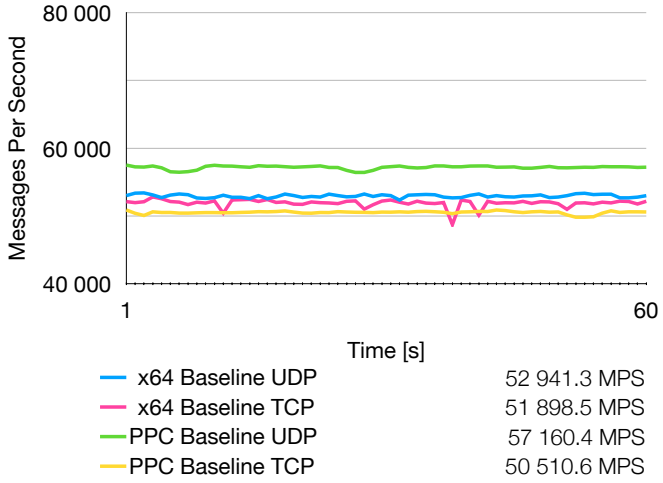


Fig. 1. Comparison of the TCP and UDP baseline measurements on the x64 and POWER9 servers over one minute with 10 bytes long messages.

trip time: 19 μs). For TCP the order is reversed, i.e. the x64 servers perform better than the POWER9 servers. Note, that in this scenario TCP is just slightly slower than UDP.

B. Busy waiting

By default `recvfrom()` blocks when there are no messages available for the application and the process is put to a sleeping state by the kernel and only waken up when there is a new message [1]. As [3] suggests the context switches could be avoided by busy waiting, where the socket is continuously polled by passing the `MSG_DONTWAIT` flag to the `recvfrom()` call. This modification was applied both to the server and to the client which increased the messaging rate by about 35-40% (Fig.1) at the cost of increased CPU usage.

The kernel's scheduler assigns the threads to a logical CPU to run it, but the threads could be rescheduled during runtime. To avoid this change, both the server and client were pinned to the 0^{th} logical CPU. In case of the POWER9 servers this didn't change the results significantly, but the variance decreased. However, the performance dropped on the x64 servers by about 14% (Fig.2).

C. Boost.Asio

Boost.Asio is a cross-platform C++ library for network and low-level I/O programming with both synchronous and asynchronous programming models. Boost.Asio also provides the interfaces and functionality specified by the "C++ Extensions for Networking" Technical Specification [4].

1) Synchronous API: The Boost.Asio library includes a low-level socket interface based on the BSD Socket API [4]. Since it is a thin layer over the API used in the baseline measurement (Section III-A), similar results were measured (Fig.3) with about 3.5% decrease (which could be due to different network utilization by other users).

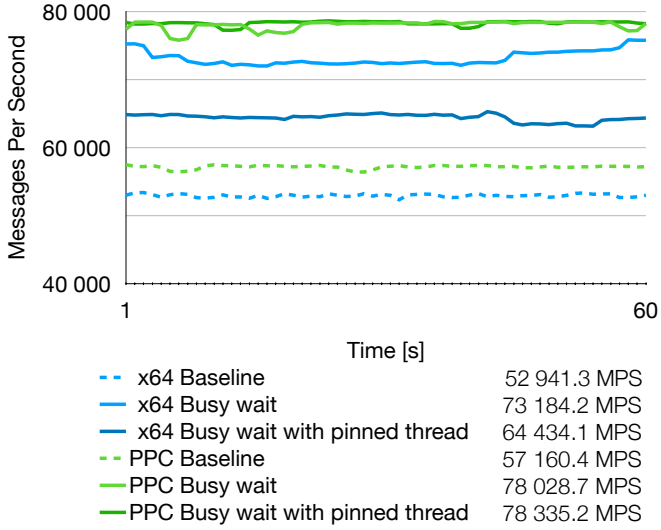


Fig. 2. Comparison of the busy waiting measurements on the x64 and POWER9 servers over one minute with 10 bytes long messages.

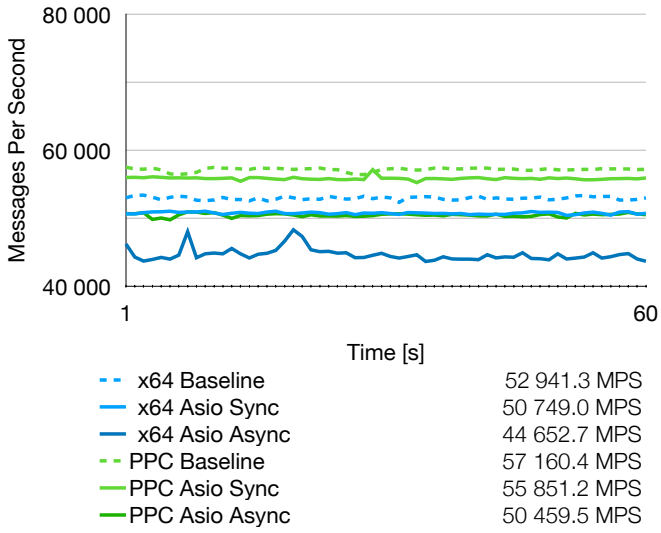


Fig. 3. Comparison of the Boost.Asio synchronous and asynchronous APIs based measurements on the x64 and POWER9 servers over one minute with 10 bytes long messages.

2) Asynchronous API: The Boost.Asio library also provides asynchronous support based on the Proactor design pattern [4]. The same *echo protocol* was implemented with this API. The server starts an asynchronous receive operation, and when it completes successfully an asynchronous send operation is started. When the send operation is finished a new asynchronous receive operation is started. It performs about 15% worse compared to the synchronous APIs (Fig.3).

Since the asynchronous model requires more housekeeping (both in the application and in the kernel) it is not worth it for a single socket. However, it provides a clean and event driven programming model which allows to handle many concurrent sockets on a single thread. It could easily be scaled to multiple threads, which could be beneficial especially when

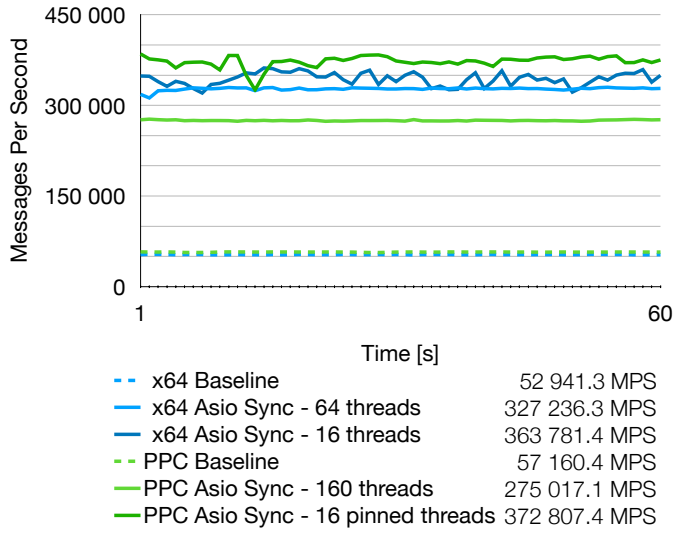


Fig. 4. Comparison of the multi-threaded measurements on the x64 and POWER9 servers over one minute with 10 bytes long messages.

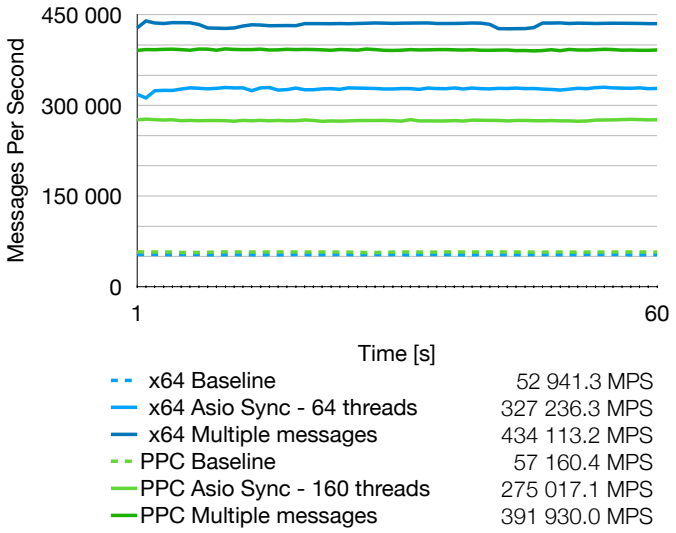


Fig. 5. Comparison of the multi-threaded and the multiple message APIs based measurements on the x64 and POWER9 servers over one minute with 10 bytes long messages.

the completion handlers of the asynchronous operations do more processing.

D. Multi-threading

A single socket was used from multiple threads via the Boost.Asio synchronous API (as in section III-C1). The implementation was done in a simple way, so it is possible that in the client a different thread will receive the response of another thread's message. However, all threads send the same message and this way the underlying network stack's raw performance was measured.

First as many threads are started as the number of logical CPUs, i.e. 160 on the POWER9, and 64 on the x64 servers. The x64 servers outperformed the POWER9 servers (Fig.4) by about ∼50 000 MPS which is surprising because in the baseline test the POWER9 servers performed better.

Since many threads access the same socket, the kernel spends more time waiting for the internal locks belonging to the socket. When the number of threads were limited to 16 the performance increased (328 419.1 MPS for POWER9 and 363 781.4 MPS for x64). When these threads are pinned to separate logical CPUs, the performance increased for the POWER9 servers to 372 807.4 MPS, but decreased for the x64 servers to 343 920.6 MPS (as seen in section III-B). On Fig.4 the better results are shown. Note that the variance increased significantly with the 16 threads compared to the maximum number of concurrent threads supported by the CPU.

E. Multiple messages

Since Linux 3.0 with the `sendmmsg()` and `recvmmsg()` system calls it is possible to send multiple messages on a socket to different endpoints with a single system call [5] [6].

Sending and receiving 64 messages with a single system call outperforms the multi-threaded approach (section III-D) with less CPU utilization. The x64 servers go beyond 400 000 MPS and the POWER9 servers approaches it (Fig.5).

IV. Comparison

Fig.6 provides an overall comparison of the previously described methods with different payload sizes (10, 100, 1 000, 2 000 bytes) with one client and one server measured on the POWER9 servers.

Since the Ethernet MTU (Maximum Transmission Unit) was set to 1 500 bytes on the servers used in the measurements, the larger messages (i.e. 2 000 bytes) were fragmented into multiple IP packets. The effect of this is clearly noticeable in the measurements too (Fig.6).

Fig.7 shows the results when four clients message one server at once measured on the POWER9 servers. For the *baseline*, *busy waiting*, *Asio Sync* and *Asio Async* methods the server's transmission rate scales almost linearly with some drop per client compared to the single client case.

However, for the *multi-threaded* and *multiple messages* tests the messaging rate of the clients are not distributed evenly. Furthermore, in some cases of the *multiple messages* test one or two clients' messages were dropped and didn't reach the test applications.

Based on the tests on the POWER9 servers the upper limit of a socket's receive rate seems to be around 400 000 MPS by default.

V. Conclusion

In this article some networking methods were presented and their effect on the performance measured by the total transmitted Messages Per Second (MPS) on a single UDP socket. It was observed that the method used for communication has a large impact on the overall performance. Starting with ∼50 000 MPS as the *baseline*, up to ∼400 000 MPS could be achieved with the *multiple messages* method. However, there is a place for future improvements too.

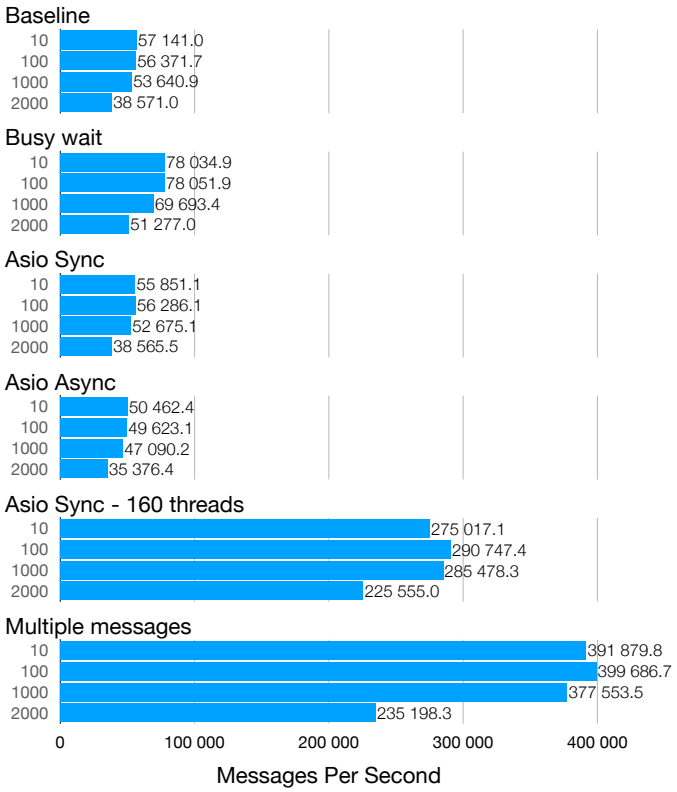


Fig. 6. Comparison of different message sizes with different methods on the POWER9 servers with one client and one server.

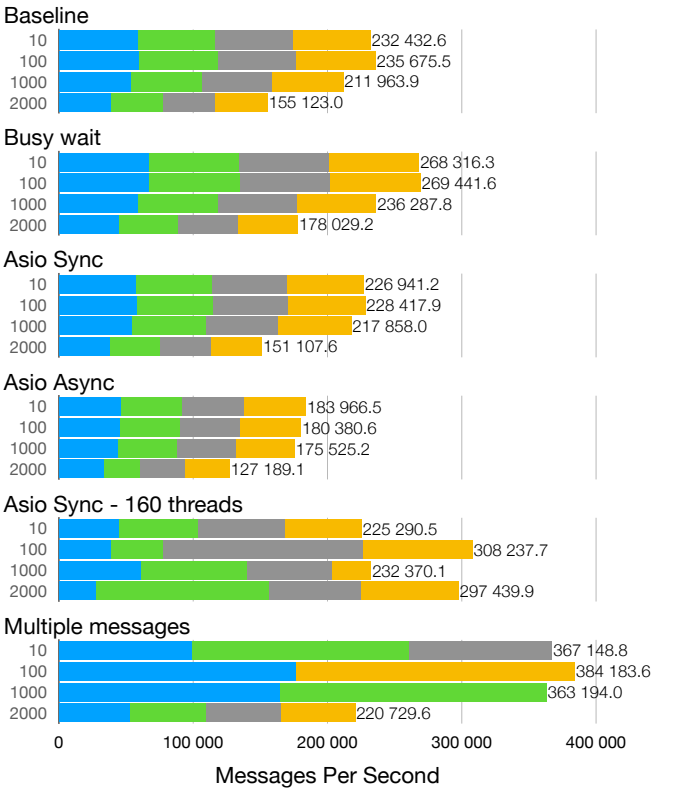


Fig. 7. Comparison of different message sizes with different methods on the POWER9 servers with four clients and one server. Each color represents the messaging rate of a POWER9 client. Note, that for the last two methods they are not evenly distributed.

A. *Future work*

The tests except the *multi-threaded* and *multiple messages* examined synchronous communication, i.e. the client waited for a reply before sending a new message. It should worth investigating having multiple pending messages, i.e. the client sends N messages at once, and only sends a new one when it receives a reply. This way there would always be N pending messages on the network. Comparing a manual implementation with the *multi-threaded* and *multiple messages* case would be interesting.

In the measurements carried out there was only one socket in the client, and one socket in the server. By having multiple sockets, the overall performance could be improved significantly [2]. By using the `SO_REUSEPORT` socket option multiple sockets could be bound to the same address and port, i.e. the clients don't need to address different ports. By having a clear understanding how the networking stack handles the packets [2] [3] [7] and fine tuning some kernel and NIC specific parameters, on a 10 Gbit/s network 7 010 000 Packets per Second (9 310 Gbit/s) could be achieved [8].

Currently the measures were only carried out with C++ implementations, however it would be interesting to see the impact of other languages (e.g. Java, C#, Python) as well.

Up to this point the networking was done via the kernel, however by having a user space network stack it could be omitted. This could give higher performance for some specific applications (e.g. in high performance trading) [9] however, the network interface would be exclusively used by only one application [10].

References

[1] W. R. Stevens, B. Fenner, and A. M. Rudoff, *UNIX Network Programming, Vol. 1*, 3rd ed. Pearson Education, 2003.

[2] M. Majkowski. (2015, Jun. 16) How to receive a million packets per second. Cloudflare. [Online]. Available: https://blog.cloudflare.com/how-to-receive-a-million-packets/

[3] M. Majkowski. (2015, Jun. 30) How to achieve low latency with 10Gbps Ethernet. Cloudflare. [Online]. Available: https://blog.cloudflare.com/how-to-achieve-low-latency/

[4] (2020, Oct.) Boost C++ Libraries. [Online]. Available: https://www.boost.org

[5] *sendmmsg(2) - Linux manual page*. [Online]. Available: https://man7.org/linux/man-pages/man2/sendmmsg.2.html

[6] *recvmmsg(2) - Linux manual page*. [Online]. Available: https://man7.org/linux/man-pages/man2/recvmmsg.2.html

[7] T. Herbert and W. de Bruijn. Scaling in the Linux Networking Stack. [Online]. Available: https://www.kernel.org/doc/Documentation/networking/scaling.txt

[8] T. Makita, "Boost UDP Transaction Performance," in *LinuxCon + ContainerCon Japan 2016*, Jul. 13–15, 2016. [Online]. Available: https://events.static.linuxfound.org/sites/events/files/slides/LinuxConJapan2016_makita_160712.pdf

[9] J. Evans. (2016, Jun. 30) Why do we use the Linux kernel's TCP stack? [Online]. Available: https://jvns.ca/blog/2016/06/30/why-do-we-use-the-linux-kernels-tcp-stack/

[10] M. Majkowski. (2016, Jul. 7) Why we use the Linux kernel's TCP stack. Cloudflare. [Online]. Available: https://blog.cloudflare.com/why-we-use-the-linux-kernels-tcp-stack/

AUTONOMOUS POWER SUPPLY SYSTEMS OPTIMIZATION FOR ENERGY EFFICIENCY INCREASING

Olga Shvets
Faculty of Information Technology
D.Serikbayev East Kazakhstan State Technical University
Ust-Kamenogorsk, Kazakhstan
olga.shvets75@gmail.com

Asel Naizabayeva
Faculty of Information Technology
D.Serikbayev East Kazakhstan State Technical University
Ust-Kamenogorsk, Kazakhstan
aselnaizabaeva@mail.ru

Alibi Toleugazin
Faculty of Information Technology
D.Serikbayev East Kazakhstan State Technical University
Ust-Kamenogorsk, Kazakhstan
toleugazin.a@mail.ru

Seebauer Márta
Alba Regia Technical Faculty
Obuda University
Szekesfehervar, Hungary
seebauer@uni-obuda.hu

***Abstract* — The paper focuses on the calculation of electricity power generated by solar panel for an autonomous house. It was drawn a block diagram based on the research. The advantage of this automatic system is the elimination of moving elements. It allows the decreasing of the risk of various mechanisms failure and increasing the stability of the system.**

***Keywords*—solar panel, energy consumption, energy saving microcontroller, control system**

I. INTRODUCTION

In modern conditions the use of electronic and electrical devices has increased. This increasing of electronics and electrical appliances use and consumption has negatively impacted on unprecedented energy consumption. Due to the gap between the supply and demand, the price paid by the end-user subsequently increases annually.

As a result, there is a need to optimize and develop more energy efficient technologies and electrical systems. Because of this fact, the development of new fundamental and applied research in the field of energy efficiency and saving began. Among these research areas leading to significant changes in energy consumption are the development of integrated, improved monitoring and control mechanisms with the ability to monitor and control energy consumption effectively so that users can easily measure the energy consumption of electronic devices and optimize their use to improve energy efficiency. [1]

In this research, we have chosen a solar panel as the energy source for our automatic system. Solar energy has undergone significant changes in recent years and has also taken the lead in development in comparision to other alternative energy sources. Solar energy simply uses sunlight as energy. This does not require large generators that take up a lot of space. Another advantage of this automatic system is the absence of moving elements, which eliminates the risk of failure of various mechanisms in this system.[2]

The total energy outflowing from the sun is more than we consume at all, which means that this energy source is theoretical one of the best sources for the future. The challenge is to use and save this energy in a cost effective way in practice [3].

II. DEVELOPMENT OF A STRUCTURAL DIAGRAM OF THE FACILIT IES POWER SUPPLY

In order to calculate the required power for our solar panel, it is necessary to know electricity consumption per month. It is possible to determine the required consumed electricity volume in kilowatts per hours by looking at an electric meter (Table 1).

Table 1.Electricity consumption by devices (approximate).

№	Electrical devices	Power con-sumption, kW	kWh	kW / m
1	Refrigerator	0,04	24	30
2	TV set	0,166	3	15
3	Electric stove	1,3	1,5	58,5
4	Oven	1,8	1 (per week 3 days)	21,6
5	Washer	1	2 (per week 3 days)	24
6	Computer	0,22	2	13,2
7	Teapot	1	0,5	15
8	Dishwasher	0,47	1	14,1
9	Vacuum cleaner	0,5	1 (per week 3 days)	6
10	Iron	0,5	0,3 3(per week 3 days)	2
11	Microwave	1	0,05	4,5
12	Multicooker	0,166	1	5
13	Bulb (10pcs) 100w	0.1	3	9
14	Heater	2	1	60
15	Energy saving lamp (10pcs) 20 W	0,02	5	3

After finding out how many kW of electricity can be consumed in one day, it is possible to find out how much energy the solar panel needs to produce. We need the solstice and solar panel power for calculation. For example:

the time of the sun from dawn to dusk (Trz) is multiplied by the power of the panel (P): Trz * P = 542 * 2 = 1084W (Table 2) (Figure 1).

Table 2. The generated annual power of one solar panel.

January	February	March	April	May	June	July	August	September	October	November	December
1084	1190	1348	1532	1698	1822	1840	1742	1580	1406	1232	1104
1086	1194	1354	1536	1704	1824	1838	1736	1574	1400	1226	1102
1088	1200	1360	1542	1708	1826	1836	1732	1568	1394	1220	1100
1090	1204	1366	1548	1712	1828	1834	1728	1562	1388	1216	1098
1092	1210	1370	1554	1718	1830	1832	1722	1556	1382	1210	1094
1094	1214	1376	1560	1722	1832	1830	1718	1550	1376	1206	1092
1096	1220	1382	1566	1728	1834	1828	1716	1546	1370	1200	1090
1098	1224	1388	1572	1732	1836	1826	1708	1540	1364	1196	1088
1100	1230	1394	1578	1736	1838	1824	1704	1534	1358	1190	1086
1104	1234	1400	1584	1742	1840	1822	1698	1528	1352	1186	1085
1106	1240	1406	1588	1746	1841	1820	1694	1522	1346	1182	1084
1110	1246	1412	1594	1750	1843	1818	1688	1516	1342	1176	1082
1112	1252	1418	1600	1754	1844	1814	1682	1510	1336	1172	1081
1116	1256	1424	1606	1760	1846	1812	1680	1504	1330	1168	1080
1118	1262	1430	1612	1764	1847	1808	1674	1498	1324	1164	1079
1122	1268	1436	1618	1768	1849	1806	1670	1492	1318	1158	1078
1126	1274	1442	1622	1772	1850	1802	1664	1488	1312	1154	1078
1130	1278	1448	1628	1776	1851	1798	1658	1482	1306	1150	1077
1134	1284	1454	1634	1780	1851	1796	1654	1476	1304	1146	1077
1136	1290	1460	1640	1784	1851	1792	1648	1470	1298	1142	1076
1140	1296	1466	1644	1786	1851	1788	1642	1464	1292	1138	1076
1144	1302	1472	1650	1790	1851	1784	1638	1458	1286	1134	1076
1148	1306	1478	1656	1794	1851	1780	1632	1452	1284	1330	1077
1152	1312	1484	1660	1798	1850	1776	1626	1446	1278	1128	1077
1158	1318	1490	1666	1800	1848	1772	1620	1440	1272	1124	1078
1162	1324	1496	1672	1804	1847	1768	1616	1434	1266	1120	1078
1166	1330	1502	1676	1808	1846	1764	1610	1428	1262	1116	1079
1170	1336	1508	1682	1810	1844	1760	1604	1422	1256	1114	1080
1176	1342	1514	1686	1814	1843	1756	1598	1416	1250	1110	1082
1180		1520	1692	1816	1841	1752	1592	1410	1246	1108	1082
1184		1526		1818		1748	1588		1240		1084

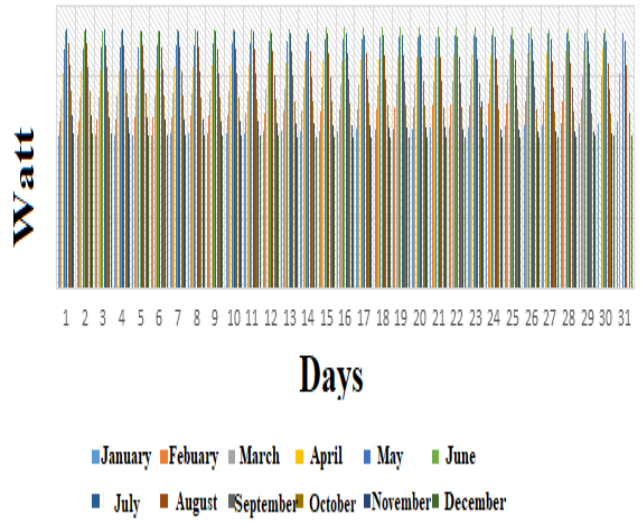


Figure 1. Graph of the generated annual capacity of one solar panel.

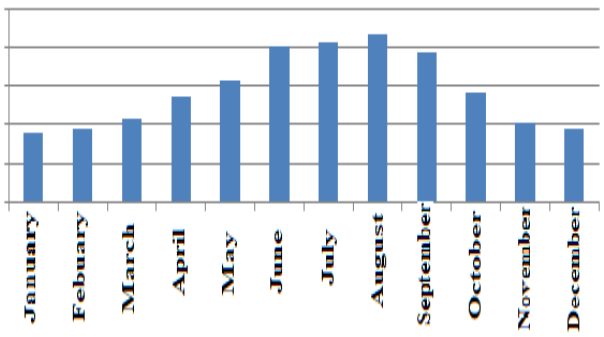


Figure 2 - Accurate energy production in one month (no cloudy days).

Then we calculate the electricity generated by the solar panel per one month. To do this, we need the average value of the generated energy in one day (Рд) and the number of cloudy days in one month (Кпас). We multiply the received data: Рд * Кпас = 1126.5 * 15 = 168975, and we get the lost power. To obtain the exact generated power in one month, it is necessary to subtract the lost power (Pp) from the received power in one month (Pm):

Pm-Pp = 34922 -16897.5 = 18024.5. (Figure 2)

Having calculated in detail the need for electricity and to fully supply the house, we need to purchase 15 solar panels with a capacity of 120W.

According to the study, a block diagram of the facility power supply was drawn up, which is shown on Figure 3. The solar battery is used for a low-voltage LED lighting system and after voltage conversion for the power supply system. [5-6]

The power of solar panels is 5 to 6 V in direct sunlight. The solar panels charge the lithium battery through the TP4056 charger module. This module is responsible for charging the battery and preventing overcharging. The lithium battery provides 4.2V when fully charged.

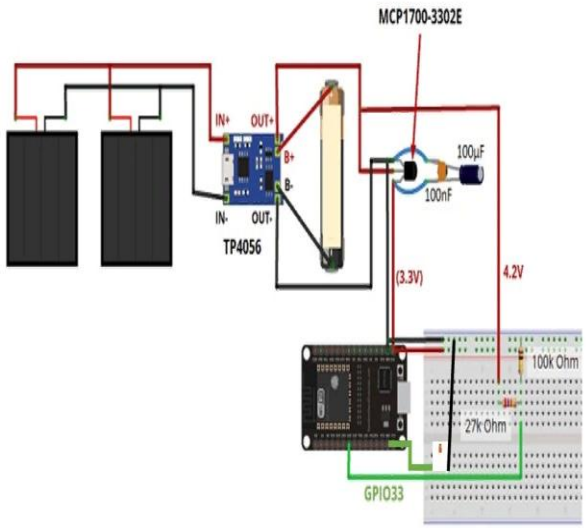


Figure-3. General diagram of the power supply monitoring and control system

Figure-4. Solar panels

We need to use a low dropout regulator circuit (MCP1700-3302E) to get 3.3V from the battery output. The voltage regulator output will power the ESP32 through the 3.3V pin.

The solar panels we use have an output voltage of 5 to 6 V. If we want our battery to charge faster, we can use multiple solar panels in parallel. In this example, we are using two mini solar panels as shown in the Figure 5.

To connect solar panels in parallel, solder the (+) terminal of one solar panel to the (+) terminal of the other solar panel. Let's do the same for the terminals (-). It might look like this.

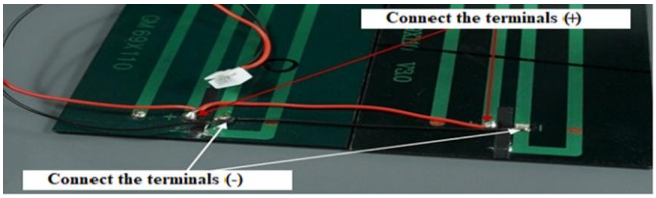


Figure-5. Parallel connection of solar panels

When solar panels are connected in parallel, we will get the same output voltage and double the current (for identical solar panels). In the following figure, solar panels are delivering approximately 6 V.

Figure-6. Measurement of solar panels with a multimeter

The TP4056 Lithium Battery Charger Module comes with circuit protection and prevents battery over-voltage and reverse polarity.

Figure-7. Charging module TP4056

TP4056 module turns on red LED when it is charging the battery and turns on blue LED when the battery is fully charged. Connect the solar panels to the TP4056 lithium battery charger module as shown in the circuit diagram below. Connect the positive terminals to the terminal marked IN + and the negative terminals to the terminal marked IN-.

Figure-8. Connecting solar panels to the TP4056 charging module

Then connect the positive terminal of the battery holder to B + and the negative terminal of the battery holder to B-.

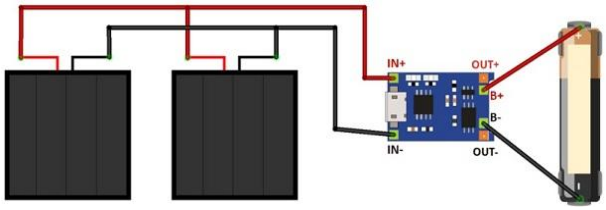


Figure-9. Connecting the battery terminals to the contacts of the TP4056 charging module

OUT + and OUT- are battery outputs. These rechargeable lithium batteries deliver up to 4.2V when fully charged (although the label says 3.7V).

Figure-10. Measurement of lithium batteries with a multimeter

We need a voltage regulator circuit to get 3.3V from the battery output to power the ESP32 through its 3.3V pin. Using a regular linear voltage regulator to reset the voltage from 4.2V to 3.3V is not a good idea, because when the battery is discharged to, for example, 3.7V, our voltage regulator will stop working because it has a high cutoff voltage. It is better use a low dropout regulator to effectively lower the voltage, or LDO for short, that can regulate the output voltage.

The MCP1700-3302E is one of the best solution. There is also a good alternative in the form of HT7333-A.

Figure-11. MCP1700-3302E stabilizer

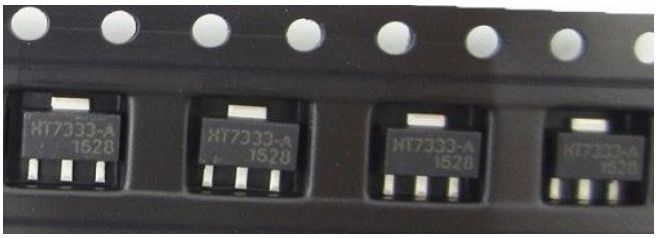
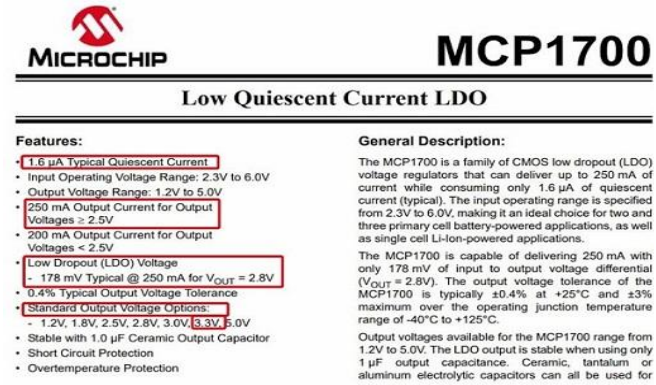


Figure-12. Alternative views of the MCP1700-3302E stabilizer

Any LDO that has similar characteristics to these two is also a good alternative. Our LDO should have similar specifications when it comes to output voltage, quiescent current, output current and low cut-off voltage. This information is at the documentation below.



MICROCHIP

MCP1700

Low Quiescent Current LDO

Features:

- 1.6 µA Typical Quiescent Current
- Input Operating Voltage Range: 2.3V to 6.0V
- Output Voltage Range: 1.2V to 5.0V
- 250 mA Output Current for Output Voltages ≥ 2.5V
- 200 mA Output Current for Output Voltages < 2.5V
- Low Dropout (LDO) Voltage
  - 178 mV Typical @ 250 mA for V_{OUT} = 2.8V
- 0.4% Typical Output Voltage Tolerance
- Standard Output Voltage Options:
  - 1.2V, 1.8V, 2.5V, 2.8V, 3.0V, 3.3V, 5.0V
- Stable with 1.0 µF Ceramic Output Capacitor
- Short Circuit Protection
- Overtemperature Protection

General Description:

The MCP1700 is a family of CMOS low dropout (LDO) voltage regulators that can deliver up to 250 mA of current while consuming only 1.6 µA of quiescent current (typical). The input operating range is specified from 2.3V to 6.0V, making it an ideal choice for two and three primary cell battery-powered applications, as well as single cell Li-Ion-powered applications.

The MCP1700 is capable of delivering 250 mA with only 178 mV of input to output voltage differential (V_{OUT} = 2.8V). The output voltage tolerance of the MCP1700 is typically ±0.4% at +25°C and ±3% maximum over the operating junction temperature range of -40°C to +125°C.

Output voltages available for the MCP1700 range from 1.2V to 5.0V. The LDO output is stable when using only 1 µF output capacitance. Ceramic, tantalum or aluminum electrolytic capacitors can all be used for

Figure-13. MCP1700-3302E documentation

This is the pinout of the MCP1700-3302E: GND, VIN and VOUT pins.

Figure-14. MCP1700-3302E pinout

The LDO must have a ceramic capacitor and an electrolytic capacitor connected in parallel to the GND and Vout to smooth out the voltage peaks. Here we use a 100 UF electrolytic capacitor and a 100 nF ceramic capacitor. Follow the following diagram to add the voltage regulator circuit to the previous diagram. The wire with the white-gray stripe must be connected to the GND.

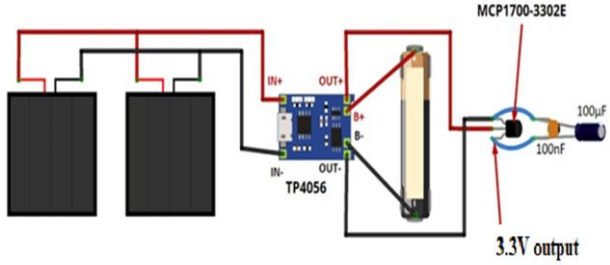


Figure-15. Connecting stabilizers and capacitors to the circuit

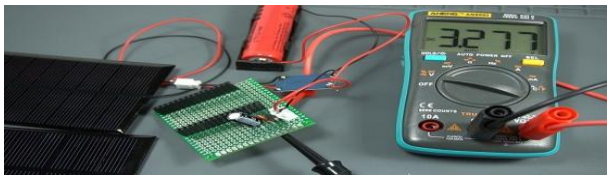


Figure-16. Measuring the voltage regulator with a multimeter

The Vout pin of the voltage regulator should supply 3.3 V. This is the pin that will power the ESP32 or ESP8266.

Finally, after making sure we get the correct voltage at the Vout pin of the voltage regulator, we can connect the ESP32. Connect Vout pin to 3.3V pin of ESP32 and GND to GND.

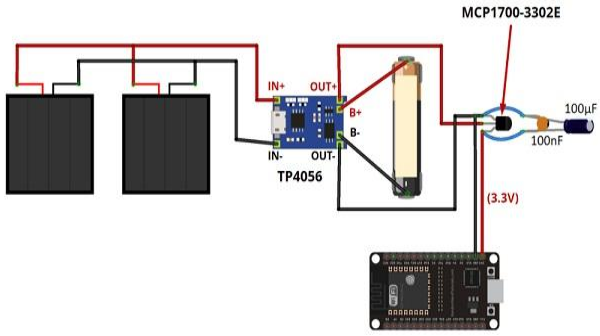


Figure-17. Connecting the voltage regulator contacts to the ESP8266 contacts

If the ESP32 is running on battery power or solar, as is the case, it can be very useful to monitor the battery level. One way to do this is to read the battery output voltage using the ESP32 analog pin. That said, the battery we are using here gives a maximum of 4.2V when fully charged, but the ESP32 GPIO lines run at 3.3V. So we need to add a voltage divider so we can read the voltage from the battery.

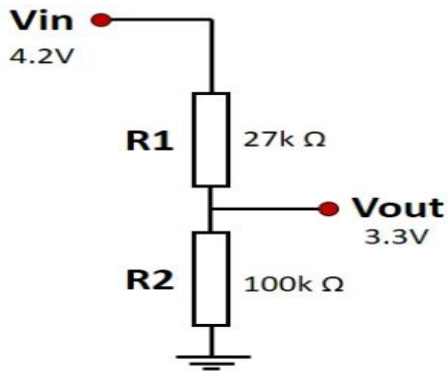


Figure-18. Adding obstacles to the scheme

The voltage divider formula looks like this:

Vout = (Vin * R2) / (R1 + R2).

So if we use R1 = 27 kΩ and R2 = 100 kΩ, we get 3.3 V. So when the battery is fully charged, the Vout output will be 3.3 V, which we can read using the ESP32 GPIO lines. Let's add two resistors to our circuit as shown in the following schematic diagram.

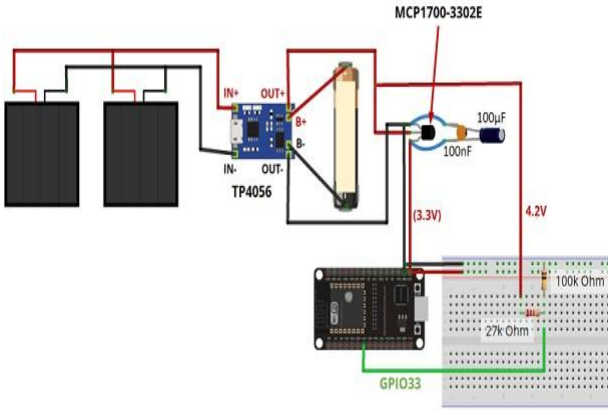


Figure-19. Connecting two resistors to the circuit

In this case, we are monitoring the battery level via GPIO33, but we can use any other suitable GPIO. Finally, to get the battery level, we can simply read the voltage on GPIO33 using the analogRead () function in our code (if we are using the Arduino IDE).

CONCLUSION

It is necessary to use alternative energy in Kazakhstan including solar energy. Based on the analysis, it can be concluded that in order to efficiently generate energy for our automated system in autonomous house.

The use of automated technologies allows an additional energy saving of 30%.

Installing an automated solar power system can have a significant impact on the environment. We can also improve our health and protect the environment by reducing our dependence on fossil fuels and installing an automated solar power system in our home or work.

REFERENCES

[1] "Effective monitoring and control of solar energy consumption by LED lighting devices with a built-in microcontroller" Naizabayeva AA

[2] "The use of alternative energy sources for energy saving" Naizabaeva AA, Tuleuғazin Ә.T, Zhenisov ES

[3] Different Types of Alternative Energy Sources // https://learning-center.homesciencetools.com/article/alternative-energy-sources/

[4] “Designing an automated control system for LED lighting with monitoring the health of LEDs” AA Naizabaeva, AE Baklanov.

[5] Derachits A.I "Control of the solar tracking system for mobile solar power plants" // Information control systems and computer monitoring (IMS KM 2013) -2013.-P.528-532

[6] Savrasov FV “Options for the construction of autonomous power supply systems using photoelectric devices and algorithms for their work” // Internet journal “Science Science”.-2013.-№6.-P.1-13.

[7] "Smart technologies for monitoring and control of power supply systems of an autonomous object" Zhaparova AT

AIS 2020 • 15th International Symposium on Applied Informatics and Related Areas • November 12, 2020 • Székesfehérvár, Hungary

Educational application of LEGO Mindstorms EV3 system by using different simulation environments

Péter Udvardy*, Bertalan Beszédes*
* Óbuda University/Alba Regia Technical Faculty, Székesfehérvár, Hungary
udvardy.peter@amk.uni-obuda.hu
beszedes.bertalan@amk.uni-obuda.hu

***Abstract*— Industry 4.0 requires wider and more sophisticated knowledge about automatization and robotics. Collaborative robots will be part of our every day soon, thus the education system must be adapted to these new requirements. The LEGO Mindstorms EV3 robotic set seems to suitable for educational purposes from the elementary school up to the universities considering the knowledge transfer level of these educational units. This robotic set can be programmed in its 'original' programming environment but also by using advanced simulation environment, too. This paper shows the robotic set, the sensors, the programming environments, the methods, and their possible application on different educational levels.**

Keywords— LEGO Mindstorms EV3, Matlab, Simulink, robotics

I. INTRODUCTION

Robots are parts of our daily life at work and in our free time. Industry 4.0 and smart solutions require more knowledge on robotics thus the educational system must be transformed into a 'smart' education from the very beginning at the elementary school.

The LEGO Mindstorms EV3 robotic sets help to involve young students into STEM world by building and programming these robots. The LEGO Mindstorms EV3 set contains the 'brick', motors and sensors as hardware elements and the programming environment is also available. Experimental results show that early stage education in technology, engineering and mathematics means higher commitment toward technical professions [1].

Matlab and Simulink are versatile simulation environments which are suitable for higher level programming tasks. Comparison between the 'original' and the higher-level programming environments is also presented. There are six types of sensors available in the set of which two sensors are highlighted as a demonstration for sensor application.

One of the main objectives of the present research is to create a new education material from the elementary school level up to the higher education level, completely. Many research results show that the LEGO Mindstorms EV3 could arouse interest for programming [2].

II. MATERIALS AND METHODS

The robot

The core of the LEGO Mindstorms EV3 robotic set is the 'brick' with 16MB of memory, Wi-Fi and Bluetooth modules and connection possibilities to motors and sensors. The brick has a monochrome display where basic information is available. There are two motor types and six types of sensors:

- infrared sensor
- touch sensor
- color/light sensor
- gyrosensor
- ultrasonic sensor
- microphone

The last two sensors are available from the former LEGO NXT sets and are compatible with the current one. Figure 1 shows the base of our robot platform.

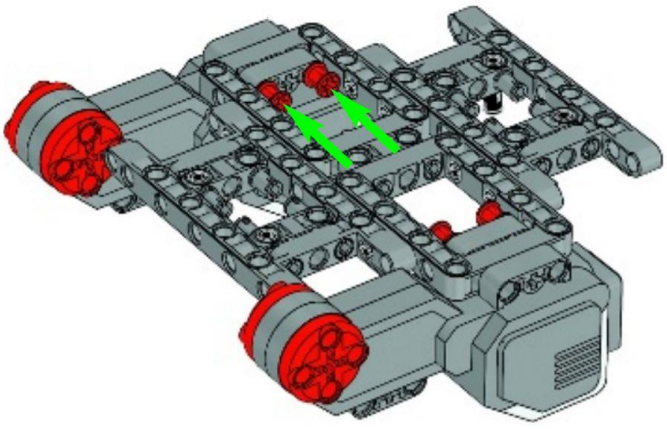


Fig. 1. The basic LEGO robot[3]

AIS 2020 • 15th International Symposium on Applied Informatics and Related Areas • November 12, 2020 • Székesfehérvár, Hungary

In our work the frame of the TRACK3R robot was used as the base robot and different sensors were put to this mobile platform [4].

The simulation environment

Two programming environments were used, one is the 'original' LEGO programmer software and the other was the MathWorks Matlab Simulink r2019b academic use version [5]. Many research results show that the use of Matlab in higher education supports better understanding of STEM related subjects [6].In order to the proper use of the programming environment the Matlab support package for LEGO Mindstorms EV3 hardware and the Simulink support package for LEGO Mindstorms EV3 hardware were installed [7]. The WiFi connection possibility was tested previously during the preparation phase but was ignored due to the inadequate communication possibilities between the robot and the computer thus the USB connection was applied [8].
In Simulink environment a model was built for controlling the LEGO robot. Two motors and three sensors were used during the simulation. Figure 2 shows the drive of the LEGO EV3 in Simulink environment.

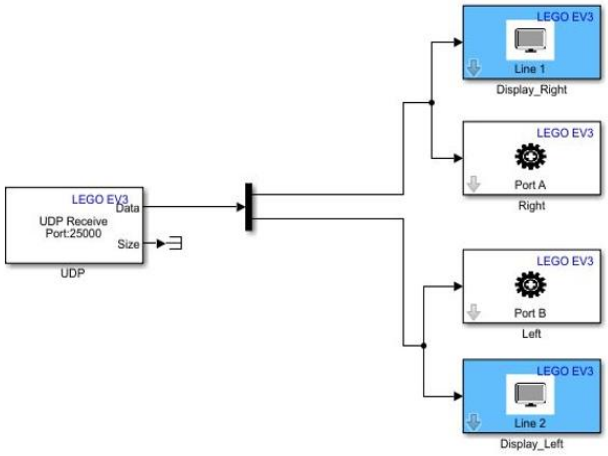


Fig. 2. Simulink drive of the robot

Educational use of the LEGO Mindstorms EV3 set

During the planning period of the educational material the age classes were determined first. For the beginners mainly the 'original' LEGO programming environment is recommended, while the Matlab and Simulink environments are suitable for advanced level users.
The first task is to move the robot by using the graphic programming environment and then the application possibilities of all sensors are shown. For the advanced level students these above-mentioned tasks with the sensors must be fulfilled by using Matlab and Simulink programming environment. Comparison between the two approaches are demonstrated during the practice.

Two sensors, namely the colour and light sensor and the gyro sensor were applied during the work for line following and for precise angle movements.

Line follower robot

Line following task is generally completed by using light/colour sensor. The robot follows the generally dark line on the light-coloured surface. The task can be solved in different ways, one solution can be the use of the difference of the colour reflectance of the surfaces or the sensor can be set to identify the black colour and if the area colour changes then the robot corrects its heading. LEGO Mindstorms EV3 colour sensor can be used for precise object identification purposes in industrial environment [9][10].
Figure3, figure 4 and figure 5 show the model and the possible solutions in the 'original' LEGO Mindstorms EV3 programming environment and in Matlab/Simulink environment. The exact setting of the robot's velocity is one of the essential points of programming as the higher the velocity is the bigger the chance of the robot to lose its line.

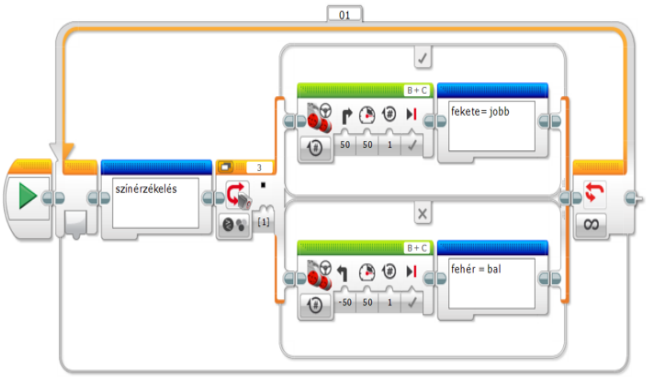


Fig. 3. Line follower robot programmed in the 'original' LEGO software

In the Simulink model reflected light intensity is measured and the dark and the light parts get own values, and these values are compared to the preset value during the program's run. The black is set to 10 out of 100 and white to 80 out of 100 which values are determined by experiments.
Movement correction is based on the difference of the velocity and the rotation direction of the two motors as the two motors velocity changes permanently while only the left motor changes its rotation direction if necessary. The LEGO robot 'brick' display shows the reflected light intensity value.

AIS 2020 • 15th International Symposium on Applied Informatics and Related Areas • November 12, 2020 • Székesfehérvár, Hungary

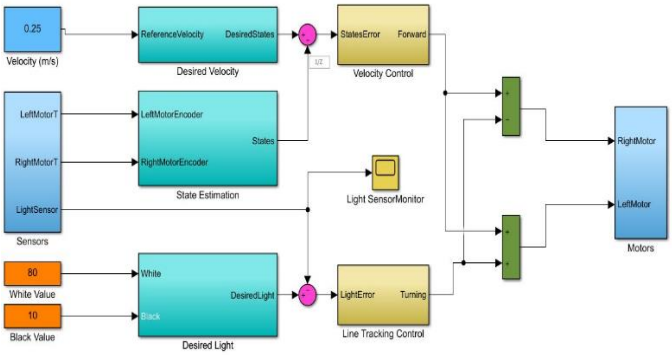


Fig. 4. Simulink model of the line follower robot

The reference velocity is set to 0.25m/s. If the reflected light intensity is low enough the LEGO robot moves straight forward and when the value goes higher the robot corrects and turns. Using the reflected light intensity method compared to the black and white colour difference method means smoother movement.

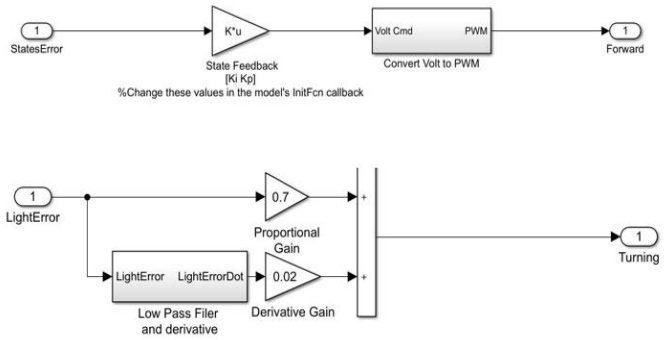


Fig. 5. Velocity and line following control in Simulink [5]

Gyro sensor application

Gyrosensor is suitable for measuring or maintaining orientation and angular velocity. Straight forward movement or exact 90 degree turning seems to be easy task for the robots but due to the environmental anomalies gyro sensor must be applied to correct these deviations.

Figure 5 shows the 'turn exactly 90 degrees' robot programming in 'original' LEGO software. The gyro sensor is set to zero first and then it checks the deviation in degrees and turns to the good direction.

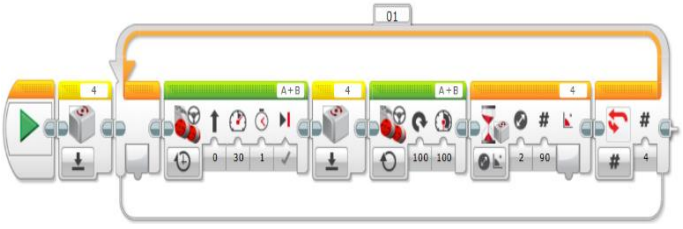


Fig. 6. Turning robot programme using gyrosensor

Figure 7 shows the Simulink model of the 'straight forward' robot where the gyro sensor permanently measures the deviation in degrees and turns the robot back to the opposite direction. The model seems to be similar to the line follower robot's model as the basic idea is the same. According to our experiences the LEGO Mindstorms EV3 gyro sensor

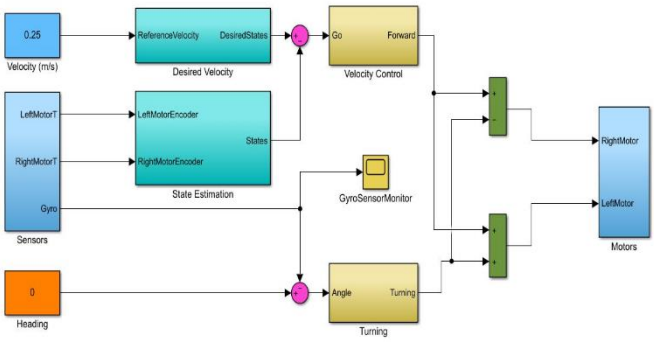


Fig. 7. Simulink model of the 'straight forward' robot

Figure 8 shows the scope of the gyre sensor while the robot turns. First the sample time of the sensor was set to 0.1 second but the scope values always turned back to zero. After setting the sample time value to 0.001 second the scope showed right values.

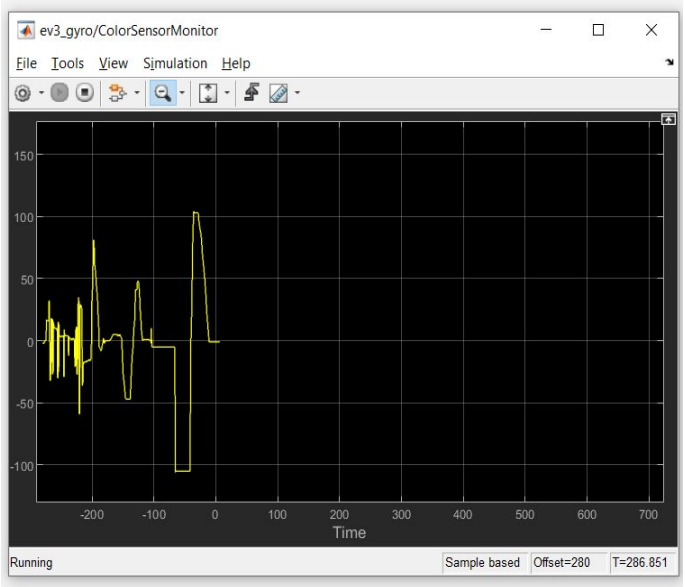


Fig. 8. Angle turning monitor

III. Conclusion

This paper was to show how LEGO Mindstorms EV3 robotic set could be applied in education from the elementary school level up to the higher education. LEGO Mindstorms EV3 set is widely used for promoting STEM.

AIS 2020 • 15th International Symposium on Applied Informatics and Related Areas • November 12, 2020 • Székesfehérvár, Hungary

A complete educational material were created from elementary level which contains the basics of robotics, helps to improve LEGO robot building skills, introduces the first steps of robot programming and shows the steps from the graphic programming method up to the high level Matlab and Simulink programming environment.

The paper showed the model used in our work which is one of the LEGO demonstration models with some minor changes and the applied sensors. The programming and application possibilities of the colour and light and the gyro sensor were demonstrated in the 'original' LEGO Mindstorms EV3 programming software and in the Matlab and Simulink environment.

It can be concluded that LEGO Mindstorms EV3 robotic set is highly suitable for demonstration and for educational purposes and can be an effective instrument for STEM popularisation.

REFERENCES

[1] V. Chaudhary et al.: An Experience Report on Teaching Programming and Computational Thinking to Elementary Level Children Using Lego Robotics Education Kit 2016 IEEE Eighth International Conference on Technology for Education (T4E), Mumbai, 2016, pp. 38-41, doi: 10.1109/T4E.2016.016.

[2] N. Montes et al.: A novel educational platform based on matlab/simulink/lego ev3 for teaching with robots, 2018. INTED2018 proceedings, pp. 975-980. ISSN: 2340-1079

[3] www.lego.com

[4] LEGO MINDSTORMS EV3 Programming Using Simulink (www.youtube.com/watch?v=dm428Q3GL6A)

[5] www.mathworks.com

[6]A. Behrens et al.: First Steps into Practical Engineering for Freshman Students Using MATLAB and LEGO Mindstorms Robots. : Acta Polytechnica Journal of Advanced Engineering, Volume 48, 2020, Pages 44-49, ISSN 1210-2709

[7] https://uk.mathworks.com/hardware-support/lego-mindstorms-ev3-matlab.html

[8] https://www.ev3dev.org/

[9]G. Lugaresi, N. Frigerio, A. Matta: A new learning factory experience exploiting LEGO for teaching manufacturing system integration. 10th Conference on Learning Factories, CLF2020, Procedia Manufacturing, Volume 45, 2020, Pages 271-276, ISSN 2351-9789

[10] G. Lugaresi, D. Travaglini, A. Matta: A LEGO ® Manufacturing System as Demonstrator for A Real-Time Simulation Proof of Concept. Proceedings of the 2019 Winter Simulation Conference, IEEE, Piscataway, New Jersey2019. 10.1109/WSC40007.2019.9004733.