

Pázmány Péter Catholic University
Roska Tamás Doctoral School of Sciences and Technology

Investigation of Spatial-temporal visual phenomena with biological- based algorithms

Tamás Fülöp

Theses of the Ph.D. Dissertation

Scientific advisor: Ákos Zarándy, DSc



Budapest

2018

1 Introduction

The world that surrounds us is full of phenomena, some of which we can directly detect, while others are indirect. The detection of a phenomenon is the first step of cognition as the process of perception. The quality of detection depends on what conclusions can be drawn from the occurrence of a phenomenon.

In biological systems, perception is an essential requirement for survival. A response to an incident can help, for example, avoid injury, collision, attack, or help navigate or find food. Perception can lead to conditional and unconditional reflex.

The same happens with built-in systems. For an event, like input to, the system is analyzes, processes and give results. In a simple system, it can only stop the operation (for example, do not cause further damage), while for a more sophisticated case, the goal may be to keep the system in a safe, controlled state with limited or without functionality (for example, railway systems, aircraft equipment). The aim is to start, continue, change, or stop an activity after detection.

Perception always begins with detection. In biological systems, the stimuli carry different information. Many stimulus responses include basic behavioral patterns of various biological systems for survival. These are unconditional reflexes as a “set of instructions” and able to provide good answers for typical emergencies and dangerous situations. The perception can be considered an important point when it comes to realizing what kind of emergency response.

Nowadays, one of the most expensive tasks is to solve image-processing tasks. A wide range of vision-related visual phenomena gives the human body immediate answers to hazards and emergencies. My research focused on what kind of solution is found in the visual system, how effective, how much it can be parameterized and how to use it, compared to the solution in the typical image processing chain. Such a common threat is the perception of looming objects, which plays an important role in both escape and navigation.

During my research, I also dealt with how an existing computer architecture can effectively implement such a biology-motivated algorithm.

2 Experimental methods

My research began with early results of the Botond Roska's Swiss research team at the end of 2008, where I saw the possibility one of the significant applications of Cellular Neural Networks (CNN). First, I dealt with the simple modeling of the retinal channel found by the research team. I worked with the basic modeling questions in my thesis in 2009; it outlined the possible directions of my research. In 2009, the article published in Nature [1] provided concrete cornerstones for model building and a more in-depth study of the model. The model's validation and calibration could be based on this.

Model analysis work was essential. First, I studied a deeper understanding of the function of the retinal channel, as well as specifying specific parameters of the model. Based on this, the model could be validated, and the model can be calibrated. I also examined the operating range of the model and the related timing conditions.

One of the aims is to implement the model on a CNN chip, so I examined the feasibility of this algorithm. The algorithm was first created on the Anafocus Eye-RIS v1.2 architecture.

The next issue of model analysis was to clarify the calibration with the mathematical support of the model. The fundamentals of the model calibration were published in 2012. This publication already included a comparison of the mouse retina model with other biological models.

During the examination of the model, several minor impacts were recognized. One of these was the determination of the direction of the approach, and one of the effects of changing the calibration parameters, based on which the issue of the recognition of pulsed light sources were also discussed.

3 New Scientific Results

The new scientific results of my work are divided into two main theses. Thesis Group 1 contains my results based directly on the research results of a model associated with the mouse ganglion cell group Pvalb-5 (or PV-5). The results of the study of side effects found in the model study were included in the Thesis Group 2. The scientific results have been published in the journal [J], book chapter [B], and refereed conference publications [C].

Thesis Group 1:

Thesis 1. I have proven that the model of Pvalb-5 ganglion cells discovered in the retina of mice can detect the looming of dark objects in the OFF receptive field. I analyzed the effectiveness of the model from an implementation point of view. Regarding the behavioral analysis of the model, I have shown that the speed and direction of objects can be estimated.

Related publications: [B], [J], [C1], [C2], [C3]

The Pvalb-5 ganglion cell discovered in the mouse retina is an OFF type cell. This type of cell sensitive only to approaching movement of dark objects. A Pvalb-5 ganglion cell obtains input signals from about 10% of the retina's receptive field. Change of the dark silhouette (growing when an object approaching) can activate the ganglion cell as an excitatory channel. The experiments also proved that the Pvalb-5 cell gives a highly inhibitory response in the ON type of change: the cell is entirely insensitive to lateral movements, so it does not give any neuronal response.

Cones behave simultaneously as exciting and inhibiting cells, depending on whether they detect an OFF or ON event. The stimulus arriving at each receptor cell (depending on whether it acts as an excitatory or as an inhibitory cell to the previous one) is summed up by the Pvalb-5 ganglion cell, which gives a neural cell response. The biological path of processing is illustrated in Figure 1.

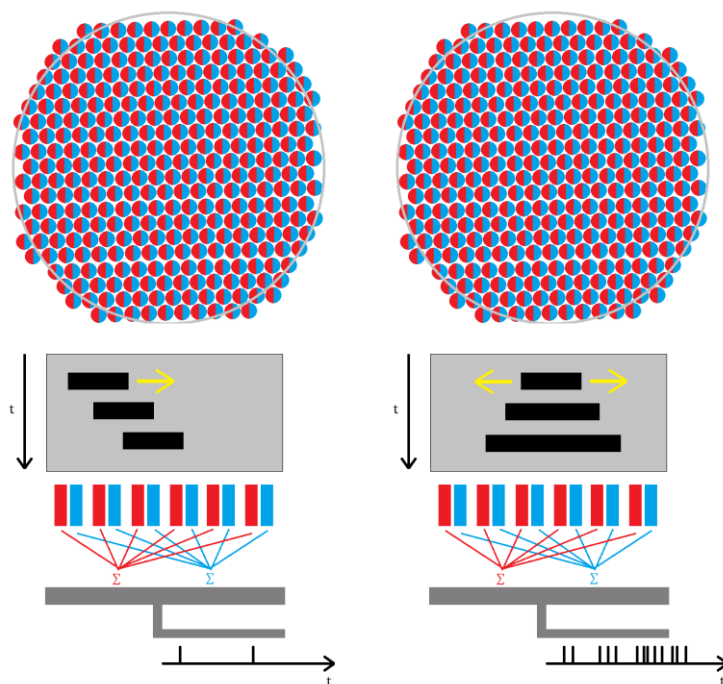


Figure 1. Pvalb-5 ganglion cell-covered receptor field. Left-side case of lateral motion, right-side looming motion (case of silhouette growth), where excitatory and inhibitory responses are aggregated on the ganglion cell.

The schematic flow of the generated model is illustrated in Figure 2. which shows that excitatory (1) and inhibitory (2) time convolution can be calculated according to the stimulus of time arriving on a receptor cell. In the case of excitation and inhibition, both linear filtering, according to (3) and (4) should be performed. The significance of this is that there is only a "positive" cellular response in nature, which is why it has a role in the inhibitory effect.

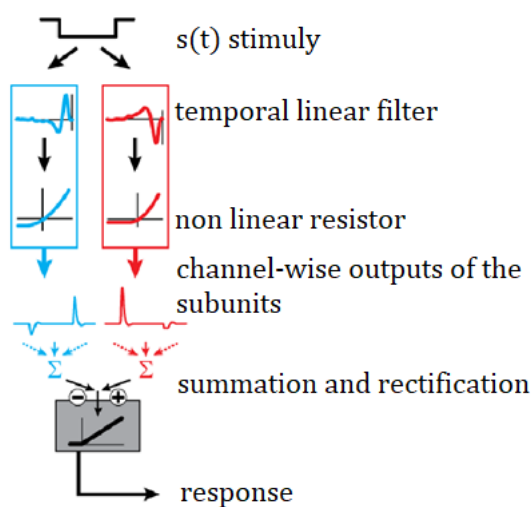


Figure 2. Pvalb-5 ganglion cell computation model for $s(t)$ stimuli.

$$e_{i,j}(t) = \sum_{n=0}^{g-1} s_{i,j}(t-n)w_n^e \quad (1)$$

$$i_{i,j}(t) = \sum_{n=0}^{g-1} s_{i,j}(t-n)w_n^i \quad (2)$$

where:

$s_{i,j}$ are the intensity values of the light reached the photoreceptors in position i,j (input);

g is the number of discrete snapshots involved in the temporal convolution;

w_n^e are the weighting factors of the temporal convolution in the excitatory channels;

w_n^i are the weighting factors of the temporal convolution in the inhibitory channels;

$e_{i,j}(t)$ is the result of the temporal convolution in the excitatory channel in position i,j (output);

$i_{i,j}(t)$ is the result of the temporal convolution in the inhibitory channel in position i,j (output).

$$h_e(x) = \begin{cases} (x+o_e), & \text{if } (x+o_e) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$h_i(x) = \begin{cases} (x+o_i), & \text{if } (x+o_i) > 0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where:

o_e is the offset value in the excitatory channel;

o_i is the offset value in the inhibitory channel;

h_e is the transfer function of the excitatory channel;

h_i is the transfer function of the excitatory channel.

$$g_{k,l} = r\left(\sum_{s,t \in N_r(k,l)} (h_e(e_{s,t}) - h_i(e_{s,t}))\right) \quad (5)$$

where:

$N_r(k,l)$ is the receptive field of the ganglion cell in position (k,l);

o_i is the offset value in the inhibitory channel;

$r(x)$ is the rectification function: $r(x) = \frac{x(\text{sign}(x)+1)}{2}$

The local results are summed up and thresholded by the Pvalb-5 cell, which also gives the estimated response to the approximation described in (5).

By examining the model, in case of the approaching the change in the receptive field changes with a positive sign (excitation), where there is no occurrence, it is neutral there. In the case of lateral movement, however, there is also a change inhibiting the excitation change.

Thesis 1.1: I have shown that the operation of the Pvalb-5 ganglion cell discovered in the retina of mice is feasible and can be effectively implemented on parallel architectures.

Related publications: [B], [J], [C1]

In formulas (1) to (5) describing the model: formulas (1) to (4) are calculations performed on a pixel, which are simple multiplications and additions, and a signum function. Formula (5) implements the spatial summary. As can be seen from the formulas, the elementary operations are simple to describe the model, and the pixel-level operations can be effectively paralleled (SIMD). The spatial summary is a function of the number of pixels and the number of available processor cores and the architectural architecture of the processor.

Thesis 1.2: I have experimentally proved that the action of mice in the retina, described on a Pvalb-5 ganglion cell can be implemented on topographic processor array with elemental CNN operations and can be effectively implemented in a speed-power-area ratio.

Related publications: [B], [J], [C1]

The Pvalb-5 ganglion cell covers 10° of the visual field, which means processing the stimulus of many cones. The field of view of the mouse is around 120-130°, covered by hundred-thousand cones, the distribution of the different types of cones in the field of view is not uniform. The difference in density is 15% per unit area.

The modeling was done on the Eye-RIS v1.2 CNN analog sensor-processor system, which contains one of the most advanced focus-plane sensor processor chips at the time of the research. The system includes a Q-Eye sensor processor chip that can take grayscale images at up to 10k fps at 176x144. By default, execution of CNN operations is supported at the command level by its functional programming language called FPP. When running a command, only the analog execution time should be counted.

The implementation can determine the ganglion cell output on the Eye-RIS at 58-98 μ s after image capture. Figure 3. illustrates the parts of each operation.

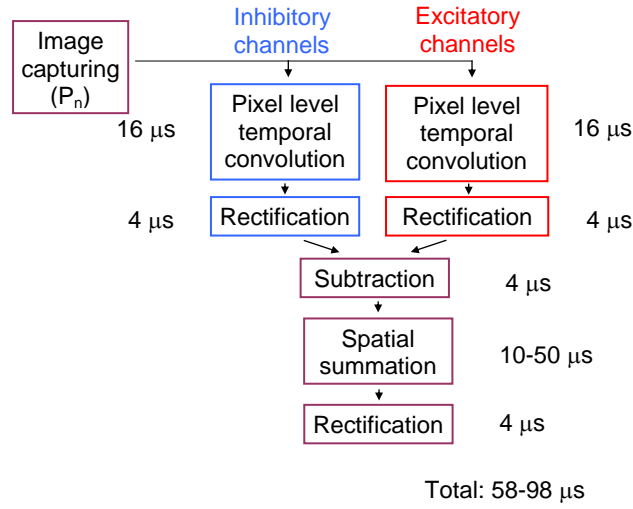


Figure 3. The flowchart of the implemented model with the time required for operations for the Eye-RIS v1.2

The ganglion cell performs a morphological summary, which is the most expensive operation throughout the process. This summary is essentially a diffusion that Eye-RIS can perform at 10-50 μ s depending on the image and other settings. Compared to digital architectures, significant gains can be made at this time. Analog diffusion time depends on background, noise, and settings. This diffuse summary implements the (6).

$$g_{k,l}(t) = h_g \left(\sum_{i,j \in N_r(k,l)} \left(h_e \left(\sum_{n=0}^s I_{i,j}(t - n\Delta t) w_n^e \right) - h_i \left(\sum_{n=0}^s I_{i,j}(t - n\Delta t) w_n^i \right) \right) \right) \quad (6)$$

Thesis 1.3: I have shown that several Pvalb-5 ganglion cells react differently on the same rapidly approaching object in different parts of the retina, and most strongly in the central center.

Related publications: [B], [C3]

The research team of Botond Roska showed [1] that the areas covered by ganglion cells overlap, that is, overlapping the receptor field. By modifying the model, I created an overlapping network and showed that it affects the middle receptive field most and overlapping in the field of visual field growth that does not exceed the size seen by the receptor field, but that the growing fields have still reacted to the growing field, thus increasing the level of activity.

Thesis 1.4. Characterized and calibrated the model of Pvalb-5 type ganglion cell function in speed-to-distance ratios

Related publications: [B], [J]

Characterization means specifying four main parameters:

1. Convolution window size (system memory)
2. Ability to eliminate small changes (what can be "dropped", thresholds for stimulation and inhibition)
3. The size of the receptor field relative to the size of the approaching object, which influences the size of the cell responses
4. The sensitivity of the system, i.e., how much does the ganglion cell respond to (threshold)

Due to the behavior and scaling of the overlapping receptor fields, I examined the relationship between the size of the receptive field, the size, and velocity of the approaching object and the distance and direction of the response to the Pvalb-5 ganglion cell response.

I have modeled experiments with approaching, receding and lateral movements and combinations of these, as well as I wrote the fundamental relationships typical of the model and analyzed the parameterization of the model as a function of these.

The model calculation was validated with the help of Plotter table by examining movements of different speed and orientation.

Thesis 1.5: I have shown that the direction of the approaching object can be estimated with the outputs of the overlapped Pvalb-5 ganglion cells.

Related publications: [B], [C3]

Since the approaching object activates the overlapping ganglion cells in varying degrees. This activity pattern happens before Pvalb-5 thresholding. It shows that this potential pattern as non-binary responses can be used to determine the direction of the approach or any other movement with overlapping receptive fields.

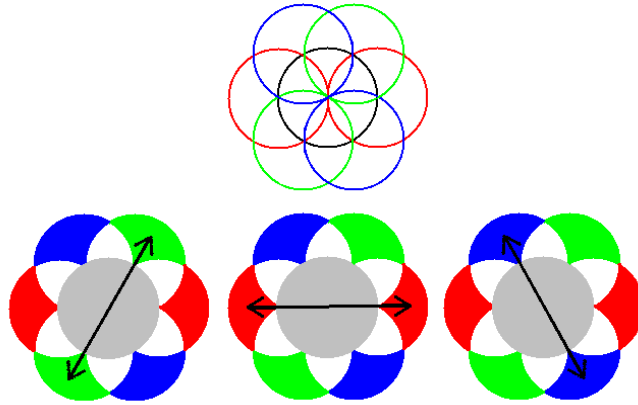


Figure 4. Potential differences between overlapping receptive fields.

This vector is essentially derived from the superposition of a receptive field and neighbor overlapping receptor fields, which can be calculated by formulas (7) and (8) for the overlapping receptor fields shown in the figure.

$$\vec{v}_e = \sum_{i \in n} a_i \cdot \vec{v}_i \quad (7)$$

where:

- a_i the potential difference between two receptive fields;
- \vec{v}_i normal vector between the center of two receptive fields.

This can be calculated in the case of the overlap shown in Figure 4. as follows:

$$\begin{pmatrix} x_{size} \\ y_{size} \end{pmatrix}_{center} = \begin{pmatrix} (rR - rL) + \sqrt{(bR - bL)} + \sqrt{(gR - gL)} \\ \sqrt{(bR - bL)} + \sqrt{(gR - gL)} \end{pmatrix} \quad (8)$$

Where r: red, g: green, b: values of blue receptive fields to the left (L) to the center or to the right (R).

I have shown that the model can indicate lateral movements and that it is suitable for the detection of lateral movements.

Thesis Group 2.

Thesis 2. I have shown that pulsed-mode LEDs emitting periodically varying light intensity can be detected by temporal under-sampling after segmentation with the HSV color range in the case of moving and stationary platform images.

Related publications: [C4]

Examination of the approach-sensitive retinal channel revealed that LEDs with periodically varying intensities and low image capture speeds influence the output of the receptive field model. This is because a continuous ON-OFF pulse hits the receptive field in a given area and thus influences the initial convolution at the pixel level. After proper color segmentation, a signal lamp can be recognized by frequency.

An important aspect of segmentation is to select as narrow a range as possible, but with all possible hits. The visible light wavelengths of LEDs are in a narrow range, so they are easier to segment and, in the case of indicator lamps, standards apply to the color and light intensity of the lamps. With LED lamps, the color of the emitted light is constant and only the brightness and saturation change.

I have examined the different color spaces for the problem of signal light recognition and found that the HSV color space is most suitable for the segmentation task. The value of H (Hue) can be considered well defined by standards, while the other two parameters depend on the environment and the imaging process.

A steady, pulse-like flashing of the LED at low frame rate means that the phenomenon is sub-sampled, the frequency of which cannot be determined due to the Nyquist-Shannon sampling theorem. However, taking advantage of the constant flashing frequency, the Goertzel algorithm can be used to detect a pulsed light source, such as a pilot lamp, even if the sample number is low.

Following the simple color segmentation algorithm, it was evident that the algorithm was capable of frequency recognition.

13,000 frames had 5,399 red markings. 86.76% of red signal images were recognized by my procedure and the number of false positive (fp) results in this simple model was around 8%. The table below gives the results

	Positive	Negative
Positive	4693(tp)	716(fn)
Negativ	1400(fp)	17304(tn)

tp: true positive, fp: false positive,

tn: true negative, fn: false negative.

This results in the following results in terms of algorithm quality for the red signal image: Precision = 0.77 Recall = 0.87

Thesis 2.1: It is also possible to provide a solution for frequency calculation on an analog CNN processor array that is capable of detecting pulsed light sources.

Related publication: [C4]

An advantage of the Goertzel algorithm is that it can be easily calculated at pixel level with sufficient amount of analog memory:

```

realW = 2*cos(2*pi*k/N); //FPGA operation
imagW = Sin(2*pi*k/N);   //FPGA operation

LAM_1 = 0;               //CNN operation
LAM_2 = 0;               //CNN operation
LAM_0 = 0;               //CNN operation
for (n=0; n<N; ++n) //FPGA operation
{
    LAM_0 = LAM_x(n)+(realW*LAM_1)-LAM_2; //CNN operation
    LAM_2 = LAM_1;           //CNN operation
    LAM_1 = LAM_0;           //CNN operation
}
outreal = 0.5*realW * LAM_1 - LAM_2; //CNN operation
outimag = imagW * LAM_1; //CNN operation

```

In the code implementing the Goertzel algorithm, the operations in the cycle are operations interpreted between LAM analog memory cells, and the cycle calls can be initiated from the FPGA, as is sinus and cosine computation. The above algorithm cannot be implemented on Eye-RIS v1.3 because the 6 analog memory cells of the analog processor array are not sufficient for computing.

4 Applicability of the results

In the field of transport informatics, like the railways, the automotive industry, or the flight, need large amounts of data. It is valid for all three systems that a completely autonomous system does not yet exist really. The closest is the railway systems, which work on a closed and well-characterized track (for example CBTC-based subways). Subways can work as an autonomous system at high speed (up to 80 kph) and ready to stop the train even ten meters before next train without radar or other distance-active technology. It uses just location data and speed measurement and knowing its own condition (speed, weight, speed curve). The railway, which is not a closed system as a subway, is a real problem to monitor the alignment of objects even when the collision cannot be avoided. In this case, the driver could escape to a more protected part of the locomotive area before a major collision warning.

For fast and accurate recognition, it is important to detect the proximity of an object in time to allow more time for a warning or intervention decision.

For autonomous airplanes, many sensors are deployed, some of which are active in radar (object sensing), but in the case of autonomous flight (UAV), the use of cameras in object research is also the subject of numerous studies.

In addition to the relatively expensive and less practical active tools such as radar and LIDAR solutions for driving support systems, in the past 15 years, countless companies have been working on solutions based on camera systems. In autonomous leadership, recognizing approaching objects is a critical and ongoing problem that needs to be addressed effectively. Most solutions currently do not dare to rely on camera detection alone, so one of the most advanced embedded EyeQ4 chip solutions on the market today includes the ability to be used with radar, although the chip itself is capable of security critical computing data supply.

For warning systems, it is important to minimize false positive signals, while autonomous systems also require high security detection, which allows only a certain tolerable level of error.

Since the beginning of my research, there has been a clear tendency for classical task-oriented segmentation to be pushed back and for deep learning algorithms and efficient processing networks to emerge.

They have similar approaches to these living organisms, but as can be seen, with the exception that certain important image processing operations are not recognized at the high (brain) level, but at a low (retina) level as with the approaching object.

The LGMD1 model, inspired by the locust and much more well known and coarser than the Pvalb-5, is still being researched in robotics for last-minute collision avoidance, probably due to its chip implementation. It is noteworthy that the LGMD2 solution, which supports more accurate recognition, is not as widespread as Pvalb-5.

Typically for Pvalb-5, the future application area could be the further development and widespread deployment of sensor sensor processor arrays, which may also be a suitable runtime environment for ConvNets, also inspired by the biological environment.

With regard to the recognition of the lamp signal image, the recognition by frequency analysis on high-speed recordings has been researched by others after my publication. Algorithms based on this recognition have a better performance than the ConvNet lamp detection methods that were born last year. My published procedure still remains unique in that it approaches the problem of light source recognition with a more traditional image acquisition device and computing apparatus, which also makes it more advantageous in computational tasks. However, the Goertzel algorithm can be implemented on a suitable analog CNN device.

My results also point out that many problems in nature are solved, and that it also aims to show the way that more and more nature can be dropped and realized.

5 Publications

Book chapter

[B] Á. Zarándy és **T. Fülöp**, „Implementation and Validation of a Looming Object Detector Model Derived from Mammalian Retinal Circuit,” in *Focal-Plane Sensor-Processor Chips*, Springer, 2011, pp. 245-260.

Journal

[J] Á. Zarándy és T. **Fülöp**, „Approaching object detector mouse retina circuit model analysis and implementation on cellular sensor-processor array,” *INTERNATIONAL JOURNAL OF CIRCUIT THEORY AND APPLICATIONS*, Vol. 40., pp. 1249-1264., 2012.

Conference

[C1] **T. Fülöp** és Á. Zarándy, „Bio-inspired looming object detector algorithm on the Eye-RIS focal plane-processor system,” in *12th international workshop on cellular nanoscale networks and their applications - CNNA*, Berkeley, 2010.

[C2] Á. Zarándy és **T. Fülöp**, „Retinal approaching object detector model implementation and validation,” in *IEEE international symposium on circuits and systems - ISCAS*, Paris, 2010.

[C3] **T. Fülöp** és Á. Zarándy, „Bio-inspired looming direction detection method,” in *13th International Workshop on Cellular Nanoscale Networks and their Applications - CNNA*, Turin, 2012.

[C4] **T. Fülöp** és Á. Zarándy, „Chromatic Spatial-temporal Frequency Analysis for Traffic Light detection,” in *4th IEEE International Conference on Cognitive Infocommunications - CogInfoCom*, Budapest, 2013.

[C5] **T. Fülöp** és Á. Zarándy, „Real-time moving object segmentation algorithm implemented on the Eye-RIS focal plane sensor-processor system,” in *International symposium on nonlinear theory and its applications - NOLTA*, Budapest, 2008.

Reference

- [1] T. Münch, R. Silveira, S. Siegert, T. Viney, G. Awatramani és B. Roska,
„Approach sensitivity in the retina processed by a multifunctional neural circuit,”
Nature Neuroscience, Vol 12., pp. 1308-1316, 2009.

6 Acknowledgement

Many patience and support for those who have been around in recent years. To get here, I could get here. It is a pleasure to have so many good people around me.

I am very grateful to Ákos for the many guidelines, suggestions, professional and personal talks, encouragement, review of my work, or the enduring of my English writings, which I have supported since 2007. In connection with the completion of our joint work, I think of Luke 15:11–32. Thank you for your continued and unconditional support.

In addition to Ákos, István Vető Teacher is very much appreciated from the professional and human side. Our valuable conversations are very missing. His humility, professionalism, attitude to engineering and his always encouraging, benevolent and humble attitude are examples to follow.

I am grateful to Professor Tamás Roska for giving me the opportunity to get to know the life of the researcher and find out what deep research and humble work are, in 2007, at the SZTAKI and in 2009 at the doctoral school. for a task. At the time of writing my dissertation I was able to experience it again. This brought back the feeling that I would not break away from the university and research circles, and that they would have a perspective on my future career.

I would like to thank Péter Szolgay and Tivadarné Vida for their continuous support, their present presence and their attention to me. Their encouragement meant a lot to finally write my dissertation.

Thanks to Zsuzsanna Vágó for failing Analysis II for the third exam. "I see you don't understand that," her words still ring in my ears. It was then and in place: my father and primary school chemistry teacher, Ágnes Kaposiné Héger, reminded me again, who always taught me to understand things, the world, and the relationships between them.

From Kalmar Secondary School to Margit Pataky, I thank you for the solid and deep computing professional backgrounds she has given.

Csilla Szombatfalvy for teaching me to write, together with loving classical music: From the Russian five, Pál Esterházy through Harmonia caelestis to Krzysztof Penderecki.

I would like to thank the teacher, the trust and encouragement of Ágnes Bérceséné Novák and Gergely Lukács. András Oláh, with whom I worked together to start university professional competition and care for young people. Attila Tihanyi and Márton Péri, who were at my disposal at any time when I asked for help in their field.

Thank you very much to my university colleagues. Particularly for those who took part and always had my fellow students as doctoral students in their everyday life, they were doing exercises with whom I could talk about professional and social issues, difficulties or even results, insights and feedback: Mihály Radványi, Miklós Koller, Kálmán Tornai, László Füredi, Attila Stubendek Gábor Tornai, András Horváth, András Kiss, Dávid Tisza, Emilia Tóth, László Kozák. I would also like to emphasize Tamás Zsedrovits, for whom I have many years of joint practice, discussion and English proofreading over the past ten years.

Thank you for taking a look at the industrial research at FETI, working together, especially for Antal Kováts, Balázs Gergő Soós and Aoyagi-san.

Thank you for supporting my current workplace, Prolan Process Control Co, and the patience of my colleagues to give me time to write my dissertation in addition to my daily tasks.

Special thanks to the support of all those who were along with me, encouraged, interested. First and foremost, Orsolya, with whom my life was bundled when I started this adventure and then became really one when I was about to finish.

Thank you very much for your patience.

To my parents. All three of my brothers. To my godparents.

To my best friends: Balázs, Csongor, Feri, Gergő.

To my close friends: Bálint, David, Erik, Tamás Fenyőfalvi, KV, Zolin Bálint, Zolin Mandácskó and my brother, Cyan.

My friends in MAVÉ and EESTEC, highlighting Franciska, Ticz Ádám, Kriszta, Krisztián, Reni, Csabi. Many times they asked me to stand with him.

I wrote the PhD dissertation in support of the European Union, using the support of the European Social Fund (EFOP -3.6.3-VEKOP-16-2017-00002).