Mammalian Retina and the CNN universal machine: Locally adaptive algorithms and examining ON-OFF interactions

Theses of the Ph.D.Dissertation

Róbert Wagner

Scientific advisor:
Tamás Roska, D.Sc.
ordinary member of the HAS

Supervisor:
Zarándy Ákos, Ph.D.

Pázmány Péter Catholic University
Faculty of Information Technology

Budapest, 2007

Introduction

The analog nonlinear dynamics based cellular neural-nonlinear network (CNN-UM) was applied for an increasingly wider range of applications in the last fifteen years. This network was especially useful in tasks where the spatio-temporal processing has an important role. Some of these applications, where the signal-processing begins already at the sensor level are of especially great importance. One of the most important branches of these sensing processing devices is the field of visual microprocessors - camera computers. In the first half of my dissertation I examine such an application: Perception of scenes with large luminance differences - high dynamic range.

Such high-dynamic range scenes have the property that their intensity range is wider than the range perceived by usual, 8-bit (256 level) cameras. Even if we use high-dynamic range cameras beside their high cost we have the problem of displaying this wide range. This is problematic because most displays have 7-8 bits, 50 dB dynamic range. Therefore dynamic range compression methods or image enhancement methods are needed. On the other hand in the perception we can exploit the fact that the high dynamic range is mostly caused by spatially unequal illumination. Such scenes are well perceivable with sensor arrays where the capturing parameters of the sensors can be set individually. The locally adjustable CNN-UM visual microprocessor is such a sensor. Using local adjustment of the pixels we can achieve that both bright and dark parts of the scene are captured with the appropriate settings. In the first part of my dissertation I discuss an integration time adjustment algorithm designed for locally adjustable sensor arrays. The method can enhance the dynamic range of any sensor by reducing the intensity differences within the image.

In the field of vision another application area of the CNN-UM wave computers is the modeling of animal, principally mammalian visual system. The architecture is suitable for this because of the similarities in the topology and processing. In the year 2001 in Berkeley based on mammalian retinal measurements, physiological and mor-

phological knowledge a novel multichannel processing of the retina was discovered. Based on this qualitative model my colleagues constructed a neuromorphic quantitative, CNN-UM model. The advantage of the neuromorphic structure is that the elements of the model are corresponding to neurons and therefore the activity of the cells can be analyzed.

By changing the weights of the connections and the characteristic of the nonlinearities in the model we can switch off retinal layers and nonlinearities. In this way we can analyze their effects on the output of the system. In the second half of my work (chapter 4.) I show such an analysis. Based on recent bipolar cell measurements I altered the earlier constructed model. Using this model I made my principal analysis regarding the role of the cross-inhibition between the ON and OFF systems. The cells of the ON pathway are sensitive to light or light increase, the OFF cells are responding to light decrease. Cross inhibition is the inhibition between the two systems.

Methods used in the experiments

In my analyzes I used the multilayer CNN architecture especially in modeling spatial processes. In the chosen applications the nonlinear, spatio-temporal computing power of the CNN-UM had an essential role.

The analysis of biological systems (e.g. rabbit retina) performing similar tasks had a principal importance for me. In the field of modeling it is indisputable the I have to use all the available information of the mammalian retina. This included anatomical, physiological, and functional knowledge as well as measurements results under different conditions. Even in the field of adaptive perception I heavily relied on the ideas originating from the retinal adaptation mechanisms.

In both of my topics the obtained results were mathematically and qualitatively analyzed. In case of the adaptation algorithms I examined the slope of the camera's response curve. The response curve shows the camera response with respect to the input light

intensity.

In some modeling tasks I used linear filter theory, Laplace transformation and Taylor series expansion.

The target architectures were simulation devices in the first line. For a fast algorithm design and testing I used Matlab environment.

Then the corrected algorithms were implemented using the Aladdin software, which enabled the portability of the code to hardware implementations. For the hardware implementation I used the ACE 16k v. 2 visual microprocessor chip. This consisted of a 128 x 128 CNN array executing the analog operations of my algorithm. In the ACE-BOX architecture this visual microprocessor is accompanied by a DSP, which executed the digital and logical operations.

In the field of retinal modeling the RefineC simulator (Dávid Bálya) was used. This simulator is also based on the Aladdin software package. The processing of the measurements done by my neurophysiologist colleagues and evaluation of the simulation results was done under Matlab framework.

Summary of the main results

Thesis 1 *Constructing CNN-UM algorithm adapting to spatially and temporally varying illumination on locally adjustable sensor array.*

Devices adjusting the capturing parameters globally can handle the intra-scene high dynamic-range with difficulty. Using local adaptation mediated by horizontal cells the human eye can handle this. In the field of CNN-UM, as the result of recent chip design, CNN-UM arrays equipped with locally adjustable sensor array were developed. Such implementations are the CACE-2k and the Xenon chips. Local adjustment means that the capturing parameters of the individual sensors can be specified independently from each other. In our case the integration time is this parameter, which means that we specify an integration time map.

My algorithm adapts to the changed illumination by changing the integration time map. As a result of the adaptation the obtained image has no areas being in dark or bright saturation. With the application of anisotropic diffusion and the local average (DC level) of the scene I extended my algorithm so, that it yields a human visual system like response. The operations requiring spatial processing were implemented on CNN-UM architecture. The method was compared to other sensors implementing local adaptation or high-dynamic range perception.

1.1 I designed a retina-inspired dynamic integration time adjustment algorithm that adapts to the spatio-temporal light intensity changes of the scene.

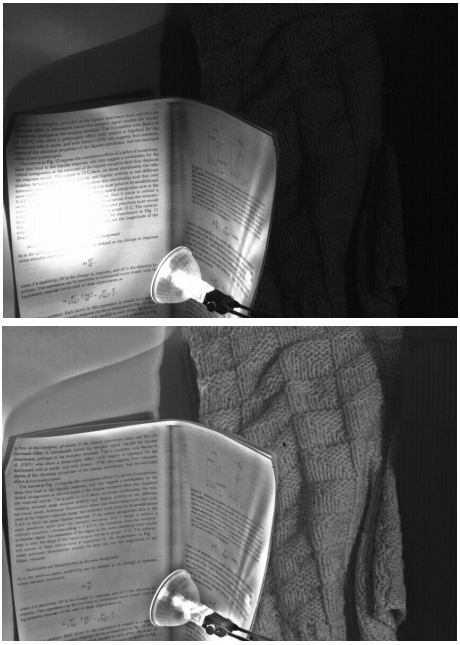
The main feature of the method is the reduction of the intensity differences of the scene's local average. Local average is the average luminance of a pixel's neighborhood, which was calculated using diffusion. The duration of the diffusion (t_d) can be calculated from the size of the image (e.g. on 128 x 128 image $t_d = 15\tau$ - CNN time constant). The σ parameter of an equivalent Gaussian convolution was 5% of the image size.

My algorithm adjusts the integration time of a sensor element in a way that the local average around that element becomes a medium gray value (half of the maximal response) (1).

$$T_{n+1}(i,j) = \overline{T_n}(i,j)\frac{V_{max}/2}{\overline{V_n}(i,j)}, \tag{1}$$

where (i,j) are the coordinates of the sensor, n is a temporal variable showing the number of the image capture. T_n and T_{n+1} is the integration time of current and the next capture respectively. $\overline{V_n}(i,j)$ is the local average of the intensity at pixel (i,j), $\overline{T_n}(i,j)$ is the local average of the integration time at pixel (i,j),

Thus the local average is driven to a gray level, its differences are eliminated. There are no regions being too dark or too bright, though the intensity differences within a region are kept. We can obtain saturated pixels, if these are far darker or brighter than the



a.)

b.)

Figure 1: Non adaptive and adaptive visual perception, using global integration time settings a.) and locally varying integration time map b.)

surrounding, but this is analogous to human perception (headlight in the night). Furthermore regions being darker or brighter than the sensor's perceivable range are obviously not adapted to gray level.

An adaptation result can be seen on Figure 1.

The method adjust the integration time continuously, hence it adapts temporally to changed illumination conditions. The flow diagram and the UMF (universal machine on flows) diagram of the algorithm can be seen on Figure 2 and Figure 3.

The adjustment of capturing parameter was chosen because in

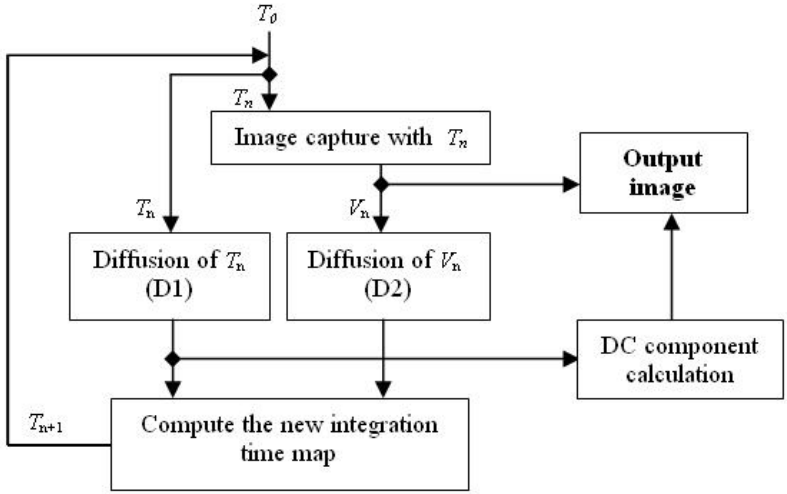


Figure 2: Flow chart of the algorithm. We capture the image with the integration time computed in the previous step (initially $T_0 = 0.2sec$). Based on this we calculate the next capture's integration time (T_{n+1}). To do this the integration time (T_n, diffusion D1) and the captured image is diffused (V_n, diffusion D2). The captured image (V_n) is retrieved, but we may add the DC component calculated from the integration time.

case of high dynamic range post processing image enhancement methods do not work because of saturation. Reducing the intra scene intensity differences correspond to the operation of the horizontal cells, the temporal adaptation is analogous to the cone adaptation.

The operation of the algorithm was demonstrated with simulations. The local adaptation was simulated with the capturing of a series of images of the same scene. The steps requiring spatial processing were implemented on analog-logical (mixed-mode) visual microprocessor (ACE-16k v.2), hence a real-time system was cre-

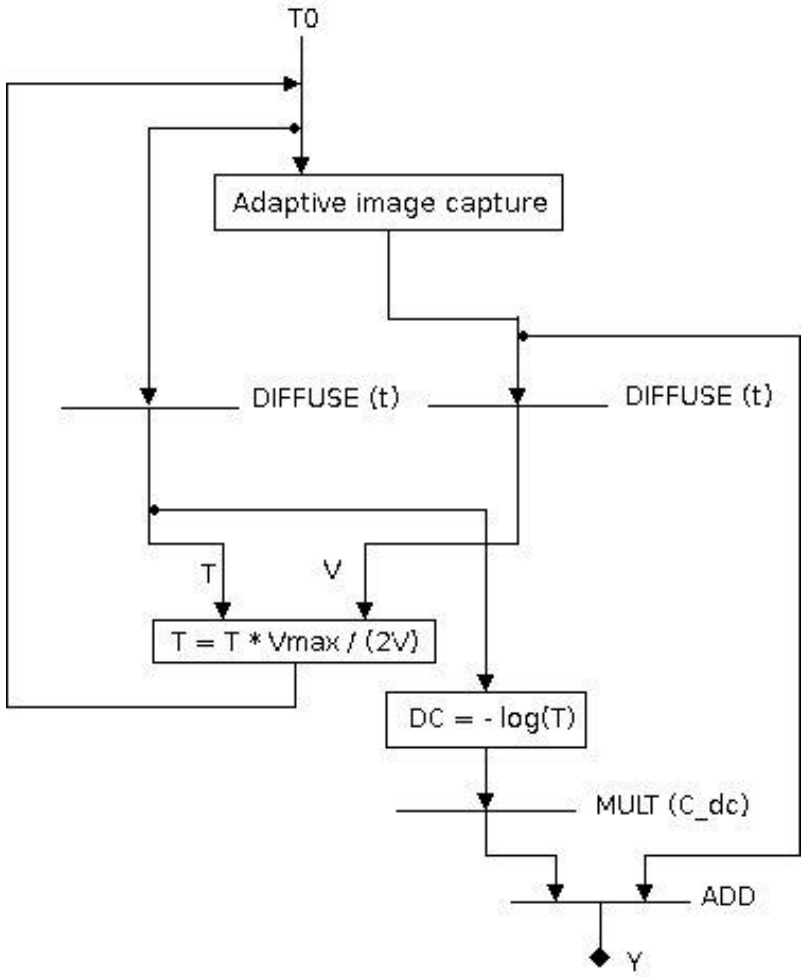


Figure 3: The UMF diagram of the algorithm. This is a detailed version of Figure 2, where the CNN templates can be seen as well. The formula computing the integration time of the next capture and computation of the DC component are also shown.

ated. Here we obtained a result video-flow with the usual 30 frame per seconds (fps)

As a result of the adaptation, the differences of the local average (DC level) were abolished. The changes in the illumination are mostly expressed in the differences of the DC component. These are eliminated by the method. As a human observer also perceives illumination changes, it is often advantageous if these are kept without saturating any part of the image. The eliminated DC component was restored from the integration time map and it was added to the captured image in a reduced manner (2) and Figure 2). As a result an image is obtained, on which the human perceivable intensity differences are visible, but the image has no saturated areas.

$$V_{DC}(i,j) = -log\left[T(i,j)\right] \tag{2}$$

As illumination often changes suddenly along shadow borders, anisotropic processes were applied for the computation of the diffusion. (D1 and D2 on Figure 2) Here the integration time difference between neighboring pixels sets a constraint on the spread of diffusion. (3) shows the equation of the anisotropic diffusion in continuous space and time. The constraint of the spread arises from the term $1/C$. The method was simulated in discrete time and space and the distortions along the shadow borders were abolished.

$$\frac{\partial V}{\partial t} = \frac{1}{C_x}\frac{\partial^2 V}{(\partial x)^2} + \frac{1}{C_y}\frac{\partial^2 V}{(\partial y)^2}, \tag{3}$$

C_x and C_y showing the contrast and

$$C_x = \frac{\frac{\partial T}{\partial x}}{T}, \qquad\qquad C_y = \frac{\frac{\partial T}{\partial y}}{T}, \tag{4}$$

where T is the integration time and V is the intensity.

1.2 Comparing my method to other high-dynamic range perception methods, I showed that it compresses the same dynamic range and furthermore it performs the dynamic range compression of the scene.

Some of the alternative methods use global adjustable sensors, where local adjustment is mimicked by capturing an image series with different settings. Another possibility is to have several sensors with different setting to a pixel. These methods loose spatial or temporal resolution. Other methods use integrating to a certain level, here the dynamic range is coded in the time domain. These methods result in high-dynamic range images, which need to be processed for displaying. My algorithm retrieves an image with perceivable dynamic range. In my method dynamic range compression occurs in a way that globally the intensity differences between the regions are reduced, but locally (within a region) the contrast is kept. Hence there is no loss of information about the texture.

The spanned dynamic range of my method depends on the used sensor. This range is the sum of the dynamic range of the perceived image at a given global parameter setting and the dynamic range of the possible integration time values. In case of spatially varying luminance the method can span 10 dB dynamic range in σdistance.

The comparison of a high dynamic range logarithmic sensor and the locally adaptive algorithm is illustrated in Figure 4.

We can consider that both the logarithmic sensor and the locally adjustable sensor can cover the whole dynamic range, but the latter has a higher rise and gives better contrast.

Compared to locally adaptive algorithms designed on CNN by M. Brendel my methods yield better contrast in originally dark regions and avoids saturated dark and/or bright regions.

Thesis 2 *Qualitative and quantitative analysis of the retinal function using multilayer-multichannel CNN mammalian retina models.*

Beside the retina inspired CNN-UM image capture, I exploited the retina modeling power of the CNN-UM. Earlier the multichannel retina model was discovered based on ganglion cell measurements and morphological knowledge. An approximative CNN-UM model was also created.

Based on bipolar cell measurements I refined the model of cone-

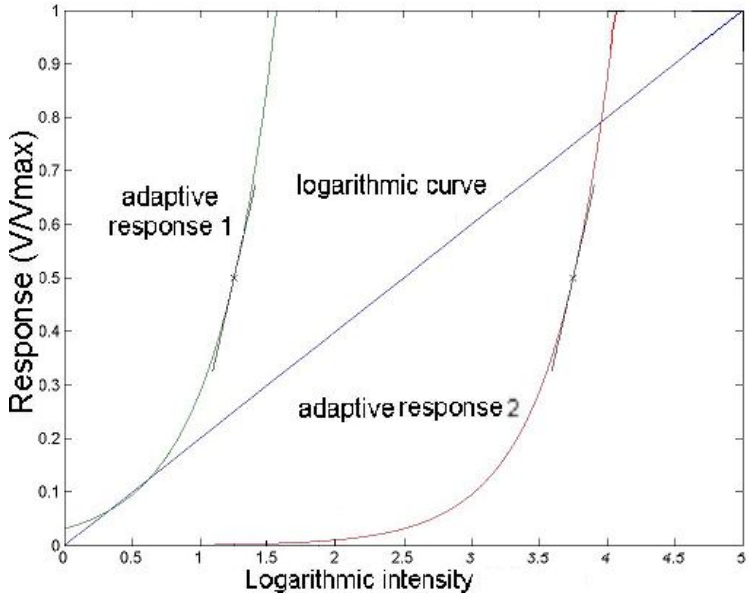


Figure 4: Response curves of a logarithmic sensor and the locally adaptive sensor at different adaptation states. The logarithmic sensor spans the whole dynamic range with a uniform rise. The linear adaptive sensor gives a useful response only around the adaptation state but there its rise is higher hence a higher contrast is achieved.

bipolar pathway with the inclusion of nonlinearities. I also extended the multichannel retina model with the cross-inhibition (out-of phase inhibition) between ON and OFF systems. Using simulation I showed the connection between the cross-inhibition and the cone-bipolar and bipolar-ganglion nonlinearities. I determined experimentally the qualitative difference between the output of the multilayer CNN-UM retina model and a model built up of linear filters.

2.1 I showed that ON-OFF cross inhibition reduces

the distortions caused by the cone-bipolar and bipolar-ganglion nonlinearities and therefore it enhances the accuracy of the ganglion cell response.

Based on rabbit bipolar cell responses to different temporal frequency sinusoid stimuli my colleagues constructed the frequency spectrum of the bipolar cell.

Based on these data I modeled the bipolar cell's excitation with a second order temporal linear filter (two poles and a zero) and a memory free nonlinear function. This corresponded to the processing of the cone-bipolar synapse. From the fundamental harmonics of the response to different frequencies I derived a linear filter. The parameters of nonlinear function's Taylor series were calculated based on the upper harmonics. The target of the calculations was to obtain a proof for the existence of cone-bipolar nonlinearities. Based on these models I extended the already existing CNN-UM retina model with the cone-bipolar rectifying nonlinearities. A simplified sketch of the retina can be seen on 5.

Rectifying nonlinearities occur mostly at synapses because the activation curve of Ca channels is exponential. Cells whose normal potential lies at the bottom of the curve can hardly give a response to negative changes. Rectifying nonlinear synapses can be found between the bipolar and the ganglion cell as well.

The existence of rectifying nonlinearities gives a proof that the retina cannot be modeled with spatial and temporal linear filters and an output nonlinearity. Such a model cannot explain certain spatial blurring mechanism. These were caused spatial low-pass filtering (ganglion dendrites) preceded by a nonlinearity (cone-bipolar synapse). In my dissertation I discuss the abolition of this blurring by the out-of-phase cross inhibition.

Out-of-phase inhibition is the cross inhibition between the ON and OFF pathways. This is realized by amacrine cells connecting ON and OFF cells. It can also act as feed-forward amacrine cells connecting the bipolar cell with the opposite systems ganglion cell. In both cases it operates through small dendritic tree glycinergic

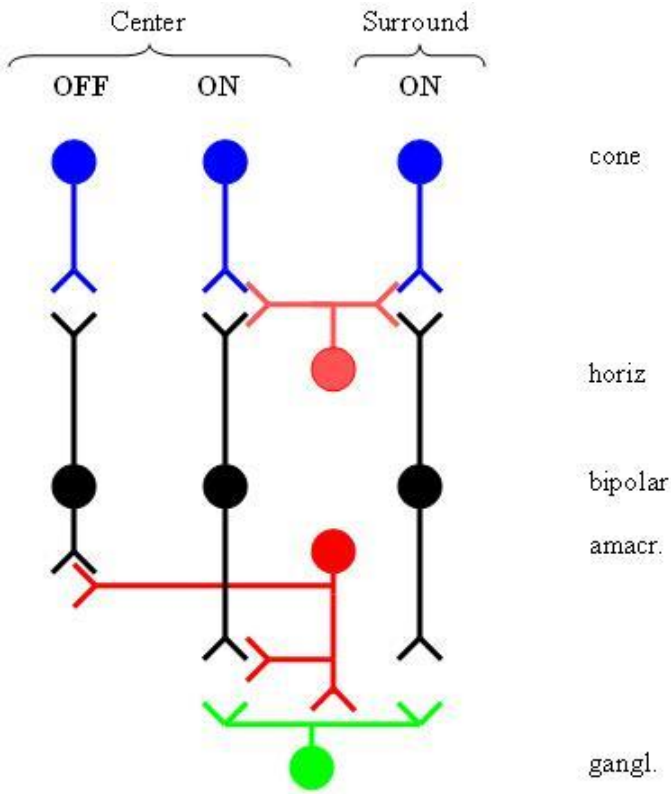


Figure 5: Sketch of the retinal cells. CNN layers correspond to the layers of retinal cells. Cells in the two right columns are the cells of ON pathway, the left column is the OFF pathway. We can see the cone - bipolar synapse, which was modeled with a linear filter and a nonlinear function. We can see the large dendritic field of the ganglion cells, which collects the output of bipolar cells thereby blurring the neural image. We can also see the amacrine mediated cross inhibition from the OFF bipolar cells.

amacrine cells. (see Figure 5)

Proof of the existence of cross inhibition are bipolar and ganglion cell measurements. In these the excitation and inhibition were measured. Comparing these two 30-40% of these cells receive inhibition originating from the other system.

The interaction of cross inhibition and rectifying nonlinearities was analyzed using flashed square and modulated sinusoid stimuli. The simulations resulted in the observation that the rectification distorts the outer-retina's spatio-temporal high-pass component enhancement. Hence the ganglion response blurs in both space and time (see Figure 6) .

Cross inhibition restores the signal distorted by the nonlinearities both at the ganglion and the bipolar level. Hence we obtain the linear signal containing the outer retina's spatial and temporal contrast enhancement. This contrast enhanced signal is more stable against diffusion, and hence the ganglion cell's response is less diffused. (see Figure 6)

Application of the results

Although both of my theses deal with vision and perception, their application is different.

The adaptive integration time adjustment algorithm can be implemented on the CACE-2k and the Xenon chips. It can also be implemented on other, locally-adjustable architecture, where a capturing parameter can be specified externally. My method can be applied for post enhancement of perceived images. In this case an enhancement gain is set instead of the capturing parameter.

The EYE-RIS chip is also a suitable architecture for the implementation of my method. Here the CNN array is accompanied by a resistive grid, where the cells can be masked which enables the approximation of an anisotropic wave propagation. On the chip's sensor array pixel values can be read out non-destructively, hence we can capture a scene with locally varying integration time. Some other

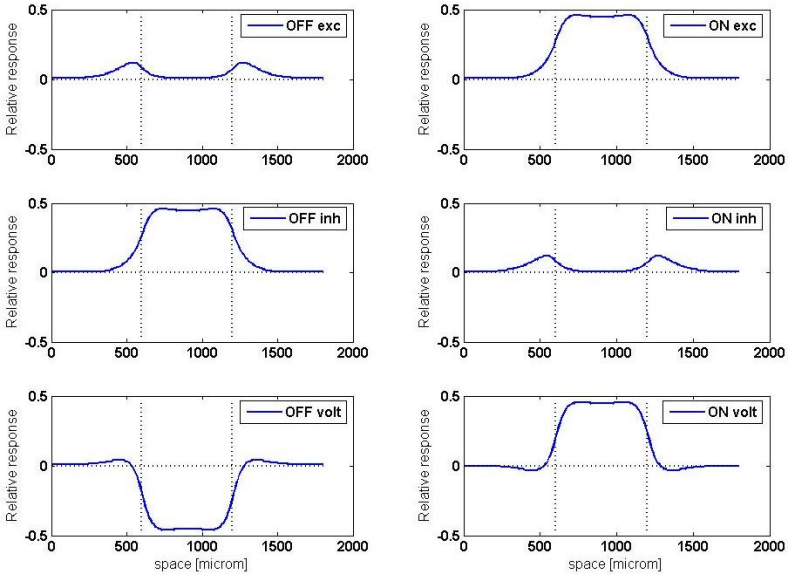


Figure 6: Ganglion outputs and cross-inhibition. The left column shows the OFF, the right the ON ganglion cells signals. Vertical dashed lines show the frames of the flashed-square stimulus. Vertical axis shows the relative response with respect to the maximal response. Top row shows the ganglion output with switched-off cross inhibition. Both in case of ON and OFF system the signal crosses the stimulus frames. The cross inhibition is showed in the second row. The last row shows the output as the result of cross inhibition. This follows the frames of the stimulus more accurately.

operations needed for the implementation of my algorithm, addition and multiplication of images are also feasible on this chip. Logarithmic characteristic needed for the computation of the DC component can be realized on the accompanying digital chip. Thus the EYE-RIS chip is an architecture, on which the entire algorithm is feasible.

The results of retinal analysis can be applied in the retinal research. The analysis of the retinal function illustrated and helped to understand a partially measured assumed operation. Using the knowledge about the output of the individual cells and their interactions, electro-physiologists can obtain new ideas to the measurement, to the used stimuli. The alteration of the model can improve its output with regard to the retinal output, and this can improve its application in retinal prostheses.

Acknowledgements

First of all I would like to acknowledge the help of my supervisor, professor Tamás Roska, who led my research with great proficiency, empathy and with unbroken enthusiasm.

I am grateful to professor Frank Werblin, who arranged me to spend two semesters at the University of California at Berkeley and gave guidance in my retinal research.

I thank Ákos Zarándy my supervisor in the field of adaptive sensing for his useful advices and his willing assistance.

I thank Alyosha Molnar the endless conversations, in which he shared his knowledge about the structure and function of the retina, I also thank Thomas Münch for his instructions.

I am grateful for the support of senior doctoral students, especially the co-work of Dávid Bálya in the field of retinal analysis. I am glad to remember the companionship of Gergely Tímár during my studies in Berkeley. I also thank his and György Cserey's help in implementation issues.

Doctoral students András Mozsáry, Attila Kis, Zoltán Fodróczi, Dániel Hillier, Gábor Vásárhelyi, Kristóf Iván, Gábor Matyi, Gábor

Hodász, Gandhi Gaurav, Gergely Feldhoffer are acknowledged for the conversation with them.

I thank the support of Péter Pázmány Catholic University (PPCU) and Computer and Automation Research Institute (SZTAKI) where I spent my PhD years and who covered the financial expenses.

I thank the guides of Zsuzsa Vágó in mathematical problems. I am grateful for the helpfulness of Katalin Keserű and Gabi Kékné from the SZTAKI and the support of the PPKE dean's office.

During my earlier studies I was supported by a number of excellent professors and teachers. Among those I would like to emphasize Gábor Horváth from the Budapest University of Technology, together with my high school teachers: Ákos Binzberger, Antal Hirka and Mátyás Rábai. Finally I would like to thank my parents for their generous upbringing, they always helped me in the best way and were patient during my education and studies.

Publications

The author's journal publications

[1] R. Wagner, Á. Zarándy, and T. Roska, "Adaptive perception with locally-adaptable sensor array," *IEEE Transactions on Circuits and Systems I*, vol. 52, no. 5, pp. 1014–1023, 2004.

[2] R. Wagner, A. C. Molnar, and F. S. Werblin, "Analysis of the interaction between the retinal ON and OFF channels using CNN-UM models," *International Journal of Circuit Theory and Application (accepted).*

The author's international conference publications

[3] R. Wagner, Á. Zarándy, and T. Roska, "Adaptive perception with local-adaptable sensor array," in *Proceedings of International Joint Conference on Neural Networks, IJCNN 2004*, (Budapest, Hungary), July 2004.

[4] R. Wagner, Á. Zarándy, and T. Roska, "High dynamic range perception with spatially variant exposure," in *Proceedings of the 8th IEEE International Workshop on Cellular Neural Networks and their Applications, CNNA 2004*, (Budapest, Hungary), July 2004.

[5] A. K. Lázár, R. Wagner, D. Bálya, and T. Roska, "Functional representations of retina channels," in *Proceedings of the 8th IEEE International Workshop on Cellular Neural Networks and their Applications, CNNA 2004*, (Budapest, Hungary), July 2004.

[6] T. Roska, D. Bálya, A. Lázár, K. Karacs, and R. Wagner, "System aspects of a bionic eyeglass," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, (Kos, Greece), pp. 161–164, May 2006.

[7] R. Wagner and M. Szuhaj, "Color processing in wearable bionic eyeglass," in *Proceedings of the 10th IEEE International Workshop on Cellular Neural Networks and their Applications, CNNA 2006*, (Istambul, Turkey), 28-30 August 2006.

[8] K. Karacs, A. Lázár, R. Wagner, D. Bálya, and T. Roska, "Bionic eyeglass: an audio guide for visually impaired," in *Proceedings of the IEEE Biomedical Circuits and Systems Conference*, (London, UK), pp. 190–193, 29 November - 1 December 2006.

Publications on topics connected to the dissertation

[9] T. Roska, "Computer-sensors: Spatial-temporal computers for analog array signals, dynamically integrated with sensors," *Journal of VLSI Signal Processing*, vol. 23, pp. 221–237, 1999.

[10] T. Roska and L. O. Chua, "The CNN Universal Machine: An analogic array computer," *IEEE Trans. Circuits and Systems, Ser.II.*, vol. 40, pp. 163–173, 1993.

[11] T. Roska, L. Kék, L. Nemes, Á. Zarándy, and P. Szolgay (ed), "CNN software library (templates and algorithms), version 7.3," Tech. Rep. DNS-CADET-15, Analogical and Neural Computing Laboratory, Computer and Automation Research Institute, Hungarian Academy of Sciences (MTA SzTAKI), Budapest, 1999.

[12] F. S. Werblin, "Control of retinal sensitivity: II. Lateral interactions at the outer plexiform layer," *The Journal of General Physiology*, vol. 63, pp. 62–87, 1974.

[13] T. Hamamoto and K. Aizawa, "A computational image sensor with adaptive pixel-based integration time," *IEEE Journal of Solid State Circuits*, vol. 36, pp. 580–585, April 2001.

[14] D. Bálya, B. Roska, T. Roska, and F. S. Werblin, "A CNN framework for modeling parallel processing in a mammalian retina," *International Journal of Circuit Theory and Applications*, vol. 30, pp. 363–393, 2002.

[15] B. Roska and F. S. Werblin, "Vertical interactions across ten parallel, stacked representations in the mammalian retina," *Nature*, vol. 410, pp. 583–587, 2001.

[16] F. R. Takahashi and C. W. Oyster, "Morphologies of rabbit retinal ganglion cells with concentric receptive fields," *Journal of Computational Neurology*, vol. 280, p. 7296, 1989.

[17] M. Meister and M. J. Berry, "The neural code of the retina," *Neuron*, vol. 22, pp. 435–450, March 1999.