Implementation and Validation of a looming object detector model derived from mammalian retinal circuit

Ákos Zarándy*, Tamás Fülöp†

*Computer and Automation Research Institute of the Hungarian Academy of Sciences, (MTA-SZTAKI) Budapest), Hungary

†Pázmány Péter Catholic University, The Faculty of Information Technology, Budapest, Hungary

Abstract The model of a recently identified mammalian retina circuit, responsible for identifying looming or approaching objects, is implemented on mixed-signal focal-plane sensor-processor array. The free parameters of the implementation are characterized; their effects to the model are analyzed. The implemented model is calibrated with real stimuli with known kinetic and geometrical properties. As the calibration shows, the identified retina channel is responsible for last minute detection of approaching objects.

1 Introduction

It is essential for a living creature to identify and avoid approaching objects, whether it is an attacking predator or an obstacle in the locomotion path. When an object is approaching, the patch caused by the projection of its silhouette on our retina is expanding. If the object is on a collision course, the expansion is symmetrical. Looming object detector neural circuit was identified in insect visual system earlier. Locusta Migratoria is exceptionally good at detecting and reacting the visual motion of an object approaching on a collision course. As it turned out, some of the largest neurons in the Locust brain are dedicated to this task [1]. After successful identification, measurement, modeling, characterization of this neural circuit of the locust, a technical team built and verified a visual sensor-processor chip for automotive application, which could detect collision threat [2] applying the same principles what the brain system of the Locust does.

The common understanding among neurobiologists was that in more developed animals (e.g. mammals) the cells responsible for detecting approaching objects are located in the higher stages of the visual pathway, most probably in the visual cortex. Therefore, it was a surprise when a looming object sensitive neuron type was identified, called the Pvalb-5 ganglion cell, in mouse retina. The identified

retina circuitry, the electrophysiological measurement results, and a qualitative mathematical model are described in [3].

We have simulated the phenomenon in Matlab environment. It turned out that the Pvalb-5 ganglion cells calculate/measure a non-linear spatial summation of the temporal brightness changes. The measured value (which depends on the approaching speed, the size, the distance, the contrast, and the color pattern of the moving object) can be interpreted as a collision threat indicator. To make quantitative analysis possible, we have generated an experimental framework using a 3D plotter, in which moving objects were recorded with known geometrical and kinematical parameters. Using these recordings, we have verified the qualitative model, calculated its parameters, and identified its operational gamut and sensitivity in different geometrical and kinematical situations.

Besides Matlab, we have made an optimized focal-plane array processor implementation of the mathematical model on the Eye-RIS [4] system as well. In this way, we made a visual approaching object detector what we can also call collision threat sensor. This device makes possible to perform real-time experiments, which is very important to characterize the model under different circumstances, because the response of the Pvalb-5 ganglion cells depends on many parameters of the approaching object. Other advantage of the looming sensor device is that it makes possible to predict the architecture of a higher level neural circuit, which evaluate the output of the modeled ganglion cells. In the future, the continuation of these studies may lead to a micro sensor devices – similar to the mentioned Locust collision warning chip [2] – which can call the attention to approaching objects.

In this chapter, the neurobiological architecture, the key physiological experiments, and the original qualitative model are shown. Then, the experimental setup, and the characterization, verification, and the sensitivity calculation are described. Finally, the efficient Eye-RIS implementation is introduced.

2 The retinal circuit

Botond Roska's neurobiologist team has identified a ganglion cell called type, Pvalb-5, in the mouse retina, which responses to dark looming objects, while it does not respond to lateral or recess motion, or static stimuli [3]. To be able to measure these cells, they had isolated a transgenic mouse line, in which only the Pvalb-5 ganglion cells were fluorescently labeled. This means that the genetically modified (labeled) cells contained fluorescent materials, which were easily recognizable and accessible in the transparent retinal tissue under fluorescent microscope. This enabled them to morphologically identify the shape and the size of its dendritic tree. It turned out that the Pvalb-5 has huge dendritic tree (~350micron diameter), which receives visual information from ~10˚ of the visual field. Under the fluorescent microscope it was clearly seen that the density of the Pvalb-5 ganglion cells is very low (Fig. 1). They were distributed equally on a

way that the dendritic tree of a cell was just reached the soma (body) of the neighboring cell.

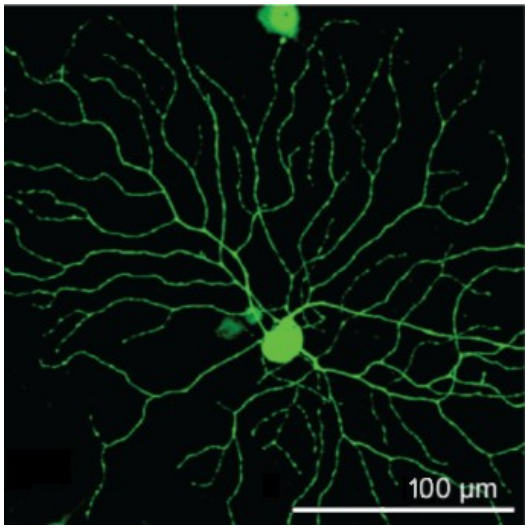


Fig. 1. The dendritic tree of the Pvalb-5 cell. The green body on the top is the soma of the next Pvalb-5 cell. The distance between the somas is roughly half of the diameter of the dendritic tree.

The behavior of the Pvalb-5 is depicted in Fig. 2. As it is shown, it selectively responses to dark looming objects, while it does not respond neither to lateral or recess motions nor to static stimuli. Since the expanding black body in the stimulus leads to an overall intensity falling (dimming) in the receptive field, the neurobiologists had to exclude that the cell provides simply a dynamic off response for the light intensity [3]. To exclude it, a specially generated pattern was projected to the retina, in which the shade of the expanding dark object was permanently lightened on a way that the overall DC level of the image did not change. Since the cell responded to this stimulus, the pure dimming detection explanation was excluded.

After extensive electrophysiology measurement series, the retinal circuitry was identified. According to the measurements, the ganglion cells average (sum) inputs coming from uniformly distributed inhibitory and excitatory channels from their entire visual fields.

3 The qualitative model

The qualitative model – proposed by the neurobiologist team – is constructed of a number of equally distributed, equal density inhibitory and excitatory channels (subunits). Each of the channels receives continuous input from one single sensor Fig. 3, hence no spatial interaction is performed at this level. The channels apply linear temporal filtering with the curve showed in Fig. 3b. The two temporal linear

filters are roughly each-others inverse. These linear filters generate inhibitory and excitatory (roughly inverted) signals inside the channels. Each channel has a rectification-like output characteristic. As it is shown, an inhibitory channel responds with a large positive signal, when its input is changing from dark to light, and generates a small negative response to negative intensity changing. As a contrast, the excitatory channel responds with a large positive signal to negative light changes on the input, and with small negative signal to positive light changes. This shows that the characteristics of the two channels are roughly opposite.

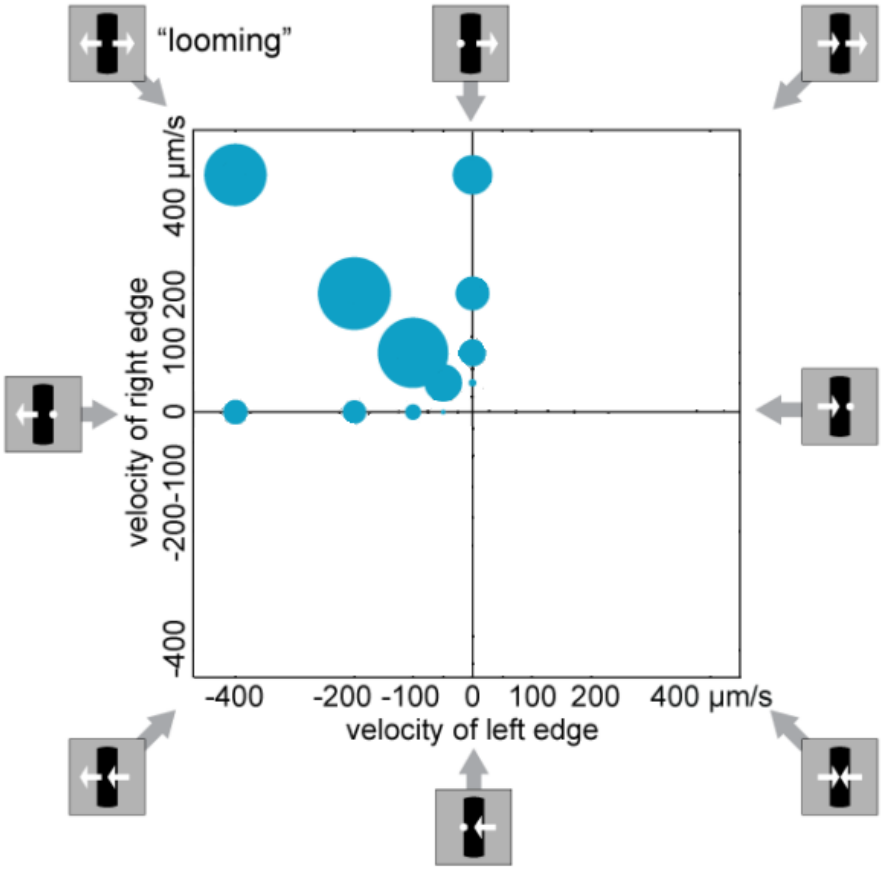


Fig. 2. Response to different dynamic patterns. The cell fired in those cases only, when the black bar against gray background were expanding (looming). It did not fire to lateral movements or shrinking (recess motion).

An engineer would ask why the retina needs two opposite channels. The most feasible answer is, that the neurons are not bipolar devices, hence the negative signals should be carried in inverted forms in off channels. Others would ask, why

the inverting channel is called excitatory, while the non-inverted is the inhibitory. The reason is, because the cell, which sums up the output of these channels is an off ganglion cell, which reacts positively to the expanding dark objects. Hence its positive input is the inverted excitatory channel and its negative input is the inhibitory channel.

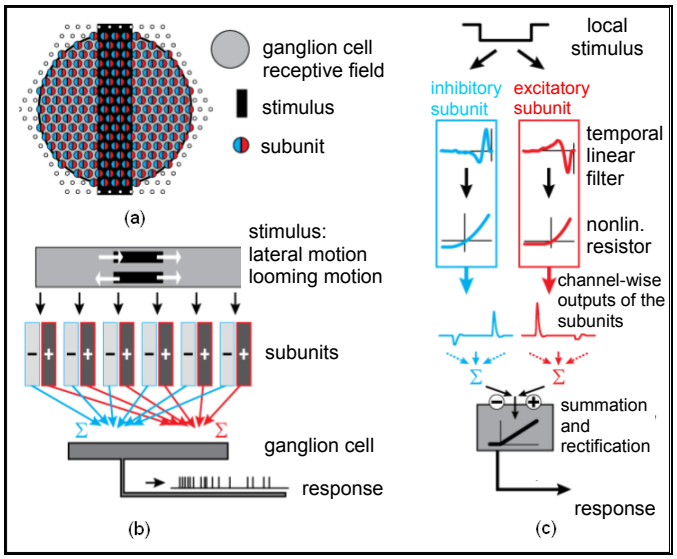


Fig. 3. The qualitative model. (a) shows the receptive field with the excitatory and the inhibitory channels, and the stimulus. (b) shows the overall behavior of the model. (c) shows the details of the channel responses.

The outputs of the channels are summed up by the ganglion cell in a circle, which covers roughly 10˚ of the visual field (Fig. 3a). In the large sum, the outputs of the inhibitory channels are taken with negative sign, while the outputs of the excitatory channels are taken with positive sign. The ganglion cell has a rectification type output characteristic also (Fig. 3c). Its output is coded in spike activity.

4 The mathematical model

The cell level signal processing in the retina can be described with mathematical equations which are continuous in time and discrete in space. Since our computers are discrete time machines we have to discretize equations in time also. In the following, we give a discrete time mathematical model which reflects the measurement results. The input of the model is the intensity of the sensed optical signal, while the output is the firing level of a Pvalb-5 ganglion cells.

The first steps of the mathematical model are the spatial filtering in both channels:

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

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

where:

- $u_{i,j}$ are the intensity values of the light reached the photoreceptors in position *i,j* (input)
- s is the number of discrete snapshots involved in the temporal convolution;
- w_n^e are the weighting factors of the temporal convolutions in the excitatory channels
- w_n^i are the weighting factors of the temporal convolutions in the inhibitory channels
- $e_{i,j}(t)$ is result of the temporal convolution in the excitatory channel in position i,j (output)
- $i_{i,j}(t)$ is result of the temporal convolution in the inhibitory channel in position i,j (output)

The spatial convolution is followed by a nonlinear transfer functions. From neurobiological aspects, the rationale of this non-linearity is twofold. First of all, it is a rectification, since the neural communication channels are unipolar. On the other hand, from signal processing point of view, it is important that it zeros the channels, which carry negative values, and those one also, which carry small positive values as well. Values around zero are generated by temporal noise, or by irrelevant slow temporal intensity changes, which should be subpressed in the spatial averaging to avoid or reduce false alarms. To simplify the mathematical model, we use a single breakpoint piece-wise linear functions mimic the measured non-linear characteristic, because this reflects rationale of this functionality. The piece-wise linear function is as follows:

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

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

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 inhibitory channel.

The outputs of the channels ($h_e(e_{i,j}(t))$ and $h_i(i_{i,j}(t))$) are spatially summed up by the Pvalb-5 ganglion cell, and rectification is applied on its output:

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

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: $r(x)=x\ (sign(x)+1)/2$.

The implementation and analysis of the mathematical model and its characteristic parameters will be shown in the next section.

5 Implementation on a focal-plane sensor-processor device

The model was implemented on a standalone vision system, Eye-RIS [4]. It is a small embedded industrial vision system, based on a general purpose focal-plan sensor-processor (FPSP) chip, called Q-Eye. The section starts with a brief description of this system, before the implementation details are introduced.

5.1 The Eye-RIS system

The Eye-RIS system (Fig. 4), developed by AnaFocus Ltd, Seville, Spain [5] is constructed of an FPSP chip (Q-Eye) [4], a general purpose processor, which is used for driving the chip and for external communication.

Fig. 4. The Eye-RIS system

The Q-EYE chip is constructed of a 176x144 sized locally interconnected mixed-signal processor array (Fig. 5). Each processor cell is corresponding to one pixel (fine-grain), hence the system can process 176x144 sized images. Each of the cells is equipped with photosensor, analog arithmetic and memory unit, and logic unit with logic memories. It can capture images, store them in its analog memories, and perform analog operation on them without AD conversion. The execution of the operators takes a few microseconds only, thanks to their fully parallel execution. Therefore the chip can perform above 1000Fps image capturing and processing (real-time visual decision making). The power

consumption of the chip is a few hundred mWs only, depending on its activity pattern. It was fabricated by 0.18micron technology. The pixel pitch is roughly 25 micron.

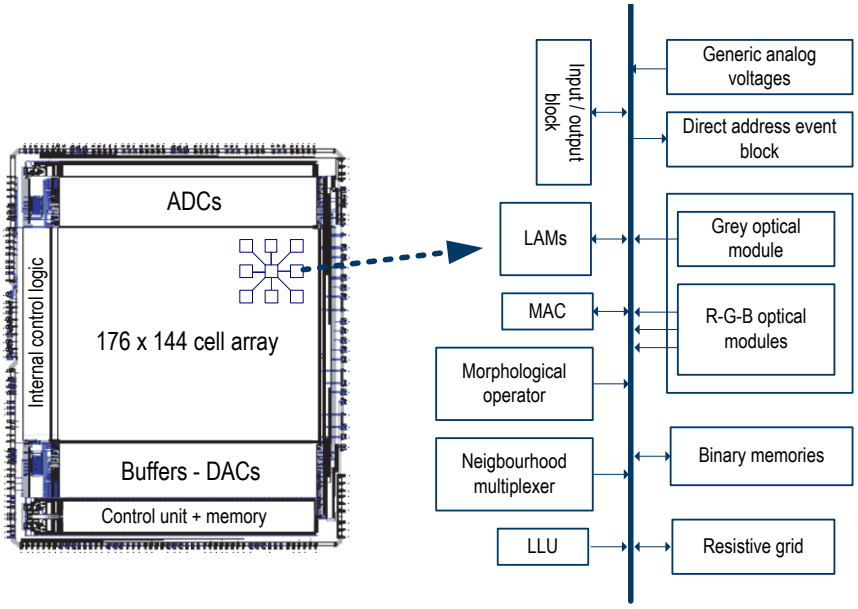


Fig. 5. Architecture of the Q-Eye chip

The functionality of the chip is summarized by the following list:

- **Grayscale**
  - Diffusion (Gaussian, directional, masked);
  - Multiple-add (MAC);
  - Shift;
  - Threshold;
  - Mean ;
  - Difference (positive, negative, absolute, signed);
- **Image capture**
  - 4 photosensors/cell for color image sensing;
  - Non-destructive repetitive readout;
  - Masking → different integration time per pixel;
- **Morphologic**
  - Arbitrary 3x3 morphologic operations;
- **Local logic**
  - AND, OR, EQU, XOR, NOT, etc.
- **Image I/O**
  - Separate grayscale and binary readout
  - Readout of a few rows

– Address event readout (coordinate of active pixels)

5.2 Implementation details

As we have seen, the first step of the mathematical model is the channel calculation. It starts with temporal convolution. We have tested various kernels and learned that the simplest convolution, which already leads to good results, is built up from the weighted summation of 3 snapshots only. We used [-½,-½, 1] weights in the inhibitory channel and its opposite in the excitatory channel (Fig. 6c). It is important to use zero sum kernels, to cancel out the DC level of the intensity of the image. Physically, this means 3 weighted pixel-by-pixel additions of the three consecutive snapshots. The temporal convolution takes 16μs on the Eye-RIS system.

Larger temporal kernels naturally lead to more accurate approximation of the measurement results. However they increase the computational complexity, require more memory and data transfer, and modify the dynamic performance of the system, because the length of the temporal convolution is increases. The length of the temporal convolution window is the first free characteristic parameters of the algorithm. The effects of the tuning of the characteristic parameters of the system will be examined later.

In the mathematical model, the second step is the application of the piece-wise linear approximation of the non-linear output characteristics. On the Eye-RIS, this is done by the addition of the offset and a thresholding, followed by a conditional overwriting of the pixels, which were below the threshold level. The operation takes 4μs. The offset values (o_e and o_i) are the second characteristic parameter of the system.

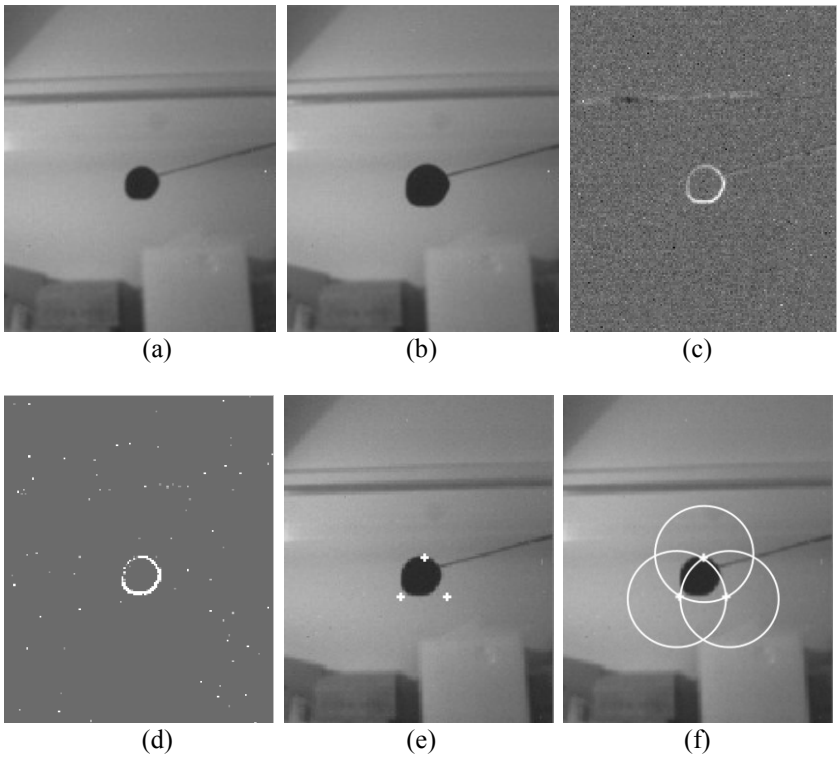


Fig. 6. Snapshots of the calculation. (a) and (b) are two snapshots of the input with an approaching black object. (c) and (d) show the response of the excitatory channels before and after the rectification. The noise canceling role of the rectification is clearly seen. (e) shows the three responding Pvalb-5 ganglion cells (r=25). (f) shows the boundaries of the receptive fields of the responding ganglion cells.

The third operation is the subtraction and the spatial summation of the output of the two channels. The spatial summation can be done in three ways.

- If the entire 176x144 image is considered as the input of a single Pvalb-5 ganglion cell we have to apply a mean instruction, which calculates the normalized sum of the whole array. This takes 12μs.

- If the receptive field of the Pvalb-5 ganglion cell is smaller than the 176x144 image, we have to calculate the summation separately in each receptive field. This can be achieved by using constrained Gaussian diffusion within each receptive field. Technically it requires the usage of the fixed state mask during the diffusion. The fixed state mask contains the boundaries of the receptive fields. Inside the receptive fields, the diffusion fully smoothens the image part, hence its DC level is calculated. In this case, the result contains the output of the multiple Pvalb-5 ganglion cells before the rectifications (Fig. 6e).
  In one step, only non-overlapping receptive fields can be calculated with this method. If we assume Pvalb-5 ganglion cells distribution as it is shown

in Fig. 7, we need to use 4 set of masks with non-overlapping boundaries of the receptive fields, to calculate the summation in each cell position. This takes 50µs.

- In the third case, Gaussian diffusion is applied to approximate the summation. The radius is controlled by the running length of the diffusion only. In this case the summation is not exact, but as an exchange, it is calculated in all the pixel positions. The calculation this way takes 10-15µs. We have measured the error of this method. (Since the exact characteristics of the diffusion function of the Q-Eye chip is not known, we could not calculate the difference analytically.) The measurement was done on a way that we calculated the spatial summation using the second and third methods, and compared the results for different running lengths. We made the comparison for different radiuses. It turned out, the accuracy is within the LSB of the system (Fig. 8), hence this fast method can be also used.

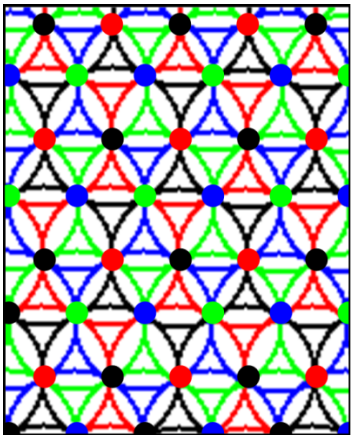


Fig. 7. The Pvalb-5 ganglion cells distribution in our model. Small solid circles are the cells, large circles are the receptive fields. Cells with non-overlapping receptive field can be calculated parallel.

The third characteristic parameter is the radius of the receptive field of the ganglion cell.

The last step of the model is the rectification. It is done on the same way as it was discussed previously. The threshold of this ganglion level rectification is the fourth characteristic parameter of the system.

The flow-chart of the calculation is shown in Fig. 9. It starts with the parallel implementation of the inhibitory and the excitatory channels. Naturally, those are calculated one after the other on the Eye-RIS, hence their time is added up. The total processing time is 58-98µs according to the selected calculation method. If the two channels use the same time convolution window, their calculation can be done in one step.

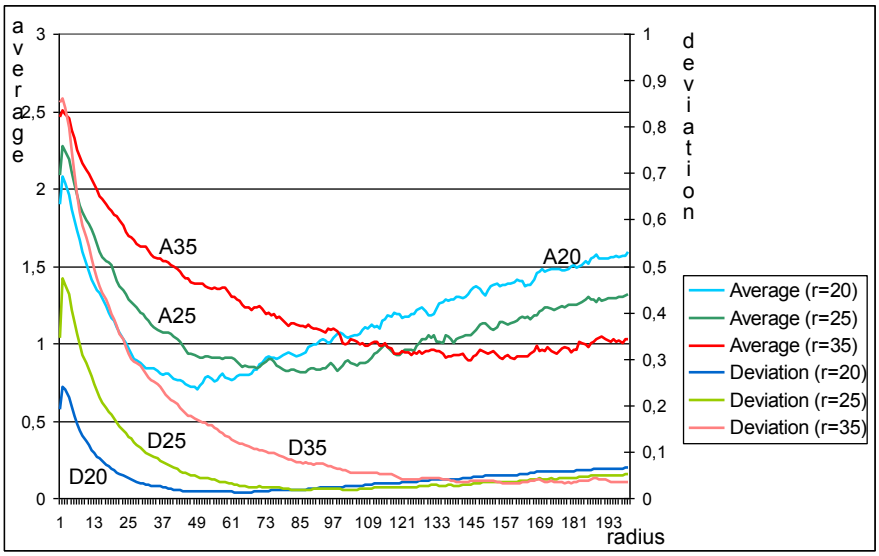


Fig. 8. Evaluation of the measurement results. The average of the absolute errors for different radiuses (A20, A25, A35) and the deviance (D20, D25, and D35) are shown. The average absolute error and the deviance is measured in the LSB of the Q-Eye chip.

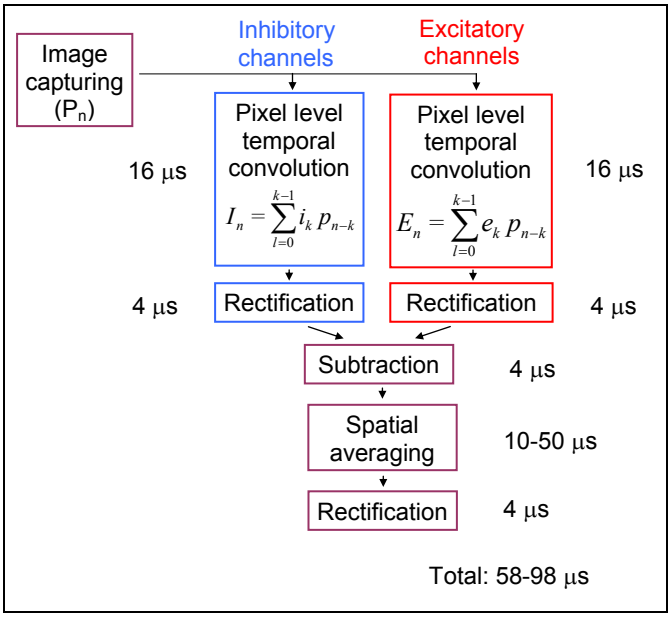
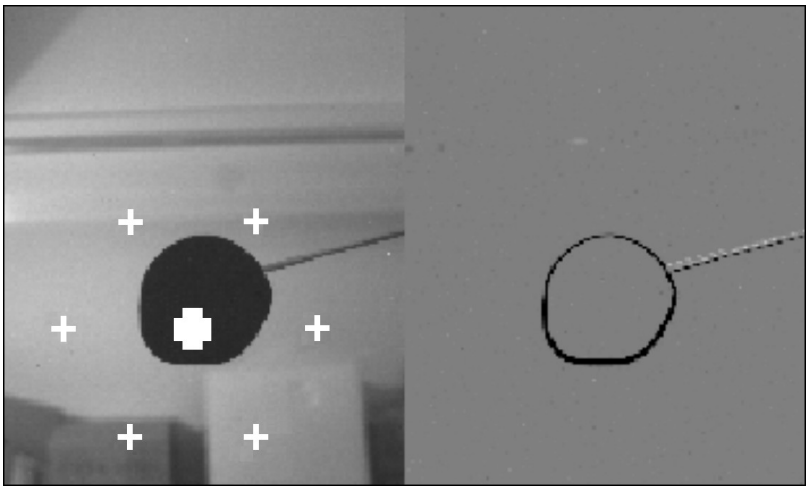


Fig. 9. Flowchart of the implemented retinal circuit model

6 Calibration

We have built a setup, which contained the Eye-RIS system and a 3D plotter, holding a black circle. By using the 3D plotter, we could generate real spatial-temporal stimuli patterns with known registered geometrical and kinematical information. Image sequences were recorded with the Eye-RIS system. We made the recordings to be able to recalculate the different stimuli patterns with different parameter sets over and over again. We did the recalculations both on the Eye-RIS system and in Matlab. Fig. 10 shows a snapshot of our calculation results. As it can be seen, there are seven ganglion cells are firing for the incoming black object. The radius of the receptive field is the distance of two neighboring ganglion cells. The ganglion cell in the middle generated the strongest respond, because all the increasing periphery of the approaching object is in its receptive field. The strength of the response is indicated with the sizes of the white '+' signs.



(a) (b)

Fig. 10. Firing ganglion cells (a) for approaching object stimulus. Combined outputs of the excitatory and inhibitory channels (b).

We have tested the model with different parameter sets, and identified the effect of the tuning of the characteristic parameters. The conclusions are as follows. The first characteristic parameter of the system (the length of the temporal convolution window) is responsible for the sensitivity and the latency. The longer the window is the more sensitive the model. However, at the same time the latency is increased with the opening of the time window.

The second characteristic parameter of the system (the inhibitory and excitatory threshold levels) is responsible for the elimination of the small changes. This parameter also tunes the sensitivity, and on the other hand, it is an excellent way to reduce the sensor noise. The effect of this threshold parameters are shown in

Fig. 11. The horizontal axis shows the intensity changes (darkening) in time, while the vertical axis shows the induced channel responses in the excitatory (upper) and in the inhibitory (lower) channels before and after the rectification. The rightmost diagram shows the combined (after subtraction) channel response. As it can be seen (Fig. 6c,d), the effects of the small changes are eliminated with the thresholds (Te, Ti). When there is a lateral movement, the same number of pixels become black at the head, as becomes white at the tail, hence they cancel out each other in the spatial summation. However, in case of approaching object, the increasing number of black pixels generates positive response only.

The third characteristic parameter of the system (the receptive field of the Pvalb-5 ganglion cells) is responsible for the size of the looming object to be detected. If it is a narrow angle, it will notice a larger distant or a small close object earlier. However it will not be able to cancel out the lateral movement of a larger object, because front and tale part of the object do not fit to the same receptive field at the same time.

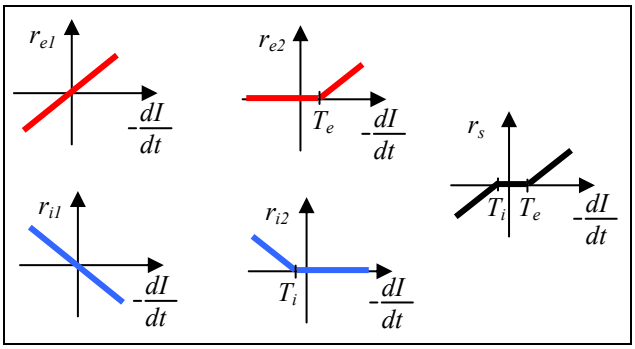


Fig. 11. The excitatory (upper) and the inhibitory (lower) channel responses to the intensity changes. The rightmost chart shows the combined response. T_i and T_e are the channel thresholds.

The fourth characteristic parameter of the system (threshold of the ganglion cells) is responsible for general sensitivity. It sets the minimal ganglion cell signal level, which is needed for the cell to fire.

From the analysis of the responses, it turned out, that the ganglion cells are not responding to lateral movement, as long as the moving object is entirely within the receptive fields. For approaching objects, we learned that the response is getting stronger as the object is approaching. Fig. 12 shows the respond characteristics to a constant speed approaching object in the function of the distance. As we can see, the response is proportional with $1/x^2$ (x is the distance from the sensor). It is not surprising, hence the response is proportional with the area increase of the projected image of the approaching object on the sensor surface, which is naturally proportional with $1/x^2$.

The strong decay of the response in the function of the distance indicates that this retina channel provides a last minute warning signal of an approaching object.

By plugging in the parameters coming from the physiological measurements made in the mice retina, and the dynamics of an attacking hawk it turns out that the ganglion cells starts responding less than 2 seconds before the predator arrives.

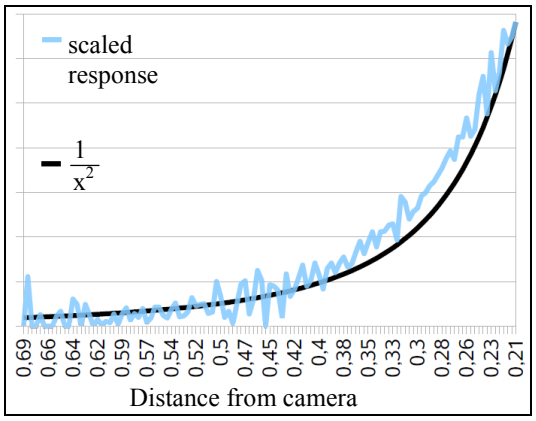


Fig. 12. Response characteristic to an approaching stimulus

It is important to note that this model and the implemented approaching object detector device responds to an approaching dark object against lighter background by nature. It is very simple to make it sensitive to approaching light objects, just by skipping the last rectification step in the ganglion cell. In that case, the large positive response is a reaction to approaching dark objects, while the large negative response indicates the approaching light object.

To understand the operation of the model, we have to discuss those situations, when the pattern on the surface of the object or the background is structured with different colors on it. In this case, the model is not behaving correctly. For example, if an object with checkerboard pattern is approaching, and the background is mid-gray, the changes in the receptive fields from gray to black and from gray to white will be in balance, hence the output will be silent as long as the individual dark areas of the checkerboard pattern dominate receptive fields. In these situations it helps, if we can somehow segment the dark or the light parts of the approaching object. In this case, we have to compute the ganglion cell response in each position (as we have seen using the third method in Section 5), and sum it app to the bright or the dark areas. However, we have to make sure that we do not include the background areas to the summation.

7 Conclusions

Recently identified mammalian retina circuit model, responding to looming object, was implemented on a mixed-signal focal-plane sensor-processor array, called Q-EYE. The steps of the implementation were detailed. The characteristic parameters of the implementation are analyzed. The implemented circuit model

was quantitatively characterized via stimulus with precisely known geometry and kinetics. It turned out that the retinal circuit is responsible for generating last minute warning signal attention call to approaching objects.

8 Acknowledgement

The explanation of the neurobiological system level background of the model by Botond Roska is greatly admired.

References

[1] C. Rind and P.J. Simmons, "Seeing what is coming: building collision sensitive neurons". Trends in Neuroscience, Vol. 22, pp. 215-220, 1999.

[2] Linan-Cembrano, G., Carranza, L., Rind, C, Zarandy, A., Soininen, M., Rodriguez-Vazquez, A, "Insect-Vision Inspired Colision Warning Vision Processor for Automotive", IEEE Circuits and Systems Magazine, Volume: 8, Issue: 2 On page(s): 6-24 2008

[3] T. A. Münch, R. Azeredo da Silveira, S. Siegert, T. J. Viney, G. B Awatramani B. Roska, "Approach sensitivity in the retina processed by a multifunctional neural circuit", Nature Neuroscience, October 2009, Volume 12 No 10 pp1308 – 1316, 2009.

[4] A. Rodríguez-Vázquez, R. Domínguez-Castro, F. Jiménez-Garrido, S. Morillas, A. García, C. Utrera, M. Dolores Pardo, J. Listan, and R. Romay, "A CMOS Vision System On-Chip with Multi-Core, Cellular Sensory-Processing Front-End", in Cellular Nanoscale Sensory Wave Computing, edited by C. Baatar, W. Porod and T. Roska, ISBN: 978-1-4419-1010-3, 2009

[5] www.anafocus.com