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

Róbert Wagner

A thesis submitted for the degree of
Doctor of Philosophy

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

Supervisor:
Ákos Zarándy, Ph.D.

Faculty of Information Technology
Pázmány Péter Catholic University

Budapest, 2007

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 PPCU 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.

I thank Cserey Zsófia the detailed linguistic correction of this dissertation.

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.

Abstract

In this dissertation (i) a retina-inspired adaptive perception algorithm is presented. The method compensates for the effects of illumination, which causes scenes to be hardly perceivable by conventional cameras. The dissertation (ii) presents an analysis of some retinal interactions using multilayer neuromorphic CNN-UM retina models. The function of these interactions was not fully understood, but could be uncovered using simulations. Both algorithms require CNN architecture as it has a large computing power and a retina-like structure.

Contents

List of Figures

LIST OF FIGURES

List of Tables

Chapter 1

Introduction

Vision is one of the most important perceptual modalities of humans. We acquire the majority of information about the surrounding environment with our eyes.

The visual pathway consists of the retinal cells, the visual areas of the brain and a thalamus which relays between the retina and the brain. Our visual impression of the world is based on the cooperation of the retina and the brain. In brief, the retina perceives the incoming light, does some basic processing steps, which enhance spatial-temporal features (e.g. motion, edges) already at the sensor level. The retina transforms the incident light into a comprehensive format for the brain. Among other processes the intensity levels of largely varying lighting conditions are projected to similar levels. The signal is coded in forms of electrical impulses, spike trains. The brain does the higher level recognition like object, and face recognition, tracking of moving objects etc.

The two main topics of this dissertation are in connection with the retinal processing:

- The first is an application of the retinal processes in the field of engineering. A locally adaptive visual sensor inspired by the retinal light adaptation is shown.

- The second targets the retina itself. Using retina models a new form of interaction between retinal cells was modeled. The quantitative role of some cells in computing the final output was clarified.

1. INTRODUCTION

My first topic originates from the retinal light adaptation. Through adaptation to luminance levels the retina performs an incredible dynamic range compression: we are able to perceive the surroundings under various lighting conditions e.g. under dim skylight or bright sunlight. We can also handle different illuminations within one scene. A striking example is when we are looking from a dark room through the window to sunlit clouds. We can perceive both the indoor shadowy part and the sunny outside. This is possible through the adaptation of the retinal photoreceptors to the observed bright or dark area. Subsequent retinal interactions enhance this adaptation.

While the simultaneous perception of bright and dark regions is an easy task for the vertebrate eye it is difficult for conventional cameras. We capture either the bright or the dark part correctly. Even if we use high-dynamic range (HDR) sensors we have difficulty in printing or displaying this HDR image on low-dynamic range (LDR) devices.

Currently there are sensors available which perceive HDR scenes. However these retrieve an HDR image, which should be rendered to LDR display devices. In this dissertation I present a method, which is inspired by the retinal light adaptation and performs the sensing of the HDR scene and outputs an LDR rendered image. In this dissertation a short comparison between my method and other HDR sensors is presented. My method can be implemented on a cellular neural/nonlinear network - universal machine (CNN-UM) equipped with a locally-adjustable visual sensor. This chip contains a network of locally coupled analog computing elements (CNN). For more information see Section 2.1. This network enables us the implementation of retina-like adaptation mechanism, due to its similar structure.

My second topic analyzes a specific retinal behavior using retina models. Earlier works (Rekeczky and Balya) resulted in the construction of a CNN-UM retina model. This model is built up of several layers of CNN cells. Each layer corresponds to a given cell type of the retina. My goal was to apply the established retinal models and exploit their power in understanding retinal behavior.

What is the use of this model with respect to retinal research? The model enables us to execute retina modification experiments in which we switch off some elements or change their parameters, thereby we observe the effects of these elements on the final output. Such experiments are very difficult to execute in "wet" laboratory conditions because pharmaceutical blockers have less accurate effects. They may affect different cells or synapses beside the desired ones.

My aim was to use the capability of the model for an exploration of the function of the cells and their role in signal processing. It was important that non trivial-interactions are examined, whose function is not fully understood. Such a process is the interaction between the retinal ON and OFF channels. ON channels respond to light while OFF channels respond to darkness, or light decrease. Cross-inhibitions between these channels were observed earlier, but their role was less clarified, because of the small overlap with the excitation of the channels. Using CNN-UM models I could show their effects in obtaining the final output, and show the distortions which are eliminated by them.

The dissertation is built up as follows: In Chapter 2 Section 2.1 discusses the structure and the basic theory of the CNN-UM and Section 2.2 shows briefly the structure of the retina. In Chapter 3 the retina inspired adaptive adjustment algorithm is shown. Chapter 4 deals with the comparison of the adaptive sensor to other methods. Chapter 5 discusses the analysis of retinal ON-OFF interactions. The References show my publications and those related to this dissertation.

Chapter 2

Used Method and Background

This dissertation has an important underlying paradigm which is necessary for the understanding: The Cellular Neural Networks (CNN). This network serves as a useful tool for modelling the retina (5) and implementing retina inspired, computationally intensive operations, which require local processing between the cells of a sensor array. The basic information about the CNN architecture is discussed in section 2.1.

Another essential background knowledge to this dissertation is the structure and the basic function of the vertebrate retina. The first part of this dissertation deals with the application of the retinal adaptation mechanism, while the second examines the interactions between the retinal ON and OFF channels.

Readers who are familiar with the concepts of CNN computers or with the retinal structure are kindly advised to pass over the appropriate sections.

2.1 The CNN Computer (a Cellular Wave Computer) - Notions and Dynamics

Cellular nonlinear/neural networks (CNNs) are regular, single or multi-layer, parallel processing structures with analog nonlinear dynamic units (cells). The state value of the individual processors is continuous in time and their connectivity is local in space. The program of these networks is completely determined by the pattern of the local interactions, the so-called template. The time-evolution of the analog transient, "driven" by the template operator and cell dynamics, represents the elementary computation in CNN (results can be defined both in equilibrium or non-equilibrium states of the network). The standard CNN equation [81] contains only first order cells placed on one layer of a regular grid and the interconnection pattern is linear.

A cellular wave computer architecture that includes CNN dynamics as its main instruction, is the **CNN Universal Machine** (**CNN-UM**, [82, 10]). CNN-UM makes it possible to efficiently combine analog array operations with local logic. Since the reprogramming time is approximately equal to the settling time of a non-propagating analog operation it is capable of executing complex analogic (analog and logic) algorithms. To ensure stored programmability, a global programming unit is added to the array and for an efficient reuse of intermediate results, each computing cell is extended by local memories. In addition to local storage, every cell might be equipped with local sensors and additional circuitry to perform cell-wise analog and logical operations.

Using the CNN-UM we are able to design and run analog and logic CNN wave algorithms. It is known that CNN-UM is universal as a Turing machine [83] and as a nonlinear operator. Therefore any kind of problem can be solved by this machine.

Its structure suggests using it for image processing in numerous applications. Beyond the classical image processing there are a lot of new methods of solving problems based on partial differential equations which need huge computational capacity. Most of these kind of problems can be transformed into CNN algorithm too.

One important scope is the biological modeling. Researchers found in early times that CNN can be used for modeling some parts of the human visual system, mainly the outer retina. Recently, a multilayer, multichannel retina model has been developed [14]. Because of the simple structure of CNN, it is realizable in real hardware. Nowadays implementations run on 64 × 64 Ace4k or 128 × 128 Ace16k chips.

2.1.1 Standard CNN Dynamics

The cellular nonlinear network (CNN) is a locally connected, analog processor array which has two or more dimensions. A standard CNN architecture consists of an $M \times N$ rectangular array of cells $C(i,j)$ with Cartesian coordinate (i,j), where $i = 1..M, j = 1..N$ (Figure 2.1)

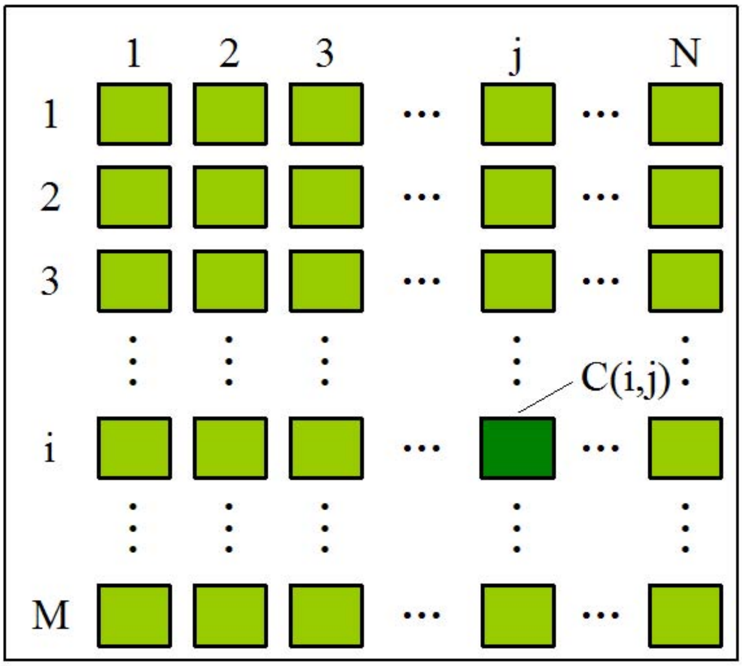


Figure 2.1: MxN representation of CNN structure

The sphere of influence, $S_r(i,j)$, of radius of r of cell $C(i,j)$ is defined to be the set of all neighboring cells satisfying the following property:

$$S_r(i,j) = \{C(k,l)|max_{1\leq k\leq M, 1\leq l\leq N}\{|k-i|, |l-j|\} \leq r\} \tag{2.1}$$

where r is a positive integer. The structure of an elementary cell can be seen on Figure 2.2.

$$\begin{aligned} I_{xu}(ij,kl) = B_{ij,kl}v_{u_{kl}}; \quad I_{xy}(ij,kl) = A_{ij,kl}v_{y_{kl}}; \\ I_{yx} = \frac{1}{2R_y}(\left|v_{x_{ij}}+1\right| - \left|v_{x_{ij}}-1\right|) \end{aligned} \tag{2.2}$$

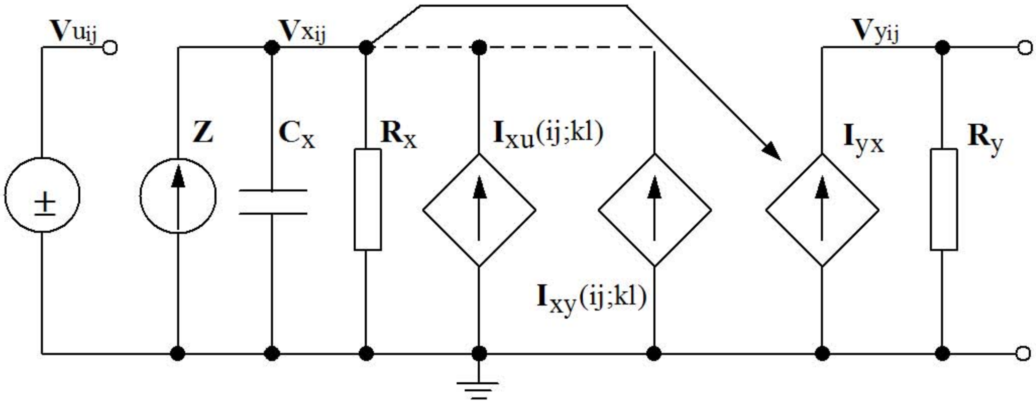


Figure 2.2: The build-up of a CNN cell

2.1.2 CNN Templates

The state of a cell depends on interconnection weights between the cell and its neighbors. These parameters are expressed in the form of the template:

$$A = \begin{bmatrix} a_{i-1j-1} & a_{i-1j} & a_{i-1j+1} \\ a_{ij-1} & a_{i1j} & a_{ij+1} \\ a_{i+1j-1} & a_{i+1j} & a_{i+1j+1} \end{bmatrix}, B = \begin{bmatrix} b_{i-1j-1} & b_{i-1j} & b_{i-1j+1} \\ b_{ij-1} & b_{i1j} & b_{ij+1} \\ b_{i+1j-1} & b_{i+1j} & b_{i+1j+1} \end{bmatrix}, z = z_{ij}. \tag{2.3}$$

A template has two main parts, a feed-forward and feedback matrix. These parts are called A and B templates. z in Eq. (2.3) is the offset (bias) term. In the simplest case the template is given by 19 numbers, 9 feedback, 9 feed-forward and one bias terms. This 19 number template is an elementary operation of CNN-UM and codes a complex spatial-temporal dynamics. An analogical algorithm might contain some templates and logical operations. The following differential equation system describes the dynamics of the network:

$$\begin{aligned} C_x \frac{dv_{x_{ij}}(t)}{dt} = -\frac{1}{R_x} v_{x_{ij}}(t) + \sum_{C(k,l)\in S_r(i,j)} A_{ij;kl} v_{y_{kl}}(t) + \\ \sum_{C(k,l)\in S_r(i,j)} B_{ij;kl} v_{u_{kl}}(t) + z_{ij} \\ v_{y_{ij}}(t) = f(v_{x_{ij}}(t)) = 1/2(\left|v_{x_{ij}}(t)+1\right| - \left|v_{x_{ij}}(t)-1\right|), \\ i = \overline{1, M}; j = \overline{1, N} \end{aligned} \tag{2.4}$$

The figure of the given function can be seen on Figure 2.3. This is called standard nonlinearity.

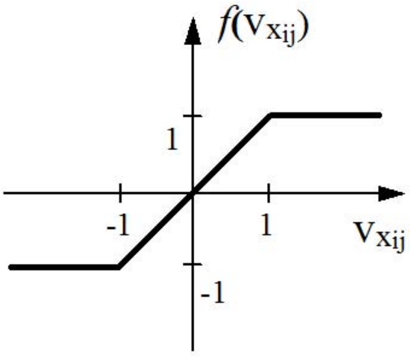


Figure 2.3: The output characteristic function of a CNN cell.

In the case where the values of $A_{ij;kl}$; $B_{ij;kl}$ do not depend on i and j, the template is space invariant. In most cases the value of the offset current does not depends on space $z_{ij} = z$. Because of the regular 2D shape of CNN, the value of a cell can be represented by a pixel of a picture. This grayscale value can be between white (-1) to black (1). Sometimes we use fixed state mask whose values allow or permit the change of the values of their cells. 3D CNN networks can connect like layers and this gives multi-layer CNN networks. Its differential equation is similar to Eq. (2.4):

$$C_{xm}\frac{dv_{x_{mij}}(t)}{dt} = -\frac{1}{R_{xm}}v_{x_{mij}}(t) + \sum_{n=1}^{L}\Big(\sum_{C(k,l)\in S_r(i,j)} A_{mn;ij;kl}v_{y_{nkl}}(t) + \sum_{C(k,l)\in S_r(i,j)} B_{mn;ij;kl}v_{u_{nkl}}(t)\Big) + z_{mij} \tag{2.5}$$

where L is the number of layers, m is the current layer, and A_{mn} and B_{mn} give the connection between n and m layers. For the solution of a given example, we have to give the input U, $x(0)$ initial state and the templates with the algorithm. The result is Y after running the transient. In most cases we can work with predefined templates that can be found in the software library [84].

2.1.3 CNN Universal Machine

The CNN Universal Machine (CNN-UM) is based on a CNN (Figure 2.4). This is the first programmable analog processor array computer with its own language and operation system whose VLSI implementation has the same computing power as a supercomputer in image processing applications [10]. The extended universal cells of CNN-UM are controlled by global analogic programming unit (GAPU), which has analog and logic parts: global analog program register, global logic program register, switch configuration register and global analogic control unit. Every cell has analog and logical memory.

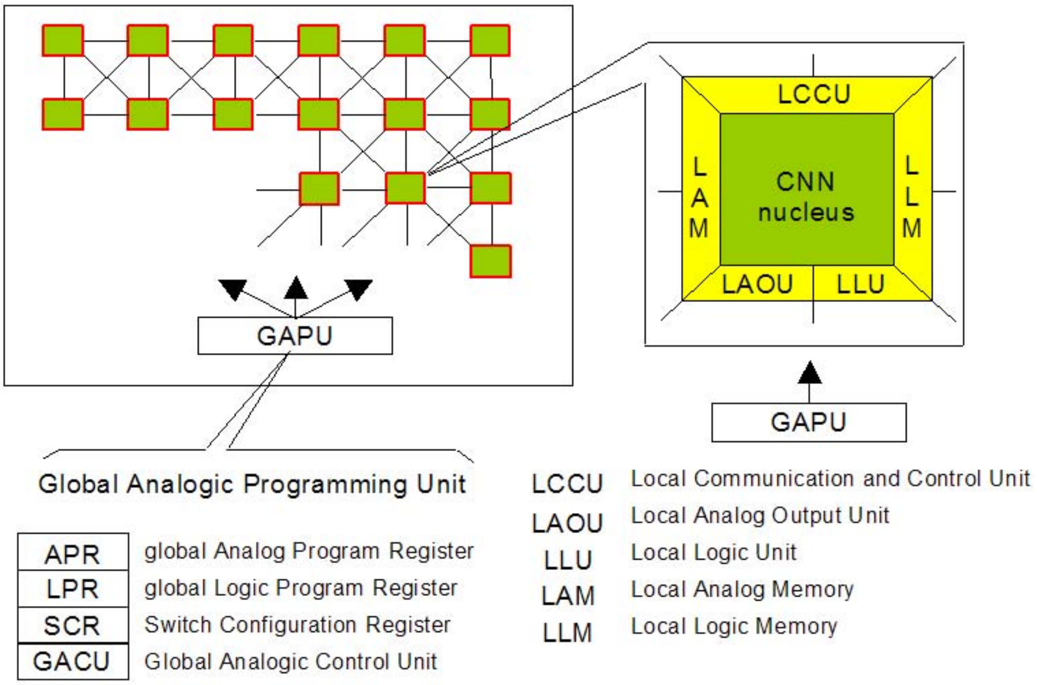


Figure 2.4: The architecture of CNN Universal Machine

2.2 The Structure of the Retina

The retina is a part of the central nervous system that transforms the incoming light to neural signals, i.e. to spikes. Besides sensing light it performs some processing tasks which facilitate the operation of the higher visual centers in the visual cortex of the brain. These processings are done by a series of cells through which the signal passes in the retina. The cell types are organized in layers which lie on top of each other (see Figure 2.5). In this chapter we give a short summary about the retinal cells and briefly mention their role.

The retina consists of three layers of cell bodies separated by two synaptic (plexiform) layers. The first synaptic layer is the outer plexiform layer (OPL), where photoreceptors (rods and cones), bipolar cells and horizontal cells interact. The second synaptic layer is the inner plexiform layer (IPL), where bipolar, amacrine and ganglion cells interact as seen on Figure 2.5.

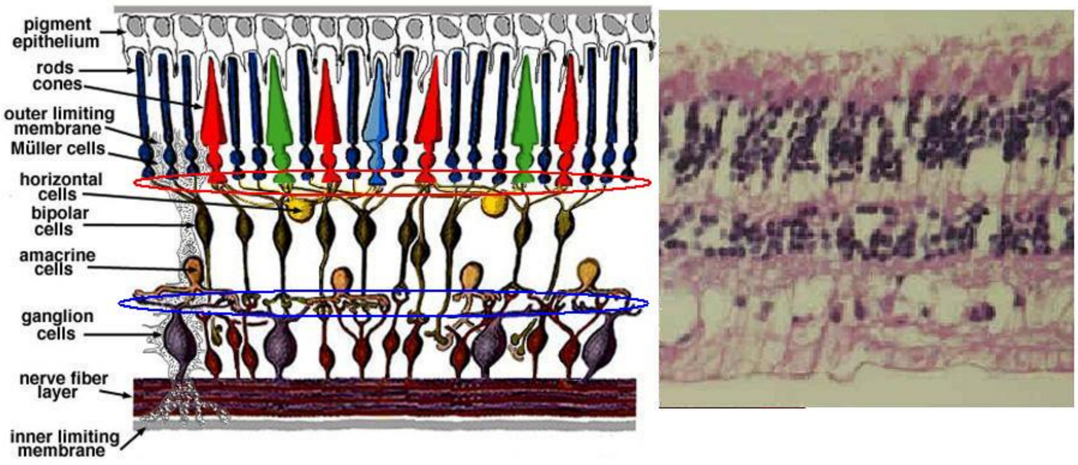


Figure 2.5: Retinal cells. (Adopted from [74, 75]) The left side shows a simple diagram of the retinal cells. Red ellipsis shows the location of the outer plexiform layer (OPL), blue ellipsis shows the inner plexiform layer (IPL). The right side shows a microscope image of a retinal slice.

Photoreceptors absorb light, they synapse to bipolars which conduct the signal to the ganglion cells. The ganglion cells are the output cells of the retina, they release spikes to the thalamus, which relays it to the visual cortex. The signal in the retina is modulated by inhibitory cells: by the horizontal cell in the OPL and by the amacrine cells in the IPL. The following sections briefly describe these cells.

2.2.1 Photoreceptor layer

Photoreceptors are responsible for perceiving light. They absorb photons and respond by a hyperpolarization (negative change) of their membrane potential. The amount of hyperpolarization is graded with light intensity. The photoreceptors have an inverse behavior: in dark they release the most transmitters in the synapse, as they hyperpolarize the transmitter release falls.

There are two types of photoreceptors, rods are active under dim light conditions (night), cones are active in daylight. Rods are monochromatic, cones have two-three types according to the wavelength (depending on the species).

2.2.2 Bipolar Cells

Bipolar cells are the cells that transmit the signal from the cones to the ganglion cells. At the cone-bipolar synapse, the ON and OFF pathways are created. The ON bipolar cells respond to an increase in the incoming light while the OFF bipolar cells respond to light decrease. This is due to the difference in sign between the cone-ON bipolar and the cone-OFF bipolar synapse (see Figure 2.6).

The OFF bipolar keeps the sign of the cone response, they hyperpolarize with the cones, this is a sign-preserving ionotropic synapse. The ON bipolar depolarizes when the cones hyperpolarize thus this is a sign-inverting synapse. Here the sign inversion is caused by metabotropic synapses in which the neurotransmitter does not cause the depolarization itself, but opens cation channels that regulate the membrane potential in a sign inverting way [74, 80].

Ganglion cells can be ON and OFF type cells depending on which bipolar synapses to them. There are ON-OFF ganglion cells as well, they receive input from both ON and OFF bipolar cells and are active at light on and off-set. They have usually a transient response (caused by amacrine cells). The ganglion cells generate spikes, which are transmitted to the higher visual centers. Spikes are released, when the ganglion cell's potential reaches a threshold level.

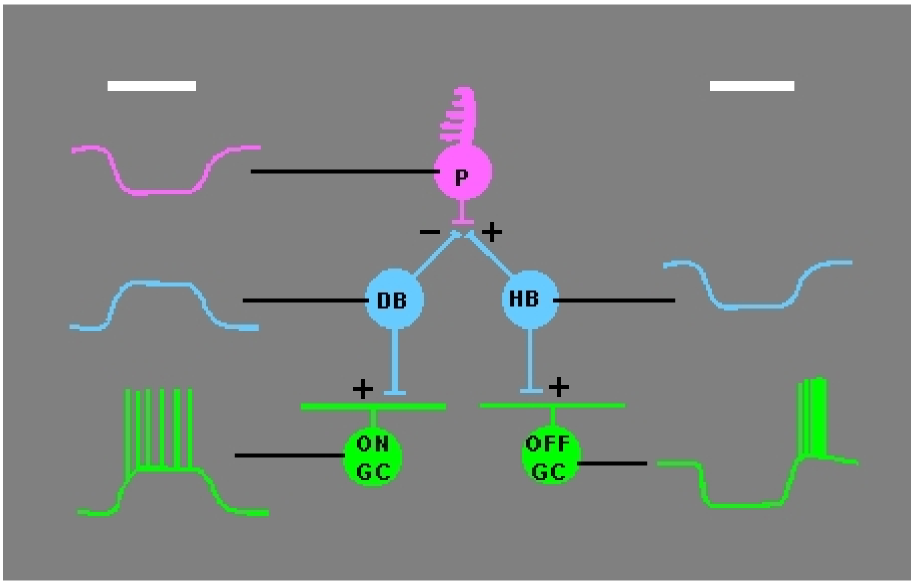


Figure 2.6: ON and OFF bipolar pathways. (Adopted from [73]) P denotes the cone cell, DB denotes the depolarizing, HB the hyperpolarizing bipolar cells. ON and OFF denote the different types of ganglion cells.

2.2.3 Horizontal Cells

The are two inhibitory cells in the retina: horizontal and amacrine cells (see Figure 2.7). The latter has several types.

Horizontal cells receive input from the cones. Horizontal cells have a wide dendritic tree. They also have gap junction between each other, which coupling increases under dim light [76]. Thus the horizontal cells integrate the output of cones over a large region, thus performs a spatial low-pass filtering of their output. This signal is fed back to the cones ([20]) and thus suppresses the spatial low-pass component of their signal and perform a high-pass component enhancement, which results in enhanced edges and the center-surround antagonism. The center is positive because there the cone are excited by the incoming light, the surrounding area is negative because the receptors only receive inhibitory horizontal cell feed-back. This is the bipolar input, on Figure 2.8 we can see the enhanced edges and the negative signal in the surround.

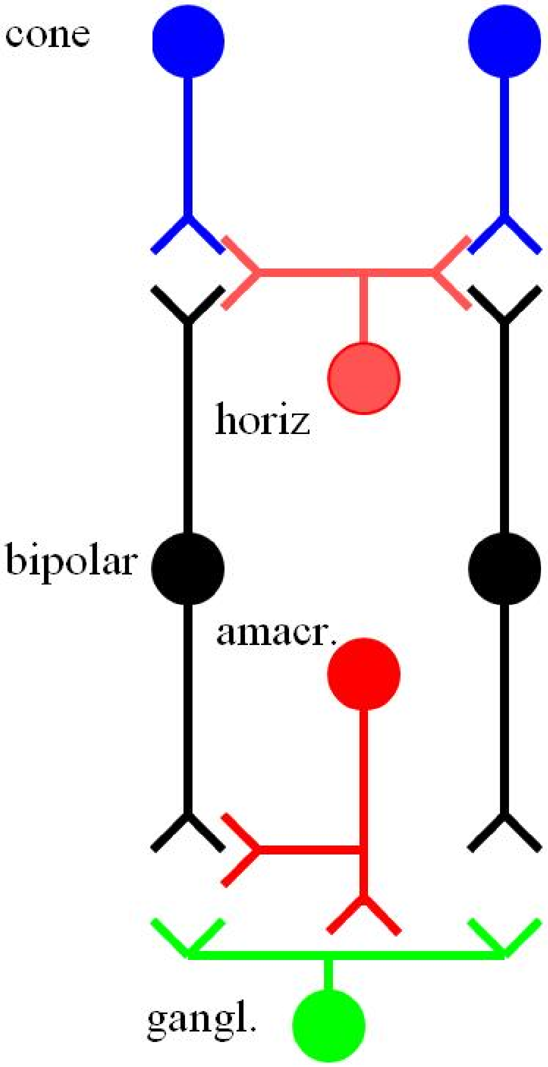


Figure 2.7: Cell types in the retina. The cells with red color are the inhibitory cells.

2.2.4 Amacrine Cells

Amacrine cells serve various functions and have several types [64, 63]. Generally they receive input from the bipolar cells and then they either make a feed-back inhibition to the bipolars or they feed-forward to the ganglions and provide inhibitory input to them.

Narrow field amacrine cells take part usually in the temporal processing, they

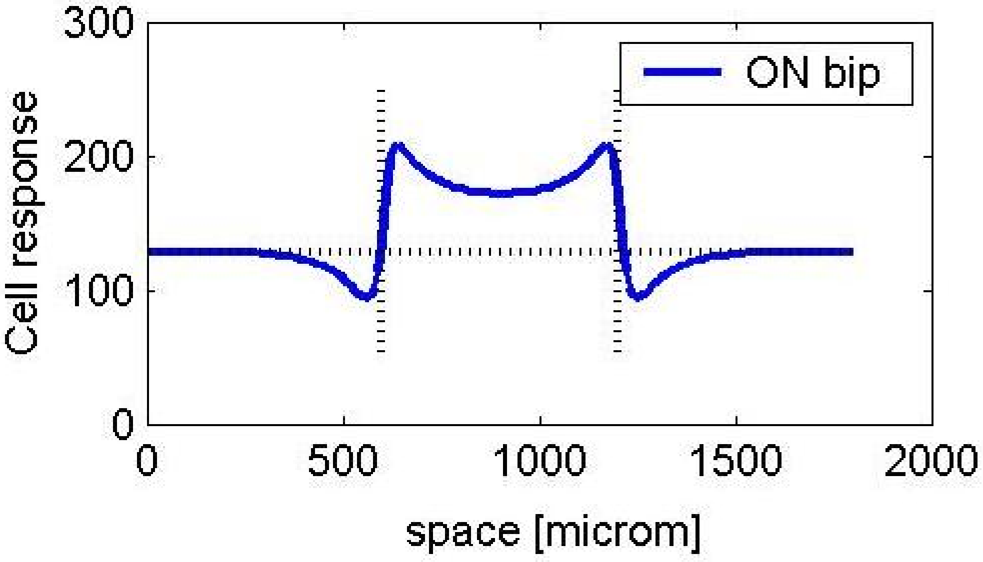


Figure 2.8: Spatial profile of a bipolar cell in response to a flashed square. The frames of the square are shown with vertical dashed lines.

make a delayed feed-back to the bipolar cells, thereby reducing their sustained response and enhancing sharp changes. Wide-field amacrine cells collect output of bipolar-cells and these cells have a role in suppressing the output in case of sudden eye-movement (saccadic suppression, [78]) amacrine cells have a role in dynamic range compression [77] as well.

2.2.5 Ganglion Cells

As noted earlier, ganglion cells are the output cells of the retina, they have ON, OFF and ON-OFF types regarding the sign of their response. Besides this several types of ganglion cells were identified based on morphology and function [16, 15].

Ganglion cells also preserve the center surround properties of bipolar cells. Given a stimulus the center is the area which excites the ON retinal cell, the surround (usually around the center) is the area which inhibits it. Surround is also used for cell which lie outside the stimulus location, and receive the signal through wide field inhibitory cells. In case of OFF cells the signals are inverse, the cell is excited by a surround stimulus and inhibited by a center (see Figure

2.9). In the temporal domain the ON cells respond to stimulus onset, the OFF cells respond to stimulus offset.

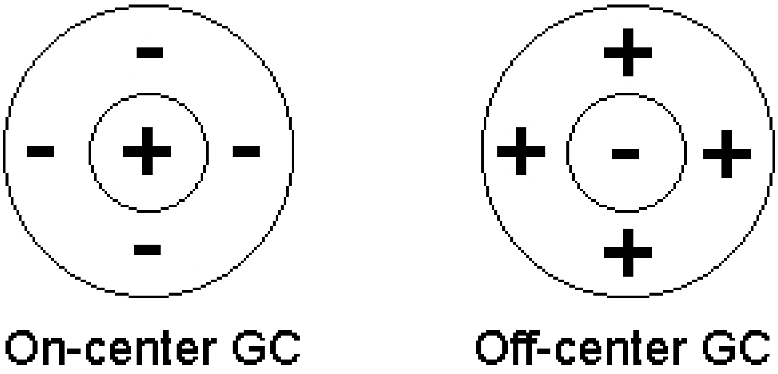


Figure 2.9: Receptive field of ON-Center and OFF-Center ganglion cells. ”+” signs show the areas which excite the cell if they are stimulated. ”-” shows inhibitory areas.

Chapter 3

Locally-Adaptive CNN-UM Algorithm

In this chapter a locally adaptive algorithm is proposed, which enables us to perceive scenes with high spatial and temporal intensity differences. The method is designed for a locally-adjustable sensor array, on which we can specify the exposure time of the sensor elements individually on the pixel level and thus adapt to spatially varying illuminations. The method is inspired by the adaptation mechanisms of the biological visual systems - i.e. the mammalian retina. The calculations required for the adaptive capture can be executed on a Cellular Neural Network - Universal Machine equipped with a locally adjustable sensor array.

3.1 Introduction

Image capture using locally adjustable sensors is crucial in some cases because there might be large intra-scene luminance differences, which is mostly due to unequal illumination. Equally illuminated scenes have 2-3 log orders (8-10 bits) dynamics only, while unequally illuminated scenes - where the illumination changes with the location - might easily bear with 5 decades (17 bits) of dynamics. Perception of these high dynamic range scenes is an easy job for our eyes (see section 3.4), however it is a great challenge for digital cameras.

In case of globally adjustable sensors every pixel has the same capturing setting. If these are sensors with linear response curve, they typically span 8bits of the input light intensity range. This means that some parts of the 17 bit input dynamic range will have constant zero or maximal response so they become black or white saturated. Usual logarithmic sensors may span 17 bits of dynamic range, but their image is noisy and of low contrast. High dynamic (16 bits) range linear sensors have better quality images, but besides their high price their drawback is low contrast and the visualization problems of the 16 bit image on a usual 8 bit screen. Similarly even if we use various ways of high dynamic range (HDR) perception ([18] or see Chapter 4) we have the problem of displaying the HDR image (of 5 or more log orders) on a screen that can display only 2 log orders. The human visual system can perceive 9-10 log orders, but this is possible because of the adaptation mechanisms in the eye (retina), which result in an output signal of 2-3 log order dynamic range. If we use similar adaptation mechanisms and dynamic range reduction in our sensors, then we can compress the HDR scene in a way that a human observer will not perceive a loss of information.

In my work such an alternative solution for these problems is presented. Using locally adjustable sensor array with linear sensor elements I will exploit the fact that the large dynamic range of these scenes is due to unequal illumination. We can eliminate these illumination differences by having locally varying exposure time settings. I present a method that adjusts the integration time parameter of the individual pixels, so that we can perceive the large dynamic range scene. The method is based on adaptation processes of the vertebrate retina and it can be implemented on a CNN-UM.

This chapter is organized as follows: Section 3.3 deals with the physical background of the adaptation algorithm. In section 3.4 the underlying retinal processes are reviewed. In section 3.6 the simulation of the locally adjustable hardware is shown. Section 3.7 shows the integration time adjustment algorithm. Sections 3.8 and 3.9 show the modification of the method with the DC component restoration and the anisotropic diffusion.

3.2 High-Dynamic Range Scenes

High-dynamic range perception as a problem can arise in everyday photography. The best photos can be made in cloudy/shadowy environment or at sunset. In case of bright sunlight shadow boundaries might be obtained with a high probability. Under artificial lights similar problems arise if both well and poorly illuminated areas are present.

Such a unequally illuminated photo can be seen on Figure 3.1 a.). Illumination differences cause dark values for large parts of the photo.


a.) b.)

Figure 3.1: Image with different illumination conditions. a.) shows the original image with capturing parameters optimized for the bright mountain slope. b.) is the result after applying a logarithmic transfer curve for the intensity values of a.). Here the dark intensity range is enhanced for the cost of compressing the bright range.

There are methods which perform enhancement operations on such images. The most fundamental operation is the so called gamma correction, where a logarithmic function of the intensity values is taken, and the result is normalized for the display range. Such curves enhance the dark values for the cost of lowering the range of bright values: see Figure 3.1 b.).

There are other operators which exploit the fact that the illumination is location dependent, and perform a selective or locally varying enhancement. These methods are further discussed in chapter 4.

However such enhancement is not always possible. If the illumination is especially uneven the dark part cannot be enhanced so easily. Figure 3.2 a.) shows such an example. In case we apply the previously mentioned gamma correction the result is not as sufficient as it was before (see Figure 3.2 b.)).


a.) b.)

Figure 3.2: Scene with large illumination differences. a.) shows the original image, b.) is the result of the gamma correction.

The enhanced dark parts are too noisy (e.g. the radiator). This is simply because there is not enough information in image 3.2 a.) and with the high level of enhancement the noisy became visible. In such cases the question of perceiving the scene and gathering the necessary information is more crucial than the visualization of the dark region. We might use several types of image enhancement methods, the basic problem remain the same, information about the dark (sometimes even saturated) parts is lacking. In case we choose higher integration time, for the capturing, we might obtain nice results for dark areas, though the bright areas will become saturated. This is already visible on Figure

3.2 b.), because as the poorly illuminated areas became visible, the bright outdoor region is driven close to saturation.

These phenomenon might illustrate for us that there is a need for a different capturing mechanism. Image enhancement methods are not sufficient for us, we need to capture dark pixels differently, because otherwise the necessary information is not available and/or we enhance the noise. A solution for this problem can be the application of retinal adaptation mechanisms. It is hard to show in images, but such scenes as Figure 3.1 and 3.2 are well perceivable for our eyes. In Chapter 3 we will present a method that is able to perceive scenes with large illumination differences. Our method is based on the retinal adaptation mechanisms. Here we try to mimic the retinal mechanism and provide a low-dynamic range percept of the HDR scene.

There are other methods and sensors to perceive HDR scenes as well. We choose the mimicking of retinal processes, to observe its benefits and because it compresses the HDR environment in way that is akin to human perception. In Chapter 4 we compare our algorithm to other sensor and dynamic range compression methods.

3.3 Physical Background

As mentioned, the large dynamic range is mostly due to the illumination differences. In this section I will describe how the illumination affects the perceived image. The incoming light falls on the objects of the scene, and then it is reflected, and the reflected light reaches our eye. There are two types of reflection, the specular and the diffuse reflection ([25]). In this thesis the diffuse reflection (or Lambertian reflection) model is used ([26]), because it is a good assumption in most of the cases and using it we can efficiently reduce the effect of the illumination. According to the Lambertian law (3.1), the diffused light from an object (I_D) is the product of the incoming light (I_L), the reflectance of the object (R) and the cosine of the angle (α) between the object normal-vector and the light (see (3.1)).

$$I_D = cos(\alpha) \cdot R \cdot I_L \tag{3.1}$$

Looking at a scene we are interested in the reflectance, because this is the measure that carries information about the objects of the scene, the illumination variations (I_L) are less important for the observer. It is important to notice that the illumination is a multiplicative factor, thus to eliminate its effect the perceived light should be divided by the illumination or the estimated illumination level.

In section 3.4 we review how this is done in the retina.

3.4 Retinal Light Adaptation

My method is based on the retinal light and dark adaptation. This adaptation compensates for huge spatial and temporal light intensity changes. To understand this we have to examine the structure of the retina first.

The retina consists of several layers, from which mostly two layers take part in the dark and light adaptation. These are the photoreceptor and the horizontal cell layers, which synapse in the outer plexiform layer (OPL) with the bipolar cells. Here we consider only this part of the retina, because this plays a major role in the light adaptation. For more information about the structure of the retina consult Section 2.2.

Our cone photoreceptors perform logarithmic compression of the input. The cones can sense a dynamic range of 9-10 log orders, and reduce it to 2-3 log orders [20]. The reduction is achieved by shifting the operation point of the response curve to the appropriate dynamic range [20]. An example can be seen on Figure 3.3. The adaptation state of the photoreceptor is shifted to the background intensity level. The background intensity level is set by a stimulus shown over a large area for several seconds prior to the actual stimulus. This level depends on the previously shown stimulus. Sudden changes evoke strong output from the photoreceptors, but if this remains prolonged the receptor shifts its response curve.

This means that the photoreceptor adapts to the changing illumination conditions. The change of the adaptation state is a temporal process: it eliminates the intensity differences due to time varying illumination.

The cone adaptation also eliminates some parts of the spatially varying illumination [49]: as we stare at the highly illuminated part of an unequally illuminated

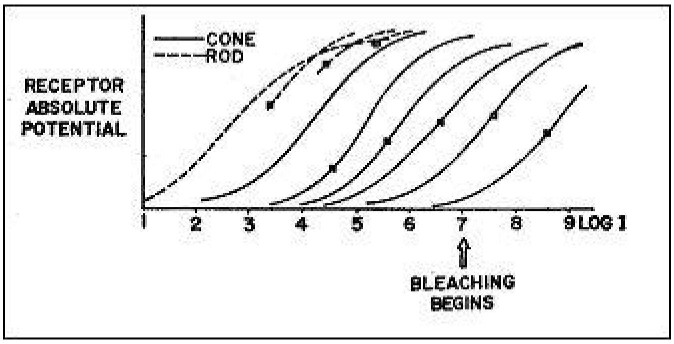


Figure 3.3: Necturus cone (and rod) response curves (response at one adaptation state to different stimuli). The adaptation state was achieved by showing a given background level - indicated by squares- to the receptors for sufficient long time to adapt. One can see the shifting of the curves. The whole dynamic range is smaller than that of humans (10 log order). (Adopted from [20])

scene, and focus on objects within this region, then our cones adapt to the luminance level at this region. With a sudden saccadic jump we may observe the less illuminated parts of this scene, and our cones adapt to the changed background conditions [20]. In our mind we have a consistent and accurate visual impression of all the parts of the scene.

If the incoming light changes suddenly the receptor gives a high response and it may saturate. At steady illumination state the receptor adapts to the changed brightness conditions and does not saturate. This implies that temporal changes are enhanced by the photoreceptors.

The second level of adaptation is the OPL, where the horizontal cells perform a spatial-temporal low-pass filtering [12, 21]. The result of the spatial low-pass filtering is the local average, which is the average luminance in a given region. Then, this is subtracted from the photoreceptor output, and the OPL acts as a spatial and temporal high-pass filter. Hence the spatial-temporal low-passed component is suppressed and the dynamic range is reduced by 1 log order.

These modifications on the visual information imply the following:

- The suppression of the spatio-temporal low-passed component eliminates the less important information. The observer is rather interested in the local differences (high-passed component) [22, 23] because these represent the objects. The suppressed low-passed component relates to the illumination. [23, 24], which caused the large dynamics.

- According to the rules of diffuse reflection the perceived light is proportional to the illumination (see (3.1)). So a proper reduction of the effects of the spatio-temporal variant illumination is a division by the illumination level (see section 3.3). The retina subtracts the horizontal cell's output from the cones, but the incoming light is already logarithmically compressed at the receptor level, so the subtraction in the logarithmic domain corresponds to a division.

Hence the retina enhances the information about the object contours, and suppresses the less important illumination components. It reacts to sudden brightness changes and later adapts to these conditions after their stabilization [20, 22].

A short note has to be made to the elimination of the effects of illumination: along shadow boundaries we have a sharp illumination change. This results in high spatial frequency components that are not eliminated. Such effects can be reduced using anisotropic diffusion. (Section 3.9)

In section 3.4.1 the similarity between the integration time and the receptor adaptation state is shown.

3.4.1 Biological and Artificial Sensors

There are several similarities between biological and artificial sensors, they both transform light to electrical stimulus and enhance certain properties of the input [39]. In this subsection, I show that using the integration time as a locally controllable parameter we can obtain retina-like image enhancement. It is obvious that the integration time is a multiplicative factor (neglecting dark current [18]). This means that integrating n times longer we get n times larger pixel value, unless the sensor saturates. Hence changing the integration time of the photo sensor, the slope of the response curve changes. This results in a shift along the log intensity axis, which is similar to the adaptation course of the cones. Integration type linear image sensor and ideal cone response curves with different integration time/adaptation stages can be seen in Figure 3.4.

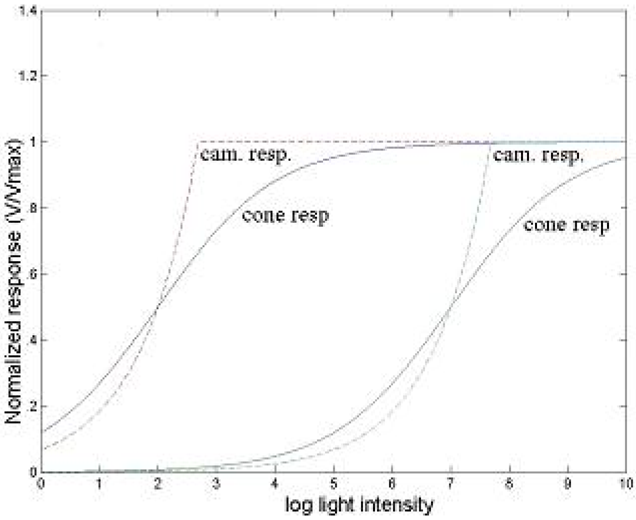


Figure 3.4: Camera responses (dashed lines) and ideal cone responses (solid lines) at different adaptation states. V_{max} denotes the maximum receptor response. Changing the integration time the camera response curve obeys similar shifting as the cone curves, thus the camera response curve can be considered as an approximation of the cone response curve.

The shifting property can be proven analytically as well:

The receptor responses can be approximated with the Michaelis-Menten law ([20] and (3.2)).

$$V = \frac{I}{I+\sigma} V_{max} \tag{3.2}$$

where V is the response measured in voltages, I is the incoming light intensity, σ is the adaptation level, i.e. the light intensity at which the response is the half of the maximal value. V_{max} is the maximal response.

The sensor response is the following:

$$V = T \cdot c \cdot I \tag{3.3}$$

where V is the sensor response in voltage, I is the incoming light intensity, T is the integration time and c is a constant multiplication factor.

For both (3.2) and (3.3) it is true that by changing the adaptation state k times, the incoming intensity I should be k times larger to have the same response

(3.4). On the log scale this means, that $log(k)$ should be added to the intensity to have the same response. This means that the curve is shifted by $log(k)$ along the log intensity axis.

$$\begin{aligned} V &= \frac{I}{I+\sigma}V_{max} = \frac{kI}{kI+k\sigma}V_{max} \\ V &= T \cdot c \cdot I = \frac{T}{k} \cdot c \cdot Ik \end{aligned} \tag{3.4}$$

Both the integration time parameter and the receptor adaptation state are multiplication factors. This is expressed by the fact that by changing both has the effect of shifting along the log intensity axis. The difference between the curves comes from the fact that the receptor response is a sigmoid type curve along the logarithmic axis, while the sensor response is a linear along the linear axis and an exponential along the log axis (Figure 3.4).

As mentioned the integration time parameter is a multiplicative factor. Hence by properly choosing its value we can eliminate (or suppress) the multiplicative effect of the illumination, as in the case of the retina. To have a good image enhancement we need a proper algorithm, which adjusts the integration time based on local image statistics. I discuss this question in section 3.7. Before that, in Section 3.5 the advantages of CNN-UM implementation and in Section 3.6 the experimental framework are shown.

3.5 Application of CNN-UM

For implementing the retinal light adaptation based algorithm, the CNN-UM architecture is chosen. There are several advantages of its usage:

1. Retina-like structure. The CNN-UM has an architecture that has similarities to the retinal networks. Neighboring cells are interconnected, this can mimic the operation of gap-junctions (cones, horizontal cells). It can also perform spatial low-pass filtering (heat diffusion), which corresponds to the low-pass filtering due to horizontal cell dendritic tree, and to the averaging effect cause by saccadic eye movements, which scan a region on the scene.

Another reason for using the CNN-UM in retina-like algorithm is its ability to model retinal functions [56, 14, 69]. Due to these works, there is a great experience in modelling and the tools for modelling the retina are also available (RefineC simulator).

2. Hardware developments. Recent hardware developments will result in locally adaptive CNN-UM chips: CACE2-k,Xenon and Q-EYE (EYE-RIS) chips [70]. There are older chips available, which enable us the verification of the portability on CNN architecture, even if they are not equipped with local adjustable sensor array (ACE-16k, [27]).

3. Software development kits. Beside the CNN retina simulator there are other simulation environments (Aladdin SDK), which enable us the testing of our algorithms on PC. The Aladdin software packet ([50]) enables us the easy transportation of our code from PC environment to CNN-UM chips.

These hardware and software environments facilitate the implementation of the algorithm.

3.6 Simulation Environment and Hardware Platforms

In this section I describe the experimental setup and the hardware platform I used for the implementation.

The algorithm I propose requires a CNN array equipped with locally-adjustable sensor array. This is so far not available, so my work had the following two phases regarding the implementation level:

1. PC simulation

2. Partial CNN implementation

In PC simulation all the calculations are implemented on a PC with a CNN simulator and Matlab.

In partial CNN implementation some operations are executed on an analog CNN array processor. These operations are those which require spatial processing. The used hardware is the ACE-BOX, which consists of a 128x128 cell CNN array, the ACE-16k version 2. chip [27] and a DSP. The full implementation of the algorithm is not possible on the ACE-16k architecture, because it cannot capture the scene adaptively. Hence we have to mimic the capturing of a scene with locally varying integration time.

3.6.1 Simulating Locally Adaptive Image Capture

Since we do not have a locally-adjustable sensor in our hand, mimicking is essential to be able to test and evaluate my algorithms.

I captured a sequence of grayscale pictures of the same high dynamic scene with different global integration time parameters $(T_1, T_2, T_3, ...T_n)$. Then, I interpolated at each pixel the value of the image. Given an integration time T at a pixel, we wish to derive the intensity V that the mimicked sensor would have perceived. We choose the two pictures with the nearest integration time values T_i and T_{i+1} in a way that $T_i > T > T_{i+1}$. V_i and V_{i+1} are the two intensity values measured in the same pixel position with T_i and T_{i+1} integration time respectively. There is a linear relationship between the perceived intensity and the integration time. Hence target value V can be computed as the weighted average of V_i and V_{i+1} (3.5).

$$V = \frac{(T - T_{i+1})V_i + (T_i - T)V_{i+1}}{T_i - T_{i+1}} \tag{3.5}$$

Using (3.5) we can synthesize the captured image based on the picture series of different integration times. This assembling technique is similar to the multiple capturing method in [18]. For capturing the image sequence, I used a Dalsa 1M28 camera [28]. The images were 8bit-depth, and the integration time was changed between 0.1 and 230 ms.

In order to make a better evaluation of the method it has to be tested on moving scenes. A new problem arises: how can we make snapshots of a given moment with different integration times, so that the scene does not change during the several captures?

I resolved this problem by making an artificial quasi-dynamic movement. This can be viewed as if I had frozen the motion at some discrete moments to create the picture series (of different integration time). Then the scene was altered slightly and a new series of snapshots was taken. As a result I had a series of snapshots at each discrete time steps. Picture series were taken at every 166ms which results in a 6fps video after assembling at each time step the result from the series.

3.7 Integration Time Adjustment Algorithm

This section deals with the local integration time adjustment. As mentioned earlier the aim of the algorithm is to reduce the high dynamic range scene and eliminate the illumination differences by suppressing the spatio-temporal low-pass component. The basis of the algorithm is the processing of the retina.

In section 3.4 we learned how the response curve of the cones is shifted to the perceived light intensity (Figure 3.3), and the spatio-temporal low-pass (DC) component is suppressed in the OPL. Both of these can be achieved by adjusting the integration time so, that the local average (spatial low-pass component) becomes the half of the maximum value - mean gray value - ($V_{max}/2$). This is advantageous for us, because this means that the local average is the same at all the locations, so the differences in the local average are eliminated and thus the high dynamic range scene is compressed. The local average or DC component is computed in a way that we blur the image (diffusion template [11]), which is similar to averaging the pixel values around each pixel (using Gaussian convolution). The size of the region on which the local average is computed is discussed later, when we write about the parameters of the method. The temporal low-pass component is suppressed because of the dynamic adjustment process, as detailed later. It is also important that by choosing an appropriate integration time we map the local average to $V_{max}/2$ (the half of the maximal response). In this case pixels brighter and darker than the local average can have a contrast to the local average because they are not saturated.

This means that the camera response curve should be shifted depending on the local brightness level. This eliminates the intra scene DC differences, because the DC values are driven to $V_{max}/2$ (mean gray value) throughout the image.

The maximal slope of the photoreceptor response curve at the local average is obtained if this value is projected to $V_{max}/2$ [29]. Similarly if the local average is projected to $V_{max}/2$ the camera curve is still steep around this level (see Fig.2.). Hence we can have large responses for the local differences, which enhances the spatial high-passed component.

Adjusting the integration time according to the local brightness level has another beneficial effect regarding the local differences. The human perception of contrast can be described according to Weber's law [22, 36, 40] (3.6).

$$C(I_1, I_2) = \frac{|I_1 - I_2|}{min(I_1, I_2)} \tag{3.6}$$

where I_1 and I_2 are the intensities of the two points for which the contrast is measured. It is important to note that this contrast measure is not only dependent on the absolute difference of the two intensities ($|I_1 - I_2|$), but it is a contrast normalized by the intensity level. Hence the Weber contrast (of two objects) does not change with the illumination (3.7).

$$\frac{|I_1 - I_2|}{min(I_1, I_2)} = \frac{|k * I_1 - k * I_2|}{min(k * I_1, k * I_2)} \tag{3.7}$$

where k is the factor, causing k times larger illumination and perceived value (3.1). Illustrated by an example: given four pixels with intensities: $I_1 = 10, I_2 = 11$ and $I_3 = 100, I_4 = 110$, the contrast between I_1 and I_2 is the same as between I_3 and I_4.

If the integration time is adjusted so that the local average is mapped to the half of the maximal value we have a Weber contrast measure. The absolute differences in a low-illuminated region might be less than in a highly illuminated one, but in both cases these differences are scaled by the adjustment of the integration time:

Assume we have I_1 as the local intensity for the poorly illuminated region and $I_2 = k * I_1$ ($k >> 1$) for the highly illuminated. Let us have two objects in both regions whose intensities are c times the local intensity ($I_3 = c * I_1$,$I_4 = c * I_2$ and $c \in [0, 2]$). If we adjust the integration time to drive the local averages to $V_{max}/2$ then the brighter region will have k-times smaller integration time ($T_2 = T_1/k$) and I_1, I_2, I_3, I_4 are perceived with the following values (3.8):

$$\begin{aligned}
I_1 &\overset{T_1}{\rightarrow} V_1 = 128 \\
I_2 &\overset{T_2}{\rightarrow} V_2 = 128 \\
I_3 &\overset{T_1}{\rightarrow} V_3 = V_1 * c = 128 * c \\
I_4 &\overset{T_2}{\rightarrow} V_4 = V_2 * c = 128 * c \qquad (3.8)
\end{aligned}$$

and the absolute contrasts are the same (3.9).

$$|V_1 - V_3| = |V_2 - V_4| = 128 * (1 - c) \qquad (3.9)$$

The equations (3.8) and (3.9) show that if we drive the local average to $V_{max}/2$ the absolute contrasts on differently illuminated regions are normalized with their local average, and the resulting contrast is proportional to the perceptional Weber contrast.

The next question is how to specify the integration time at the pixels so that the DC component is $V_{max}/2$.

One method for this is based on intra frame processing done during the image capturing. Here we capture the scene and compute the local average at the same time. The local average can be computed by averaging the voltage of the (photo-charged) capacitances. If the average reaches the desired level the integration can be stopped at that time instant. It assumes very special architecture, dedicated for this calculation. Furthermore it also requires that the averaging can be done fast enough so that even in case of short exposures the averages can be calculated. This method has a fast adaptation because the integration time is not pre-specified but depends on the local brightness level. On the other hand it requires a special hardware architecture beside the local adjustment of the sensors.

We used a more feasible method which is based on inter frame processing. It is an iterative, dynamic method, where the iteration is done through the captured frames. This dynamic-adjustment method is simpler from computational and hence implementation point of view, but as a trade off, the adaptation to a new illumination condition, similarly to our retina, takes time.

In Figure 3.5 we can see a flowchart of the dynamic-adjustment method. As a first step, we define an initial integration time map T_0 ($T_0 = 0.2sec$). Then, we repeat the followings in each iterative step. At the n-th iteration we capture an image with the integration time map T_n and get the result image V_n. Then, we perform diffusion operation on both V_n and T_n. Based on the diffused images we compute the integration time map T_{n+1} for the next snapshot. This is the adaptation phase where we change the variable parameter T_n in order to adapt to the scene.

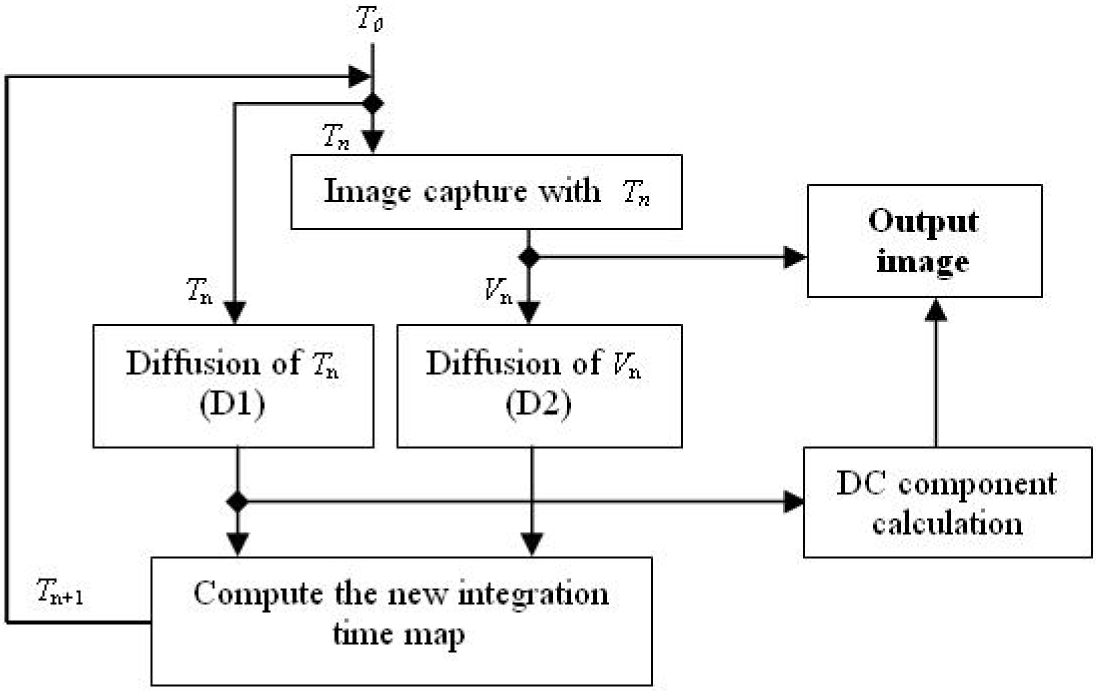


Figure 3.5: Flowchart of the dynamic-adjustment algorithm. Flowchart of the dynamic-adjustment of the integration time. The DC component can be computed based on the integration time values, and it is added to the resulting image.

The next two subsections deal with the computing of the new integration time map (integration time adjustment). We present two computations, the first is an ideal one regarding the adaptation speed, but the second one is feasible on CNN architecture. The third subsection shows results on dynamic scenes using the CNN-feasible adjustment and subsection 3.7.4 discusses hardware results.

3.7.1 Multiplicative Adjustment

The multiplicative adjustment is a way we can specify the new integration time. The adjustment of the integration time is the key of the algorithm. This is the adaptation step, where we react to the changed conditions. To describe the method some new notations need to be introduced.

At the n-th capture we have the integration time setting $(T_n(i,j))$ for pixel (i,j). With these settings the measured value $V_n(i,j)$ is obtained. The goal is to adjust the integration time at this pixel so, that the local average of V_{n+1} is driven to $V_{max}/2$. The local average of V_n is computed with diffusion (see Figure 3.5). Denote the local average for pixel (i,j) as $\overline{V_n}(i,j)$. Denote the local average of the integration time for pixel (i,j) as $\overline{T_n}(i,j)$. Change the integration time at pixel (i,j) as follows (later referenced as multiplicative formula, (3.10)):

$$T_{n+1}(i,j) = \overline{T_n}(i,j)\frac{V_{max}/2}{\overline{V_n}(i,j)} \tag{3.10}$$

As mentioned, the received intensity $V_{i,j}$ is proportional to the integration time $T_{i,j}$ (3.1). I used the heuristic that if we obtained $\overline{V_n}(i,j)$ with $T_n(i,j)$, then we have to multiply $T_n(i,j)$ with $\frac{V_{max}/2}{\overline{V_n}(i,j)}$ to get $V_{max}/2$. The heuristic assumes that the neighboring pixels are altered in the same way. This may hold for most cases, because both $\overline{V_n}(i,j)$ and $\overline{T_n}(i,j)$ are low passed components. Instead of $T_n(i,j)$ I used $\overline{T_n}(i,j)$ (local average) in (3.10), to avoid sharp changes in the integration time map. The method requires an initial integration time map. I used a uniform map with a value from the middle ranges of the possible integration time settings (64 ms.) The method calculates the proper integration time during the cycles.

The free parameter of the method is the extent of the diffusion for computing the local average. The same extent of diffusion was applied for both V_n and T_n. In case of PC simulations I computed the local average with a convolution by a Gaussian kernel. The equation of the Gaussian is in (3.11), where (i,j) is the distance from the center element, the element for which we compute the average. The σ parameter (see (3.11)) of this kernel was around 5% of the image size, maximally around 30 pixels.

$$F(i,j) = \frac{1}{2\pi}e^{-\frac{i^2+j^2}{\sigma^2}} \tag{3.11}$$

In case of the ACE-16k (see section 3.6) implementation I used 15τ (τ being the time constant of the CNN cell) as the execution time of the diffusion or 10ms on the resistive grid. The image size was 128x128 pixels, the size of the chip. For the image on Figure 3.6 a diffusion time of 250τ would be needed for the same result.

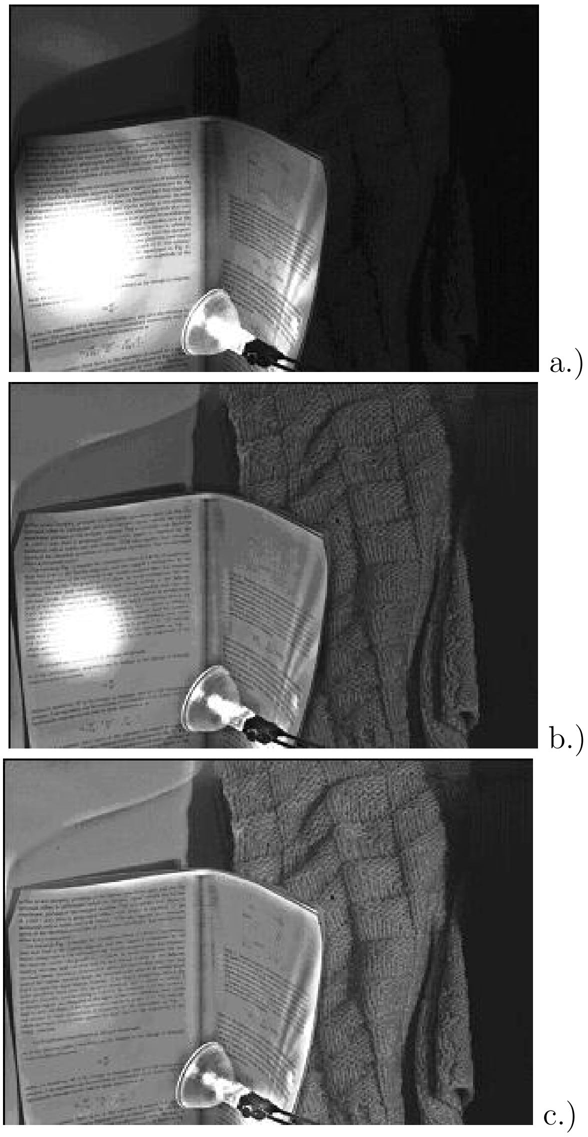


Figure 3.6: Results of the multiplicative adjustment.a.) original image with uniform settings. b.) result after the 1st adaptation cycle. c.) result after the 4th cycle.

Figure 3.6 shows some results of this method. After the first alteration of the integration time map, the dark areas are enhanced, and the saturated white zone is reduced. According to the heuristic of (3.11) all areas of the image should be in average gray. This does not hold because we had bright-saturated areas where the correction of (3.10) does not diminish the integration time enough. This is due to the fact that after reaching the saturation level of the camera (see Figure 3.4), the actual input intensity could lie anywhere above the saturation intensity. Hence the integration time is diminished according to the lowest saturation intensity.

After sufficient iterations, (Figure 3.6 c.)) an adapted state is reached where the local average is $V_{max}/2$ nearly everywhere. The right side of the image was not enhanced more for practical reasons because the integration time of the camera could not be set higher. Apart from this we can observe that the strongly illuminated areas and the dark background are driven to the same brightness level.

The multiplicative adjustment provides us a fast adaptation. On the other hand, we have to take into account some implementation aspects as well. The division in (3.10) may not be easily implemented on a CMOS CNN-UM chip, hence the additive adjustment is proposed.

3.7.2 Additive Adjustment

Additive correction specifies the new integration time based on the same data as the multiplicative adjustment. The difference is in the calculations done. Using the notations of the previous subsection (3.7.1), we compute (3.12) instead of (3.10):

$$T_{n+1}(i,j) = \overline{T_n}(i,j) + \left(\frac{1}{2} - \frac{\overline{V_n}(i,j)}{V_{max}}\right)\overline{T_n}(i,j) \tag{3.12}$$

V_n considered as $V_n \in [0, V_{max}]$. Under usual notion in CNN literature, however, $V \in [1, -1]$ so (3.12) can be reformulated as (3.13):

$$T_{n+1}(i,j) = \overline{T_n}(i,j) + \frac{\overline{V_n}(i,j)}{2}\overline{T_n}(i,j) = \overline{T_n}(i,j)\left(1 + \frac{\overline{V_n}(i,j)}{2}\right) \tag{3.13}$$

The UMF diagram of the algorithm using (3.13) can be seen on Figure 3.7, which is a detailed version of 3.5.

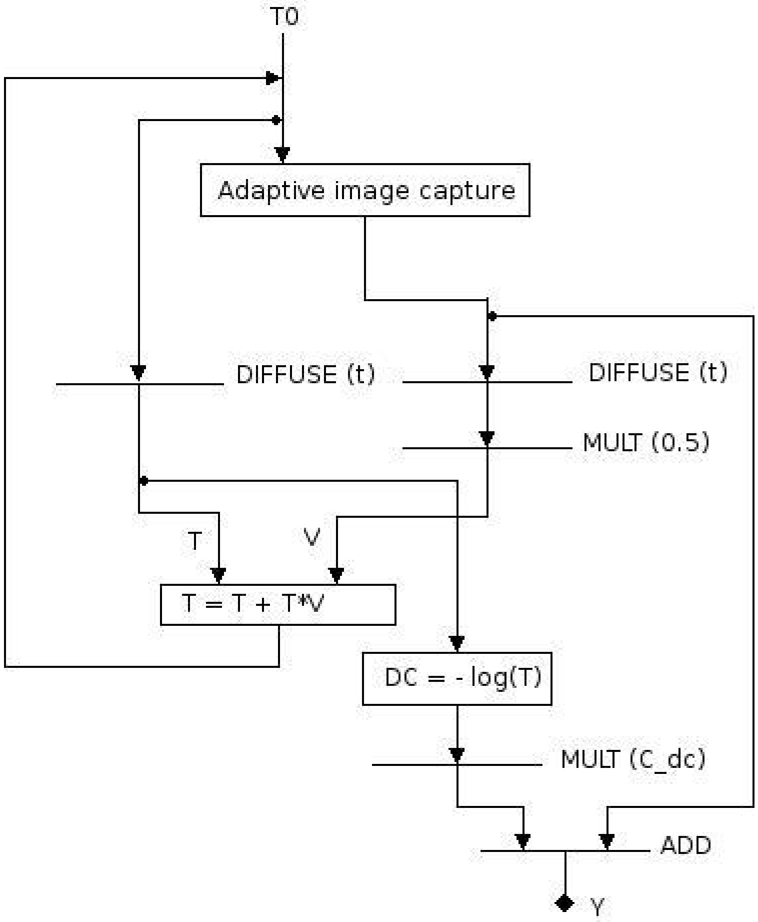


Figure 3.7: The UMF diagram of the additive adjustment algorithm. This is a detailed version of Figure 3.5, where the CNN templates can be seen as well. The formula computing the integration time of the next capture ($T = ...$) and computation of the DC component are also shown.

(3.12) and (3.13) are the same mapping, because both map T_n to $\overline{T_n}/2$ if $\overline{V_n}$ is saturated white. T_n is mapped to $1.5\overline{T_n}/2$ if $\overline{V_n}$ is black. Furthermore both (3.12) and (3.13) are linear expressions of $\overline{V_n}$. Expression (3.12) can be computed without division.

In the followings the notion $V_n \in [0, V_{max}]$ is used, because of its straightforward connection to image capture (3.3).

To understand the adjustment better we may compute the following value: depending on $\overline{V_n}(i,j)$ we multiply $\overline{T_n}(i,j)$ with a factor in either (3.10) or (3.12).

Denote this factor $M_n(i,j)$. Plotting this multiplication factor, $M_n(i,j)$ versus $\overline{V_n}(i,j)$, we can see how the integration time is enhanced/attenuated at different brightness levels. Figure 3.8 shows this integration time multiplication factor $(M_n(i,j))$ versus the local brightness level $(\overline{V_n}(i,j))$.

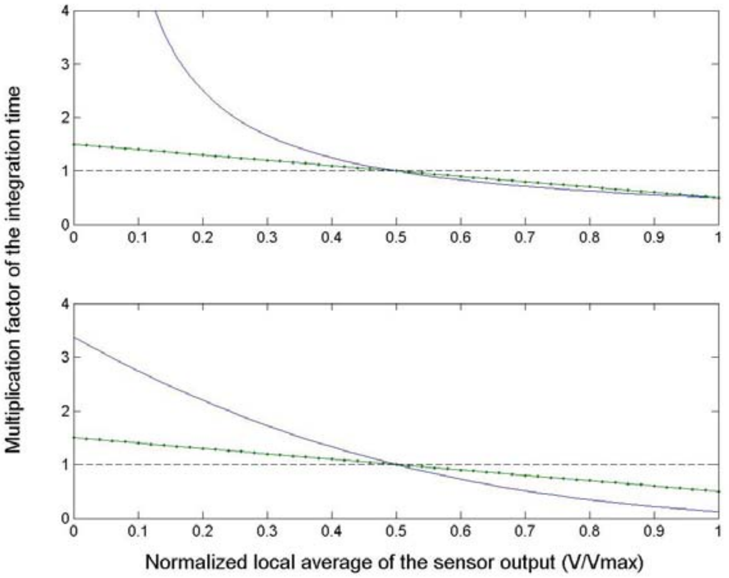


Figure 3.8: Multiplication factors of the adjustment algorithms as the function of the normalized local average of the sensor output. The factors with which we multiply the integration time at different local average values. The upper plot shows the multiplicative (solid curve) and the additive (dash-dotted line) correction. The second plot shows the curve of the fast additive correction. In this case, we took the third power of the additive factor (solid curve). For comparison, the original additive curve is shown with a dash-dotted curve. On both subplots the dim dashed line is the 1 level, which means that the integration time is kept unchanged.

The additive adjustment changes the integration time qualitatively in the same direction as the multiplicative: regions with $V_{max}/2$ average intensity are unchanged, for brighter regions ($\overline{V_n}(i,j) > V_{max}/2$) the integration time is reduced, for darker it is enhanced. This is straightforward from (3.10) and (3.12), it can also be seen on Figure 3.8.

Between the two methods there is a difference in the amount of correction. Based on the first plot of Figure 3.8 we can make a comparison between the additive and multiplicative enhancement. It is important to note that the local average at value $V_{max}/2$ is unchanged (multiplication by 1). So this is an equilibrium point of the system. In case of the multiplicative adjustment of the integration time the dark and bright regions are changed in a way that the local average would become an average gray value ($V_{max}/2$) for the next capture. This is only influenced by saturation and possible unequal change of integration time within a close (less than σ) neighborhood of a pixel.

For the additive correction the alteration of the integration time is less than this ideal amount of correction. It does not always drive non-saturated areas in the next capture to an average gray level ($V_{max}/2$). The amount of diminution of bright values is quite similar for the two methods (see Figure 3.9). There is a great difference in the enhancement of the dark values (see Figure 3.9). These are far less enhanced using the additive correction (see Figure 3.9 and 3.8). On Figure 3.8 we can see that in case $\overline{V_n}(i,j) < V_{max}/2$, the curve of the additive correction has much smaller values than that of the multiplicative correction.

In order to accelerate the adaptation of dark regions, an iterated-additive adjustment was used, further referred as fast additive correction. We can alter the integration time map k times based on the same local average values $\overline{V_n}(i,j)$ before the next capture. This is the same as if we multiplied $\overline{T_n}(i,j)$ with $(M_n(i,j))^k$ instead of $M_n(i,j)$. For $k = 3$ the resulting multiplication factor is plotted in Figure 3.8, on the second subplot with a solid line. We can see that the enhancement of the dark values became better. On the other hand, the integration times for the bright values are too much reduced. This has an oscillatory effect, because the integration time might be reduced too much for bright regions, and it would become dark, which would be enhanced to a gray level. Such a phenomenon of bright-dark-gray adaptation would be quite unnatural. Hence we need a threshold operation [11], and for the pixels brighter than $V_{max}/2$ we keep the result of the original additive correction. The data-flow of the fast additive correction can be seen in Figure 3.10.

Figure 3.9 shows adjustment results of the same initial scene as Figure (3.10). As it can be seen, the fast additive adjustment adapts faster at dark areas than the additive adjustment (compare Fig 3.9 a.) and b.)). We can also see that the fast additive correction enhances the dark areas just as fast as the multiplicative

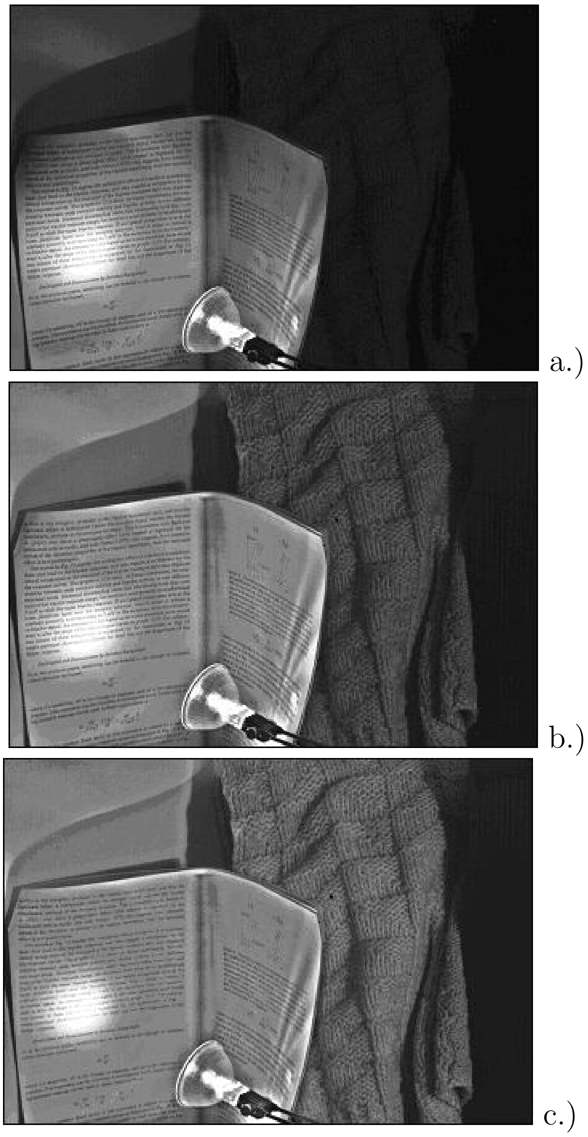


Figure 3.9: Results of different corrections after the second iteration. a.) additive, b.) fast additive adjustment and c.) multiplicative adjustment. We can see that at bright areas all methods adapt in a similar way. On the darker regions the iterative method is faster (b.) than the non-iterative additive (a.)

(Figure 3.9 c.)). The additional iterations of multiplication of the integration time with the enhancement factor do not retard the system further. These operations can be fulfilled faster on the CNN-UM architecture than the capturing of the images.

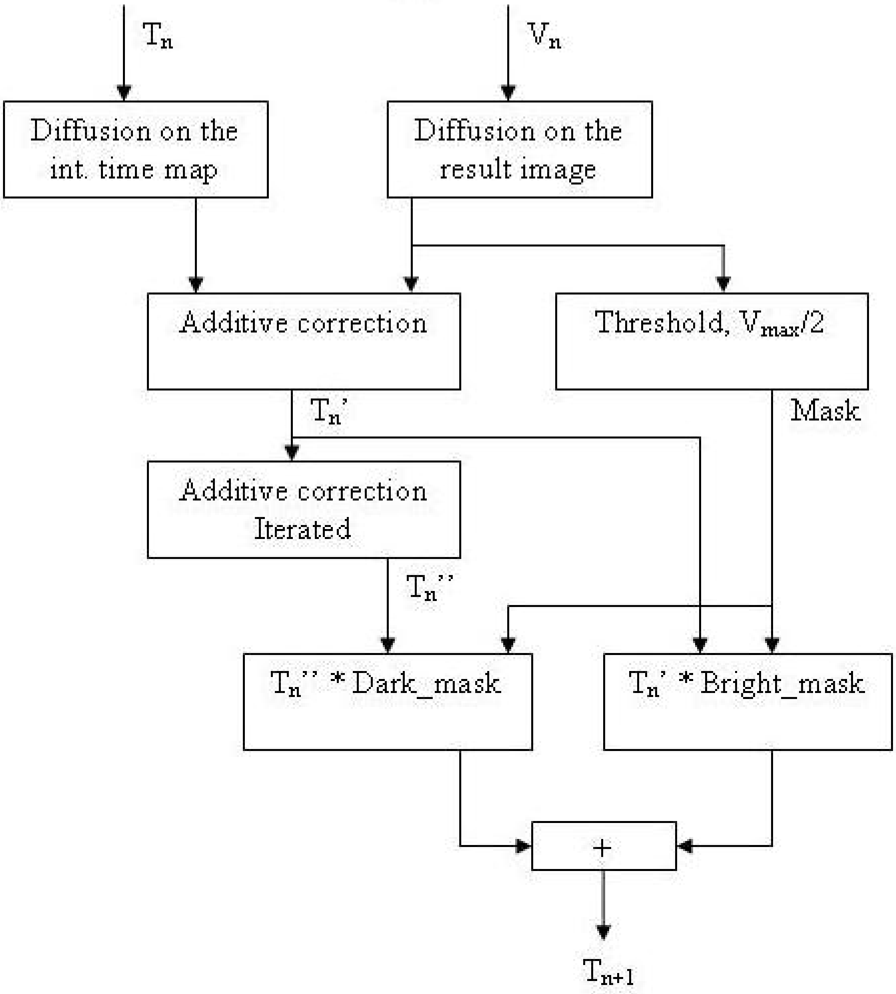


Figure 3.10: Data-flow of the fast additive correction. Detailed subpart of Figure 3.5. On the top we can see the diffusion of the obtained image and the integration time map (also present in Figure 3.5). Below this the computation of the next capture's integration time is detailed. Other steps are performed in the same way as on Figure 3.5.

The fast additive correction is feasible on CNN architecture, and beside this it has a fast adaptation as well.

The execution time of the algorithm for different images can be seen in Table 3.1.

The duration of the longest integration is independent from my algorithm, it is determined by the applied sensor and the scene. With the implementation of

Size	It. Nb.	Exec. Time [ms]	Time One It. [ms]	Max. int. time [ms]
128 x 128	15	906.3	**60.4**	1200
855 x 591	15	67280	**4485**	1200

Table 3.1: Execution time of the adaptation algorithm. A filter radius (σ) was 20. A 2.8 GHz PC with 1GB RAM was used. "Size" - size of the images. "It. Nb." - the number of executed adaptation steps. "Exec. Time" - the duration of the adaptation. Here the image capture (assembling from the series) is not considered, only the executed operations. "Time. One. It." the execution time of one iteration. The capture is not considered either. "Max. Int. Time" - Longest integration time. The duration of one capturing and processing is the sum of "Max. Int. Time" and "Time. One. It.".

the algorithm on CNN architecture the execution time might be reduced for 1/25 sec for 128 x 128 images, which enables a 25 fps flow (see Section 3.7.4).

In the following sections, this fast additive method will be used for computing the alteration of the integration time. This method has a second tunable parameter: the iteration number. However a constant iteration number $k = 3$ produces a good approximation of the multiplicative adjustment's integration time multiplication factor (Figure 3.8) and can be used independent from the scene as a fixed parameter.

3.7.3 Dynamic Results

I have applied this method to a dynamic image sequence. In this case, the input scene, captured in each iteration step, changes in time. Hence the adaptation mechanism has to follow this permanent changing. The dynamic experiments were simulated as described in section 3.6.1. According to my experiments, if the changes are not very fast on the scene, the adaptation speed is enough to produce good quality images. In case of sudden changes the method works as the human visual system: we can see the altered luminance for 1-2 seconds and then it adapts to the changed conditions. Illustrating this I present an experiment in which first the scene was unchanged, only the strong illuminant was switched on, then the position of the strong illuminant was changed.

The assembling was based on an integration time map computed with the fast additive adjustment. Some frames of the dynamic adjustment can be seen in

Figure 3.11. The parameter settings were the followings:

- The extent of diffusion was set as written earlier, to 5% of the image size.
- The iteration number of the fast additive correction: $k = 3$.

The complete image flow can be seen in [30].

To have a good comparison of the adaptation result a logarithmic-compressed image from the image series was computed from the image series at a given time instant. From the series I assembled a high-dynamic range(16 bit) image based on [31], and logarithmically compressed it.

Regarding Figure 3.11 we can see the dynamic behavior of our method. Having a change in the illumination the algorithm needs time for adaptation, to calculate the new integration time map. Before having adapted to the new conditions we get dark values where we removed the illumination (Figure 3.11 e.) the book is dark). Too bright result is obtained where the illumination was lower (Figure 3.11 e.) the right side of the image is too bright). This can be viewed as a memory; the system remembering its past. This means that we have a temporal high-pass filtering effect as well. This is similar to retinal processing, where the spatio-temporal low-pass component is suppressed and the high-pass components enhanced (see section 3.4).

This experiment illustrates that my method adapts to spatio-temporal luminance changes.

3.7.4 Hardware Results

In this section my results using the new version of the ACE16k chip [27] are shown. The specialty of this CNN chip is that it contains an isotropic resistive grid, which enables us to calculate diffusion in a very accurate way. Figure 3.12 shows the steps of the dynamic adjustment algorithm (Figure 3.5) which are realized on the ACE16k.

The currently available ACE16k chip does not contain a locally adjustable optical input. Though we could not implement the whole algorithm on this chip, we could implement one of the most time consuming steps: diffusion. The step

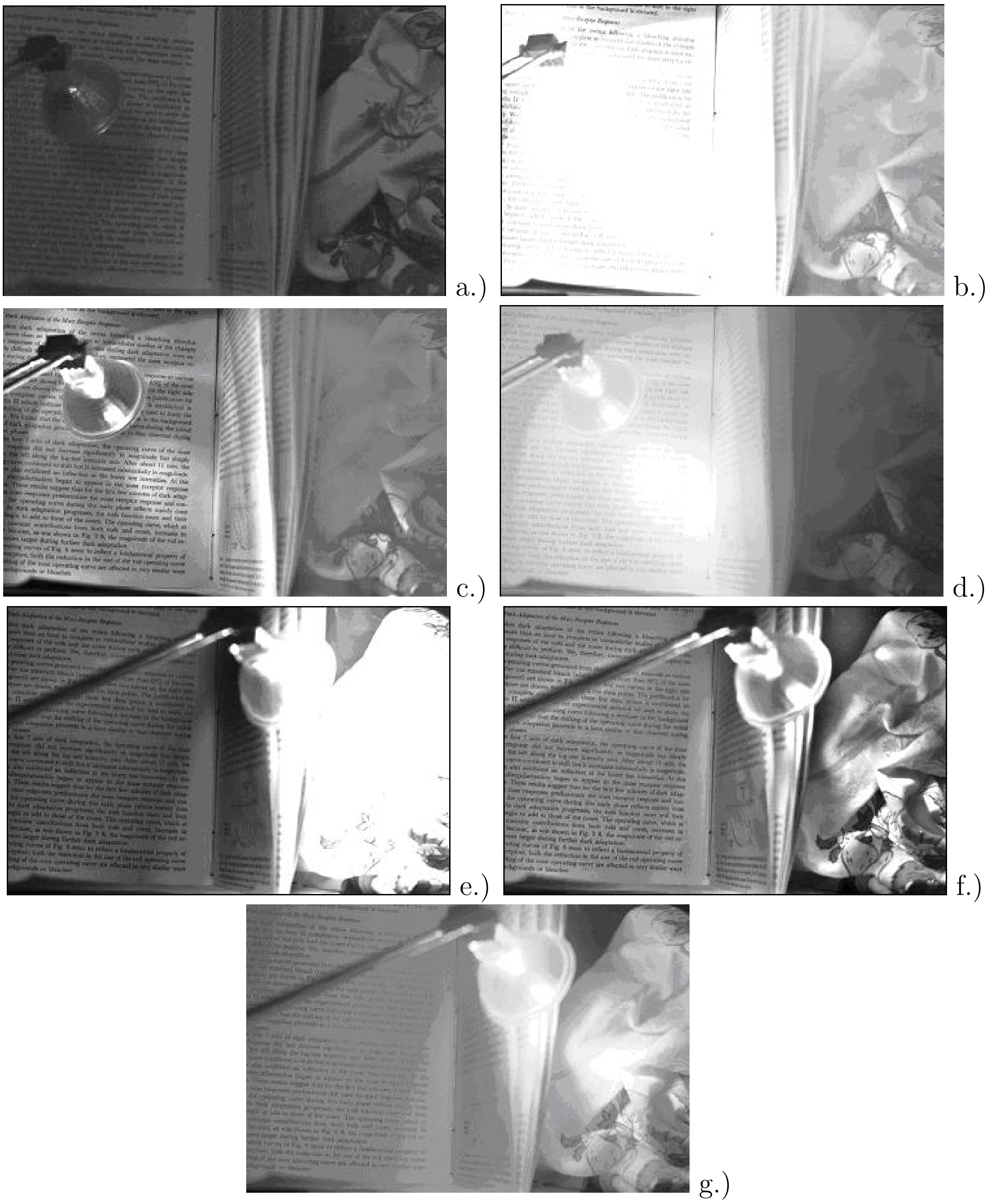


Figure 3.11: Dynamic experiments. Illumination was switched on during the first 6 frame of 60 frames. Between frames 6 and 14, illumination was unaltered, to provide time for adaptation. a.) The original, equally illuminated scene with global integration time. b.) Illumination is switched on at the 6th frame. c.) the adapted state, the 14th frame. d.) the logarithmic compression of this scene. The illuminant was moved between frame 14 and 22. e.) shows the 22nd frame. In this frame, the illuminant shifted to the right significantly. After this, I did not alter the scene until the 31st frame to provide time for adaptation (f). g.) is the logarithmic compression of this scene.

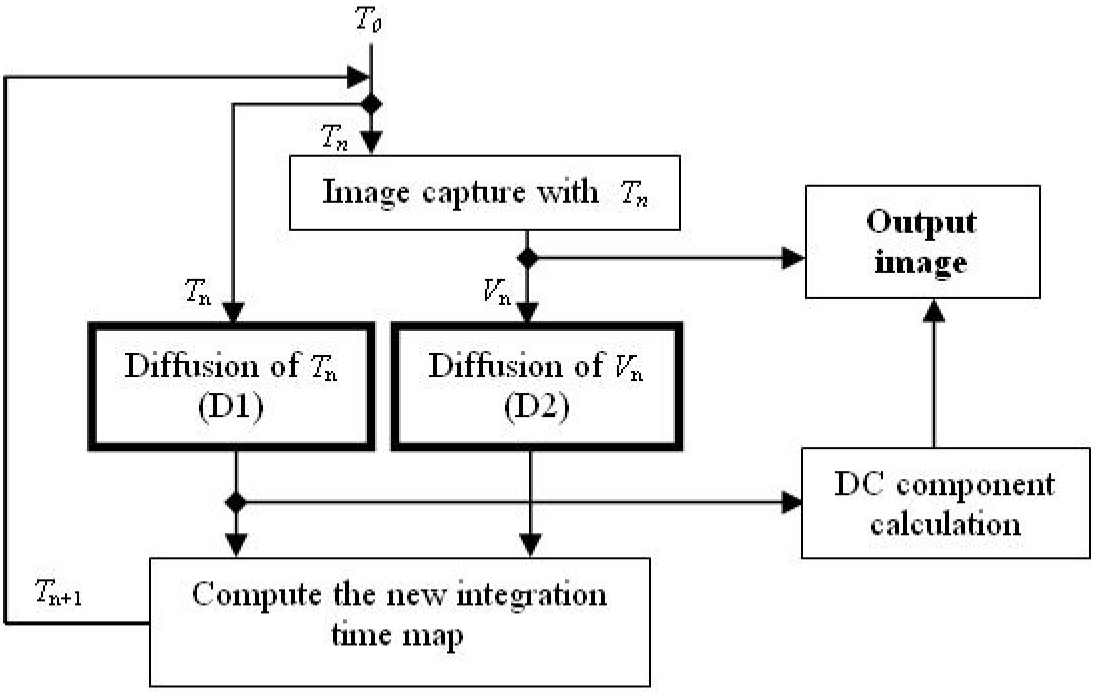


Figure 3.12: Flowchart of the adaptive perception algorithm. The box shown with thick lines are implemented on the ACE16k chip. The notations are explained in section 3.7.1.

"computation of the new integration time" was initially implemented using B-templates and currents. Later it was implemented on the accompanying PC, because there was already a need for image transfer for the simulated adaptive image capture.

It turned out that it is useful to implement the diffusion steps on the ACE16k architecture because the CNN-UM chip allows us an extremely fast and accurate diffusion of an image. Our results show, that using the ACE16k chip we can achieve satisfactory results, and based on these we can well specify the new integration time. Using the ACE16k we can calculate diffusion on 128x128 sized images only, hence we cut out a subpart of the same image that we used in Figure 3.9. The adaptation results are shown in Figure 3.13. In my example, the execution of the diffusion on a resistive grid took 5 microseconds.

The execution time of the algorithm was 24.4 ms. This time includes the calculation of the integration time map and the two diffusion operations. For the time measurements the integration time calculation was also implemented on the ACE-16k chip. On the other hand the results shown in Figure 3.13 were obtained with chip-implementation of the steps on 3.12: diffusion operations are performed on the CNN chip, the integration time is computed on the PC, because the integration time map is not calculated accurately enough.

This execution time corresponds to a 41 fps flow without considering the image capture, the duration of which can vary depending on the applied sensor and the observed scene.

The execution time of the algorithm with anistropic diffusion is difficult to estimate, because it depends on the architecture where the anisotropic extension is feasible. The most probable architecture is the Q-EYE chip, where the anisotropic diffusion can be approximated by masking of some cells during diffusion. This implementation was not simulated, but the same length for the diffusion is expected, because the strongest diffusion remains the same as in the ACE-16k implementation. On the other hand this duration should be stopped several times for masking some boundaries and thereby obtaining an unequal diffusion strength. I expect that the number and the duration of these mask-changes are not large, the challenge is rather in the determination of the masked regions.

3.8 Restoring the DC Component

By DC component the low-pass component (local average) of the input scene is meant. The elimination of the scene's low-pass component was one of the goals of the method, because it enables us the perception of the high dynamic range scene. Here we deal with the restoration of the DC component in a way that it does not drive the image to saturation. This is used for the following reasons:

The methods presented so far enhance the high-pass component of the images: edges, local differences. We eliminate the effect of the differences in the local average by driving all the averages to the half of the maximum response. This is very advantageous for example in the case of the book part of the scene, where we will have equal local averages (Figure 3.6 and 3.9). On the other hand the illumination differences are totally disappeared. Furthermore the intensity

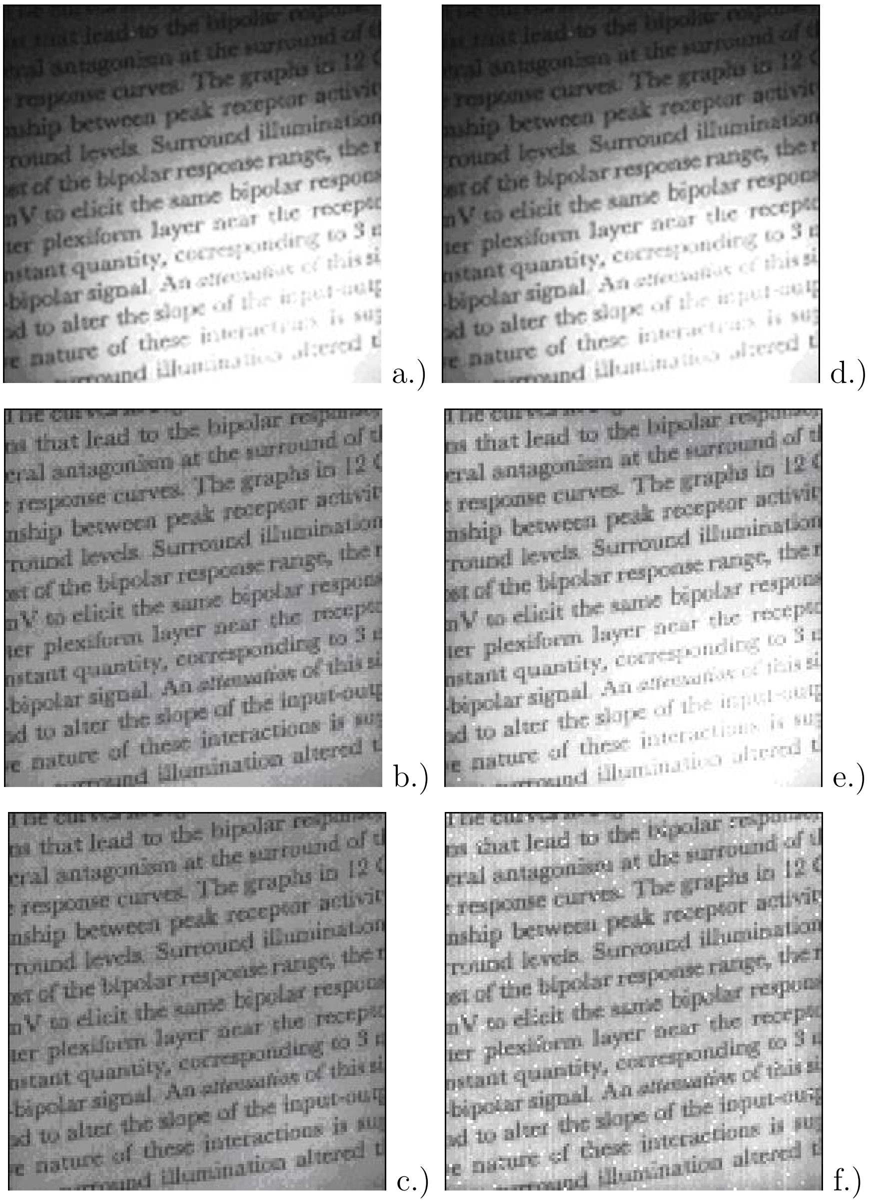


Figure 3.13: Adaptation results computed partially on the ACE16k chip. (a,b,c) show the adaptation steps using the simulator, and (d,e,f) show the adaptation steps calculated partially on the second version of the ACE16k chip.

differences between some objects also disappeared, because these were in the same order of spatial frequency band as the illumination (e.g. intensity difference between the pullover and the book). It was our goal to eliminate the illumination difference, but some observers might be interested in it.

Hence perceptually it might be better to preserve the DC component and add a downscale component of it to the resulting image. Downscaling is essential, since we do not want to reconstruct saturated areas. We can retrieve the information about the DC component from the integration time map, because it carries the information about the local averages.

During adaptation the criterion is to drive the local average of the perceived image to an average gray value ($V_{max}/2$) As a result the local average of the input scene is projected to the half of the maximum value (3.14):

$$\overline{I_n}(i,j) \cdot T_n(i,j) \cdot k \approx \frac{V_{max}}{2} = constant, \tag{3.14}$$

where $\overline{I_n}(i,j)$ is the local average of the input intensity, and k is an appropriate constant.

This means that the original input intensity multiplied by the integration time is a constant value all over the scene. So the local average of the input intensity is inversely proportional to the integration time (3.15):

$$\overline{I_n}(i,j) \propto \frac{1}{T_n(i,j)} \tag{3.15}$$

In adapted state this holds for every pixel (i,j) throughout the image. Hence we obtain a measure representing the low-pass component of the original image by taking the reciprocal of the integration time.

How can we calculate a DC component from the integration time, in a way that the image will not become saturated? In the visual system every information is logarithmically compressed so (3.16) was chosen for computing the restored DC level.

$$\overline{V_{DC}}(i,j) = -log(T_n(i,j)) \tag{3.16}$$

After that this value should be normalized between $-V_{max}/2$ and $V_{max}/2$. With these values the local average of the darkest areas will be shifted from $V_{max}/2$ to 0 and at the brightest to V_{max}. In order to avoid the saturation of the darkest and brightest areas I use a constant scaling parameter: $c_{DC} \in [0,1]$, which is a multiplicative parameter, and specifies the amount of DC component in the resulting image.

$$V_{DC,n}(i,j) = normalize\left\{-log[T_n(i,j)]\right\} * c_{DC} \tag{3.17}$$

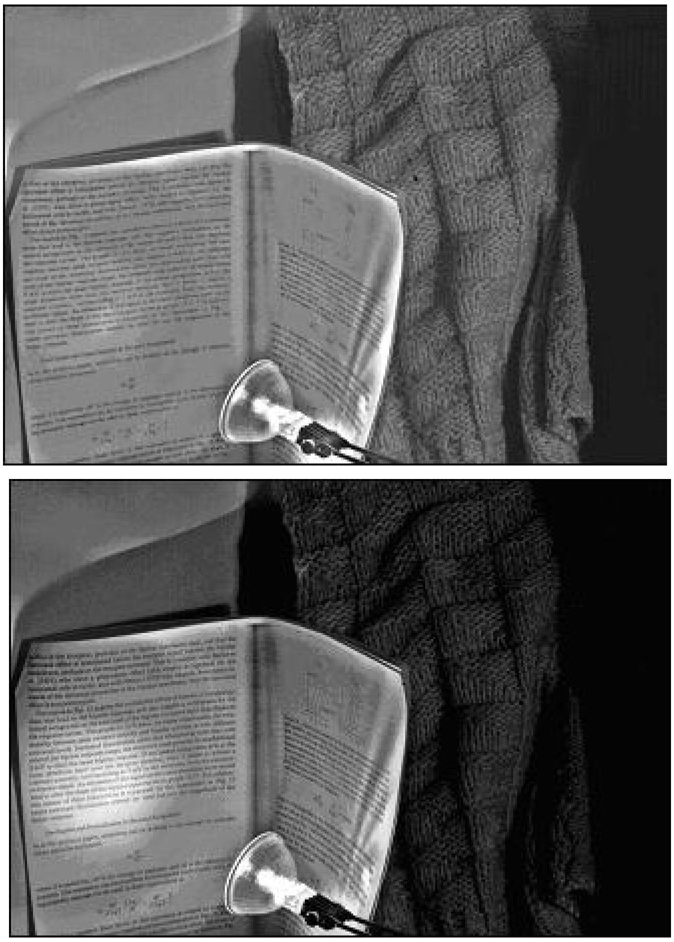


Figure 3.14: Effect of the DC component. Result images after the 8th adjustment. (a) without the DC component, (b) DC component added. c_{DC}=0.5

Figure 3.14 shows result with and without DC component. We can see that the illumination differences are visible, but none of the regions are driven to saturation.

The DC component restoration is meant to be an optional function. The user can decide, whether he is interested in eliminating all differences, or the low-pass component should be preserved in a reduced manner.

Using CNN-UM architecture the DC level can be computed by interpolating the logarithmic curve with some circuit elements.

Finally our proposed method has three free parameters:

1. neighRad: the neighborhood radius of the low-pass filtering (usually set to one tenth of the image size). Using diffusion template operation the execution time corresponds to this parameter.
2. Iteration number (k): iteration number of the additive adjustment. (it was set to 3 in the simulations)
3. c_{DC}: Ratio of the DC component. Usually 0.7 was set, when the DC component was needed.

3.9 Simulations with Anisotropic Diffusion

After adapting to the changed illumination conditions we obtain a well perceivable image. On the other hand this result can still be improved. The adaptation algorithm reduces the spatial low-pass components and the effects of the illumination are assumed to be reduced as well. However, the illumination might have sharp changes, spatial high frequency components. Figure 3.14 shows such an image. We can see a bright stripe along the edge of the paper. This is due to the fact that there is a great luminance/illumination difference between the two sides of the sheet's edge. To drive the local average to $V_{max}/2$ the integration time is set to a mean value along the edge. This integration time is longer than at other parts of the bright side while it is too short for the dark part. The bright side will have an extra white stripe, while the dark part has a black one (Figure 3.14).

This is the "halo" effect [24]. However this is similar to human vision, in that we perceive a bright object brighter if the surrounding is darker (Mach band effect [22]), it would be advantageous if we eliminated or reduced this phenomenon.

Choosing a smaller filter radius does not eliminate the problem, it only reduces the width of these white stripes. Figure 3.15 a.) and b.) show the result of the algorithm after the 5th iteration. a.) is the result in which the average was computed with a diffusion corresponding to a Gaussian kernel of $\sigma = 15$, in b.) $\sigma = 3$. We can see that the "halo" effect is reduced, but it still exists. On the other hand some low-frequency parts of the texture have disappeared or smoothed, like the beam of the lamp.

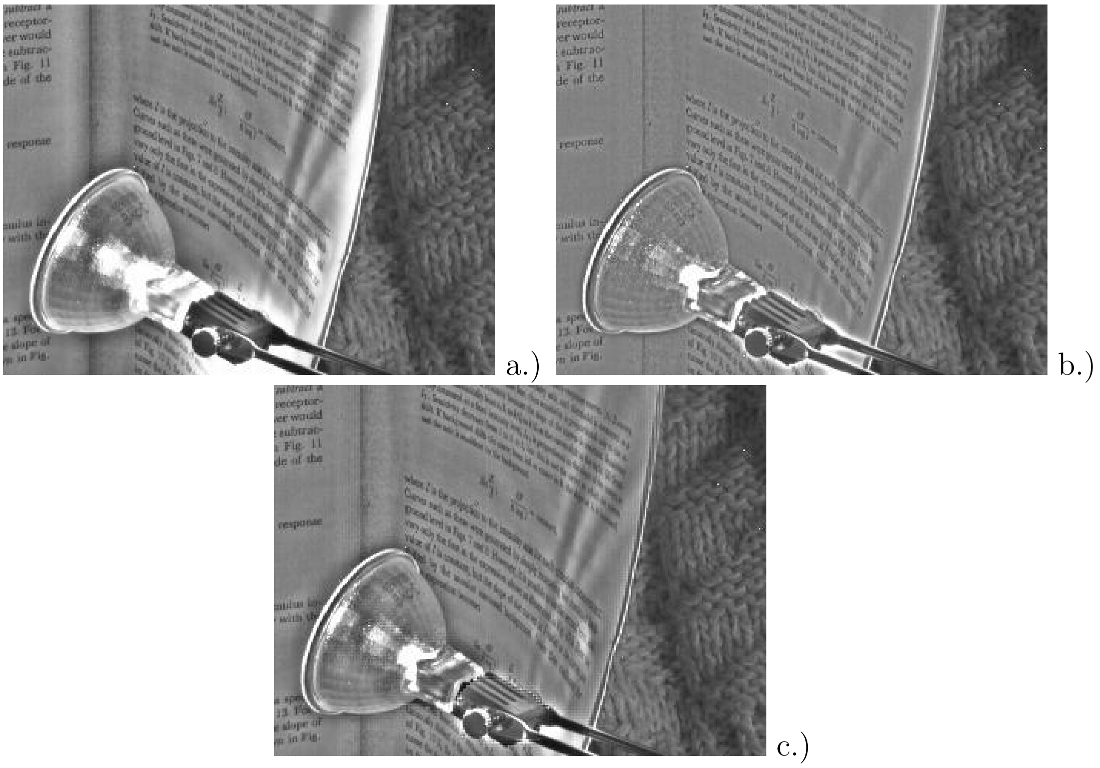


Figure 3.15: The use of anisotropy. Magnification of a subpart of the images on Fig.14. a.) shows the results using large filter radius (σ=15), we can see that homogeneous areas are well perceived due to the smoothing. The borders are still strong. Image b.): using smaller filter radius (σ=3) we obtain good results at the borders, where the different illuminated regions do not influence each other, but we have worse results in homogeneous regions. On c.) we can see the result of the digitally simulated anisotropic resistive grid. Here we obtained large filtering-like results at homogeneous areas, and small filtering-like results along the borders.

The solution for this problem is a diffusion which is widespread on homogeneous areas and restricted along sharp intensity transitions. Nonlinear diffusion models [32] are well suited for such task, in that they preserve sharp edges, and perform a smoothing in homogeneous areas. My method is based on the one published in [24], which has similar effects: it computes the local average on a resistive grid, which has a lower conductance along the edges.

The conductance term of [24] was computed as follows (3.18):

$$w_{1,2} = \frac{I}{\Delta I} = \frac{min(I_1, I_2)}{|I_1 - I_2|} \tag{3.18}$$

where $1, 2$ represent indexes of neighboring pixels and I is the intensity.

From the denominator of expression (3.18) we can see that if two pixels' intensity has a large difference then we have a low conductance. Hence there is less diffusion if neighboring pixels have different intensity. It is also important to note that a relative difference is computed (Weber contrast, section 3.7), this means that we have larger conductance if a given difference is between larger values (e.g. $I_1 = 1, I_2 = 10$ vs. $I_1 = 100, I_2 = 110$ in the latter case the proportional difference is far less). This is expressed by the nominator of (3.18).

In my method integration time is used to set a constraint on the diffusion of the perceived image. Integration time represents the illumination differences between the regions and is not influenced by the intensity differences evoked by small textures.

The well known heat-diffusion equation is modified by a contrast term which reduces the diffusion at high contrast regions (3.19):

$$\frac{\partial V}{\partial t} = \frac{1}{C_x} \frac{\partial^2 V}{(\partial x)^2} + \frac{1}{C_y} \frac{\partial^2 V}{(\partial y)^2} \tag{3.19}$$

V denotes the perceived image, t the time as before, C_x and C_y are the contrasts along the x and the y axes (3.20).

$$C_x = \frac{\left|\frac{\partial T}{\partial x}\right|}{T}, \qquad\qquad C_y = \frac{\left|\frac{\partial T}{\partial y}\right|}{T} \tag{3.20}$$

where T is the integration time.

It is important to note that this is also a proportional contrast term, and the difference is divided by the value of the integration time.

We need to discretize the diffusion equation (3.19) in space because of the structure of a CNN array: the second order spatial derivatives for pixel (i, j) become the followings (3.21):

$$\begin{aligned}\frac{\partial^2 V}{(\partial x)^2} &\rightarrow \frac{\frac{V(i+1,j)-V(i,j)}{\Delta x}-\frac{V(i,j)-V(i-1,j)}{\Delta x}}{\Delta x} = \\ &= a\cdot\{[V(i+1,j)-V(i,j)]-[V(i,j)-V(i-1,j)]\}\end{aligned} \tag{3.21}$$

where a is an appropriate constant. In the y direction the j index changes respectively. We have to divide all the terms with the contrast which can be calculated in the following way (similarly to (3.18)) (3.22)

$$\begin{aligned}C_x = \frac{\left|\frac{\partial T}{\partial x}\right|}{T} &\rightarrow \frac{\left|\frac{T(i+1,j)-T(i,j)}{\Delta x}\right|}{min[T(i+1,j),T(i,j)]} = \\ &= b\frac{|T(i+1,j)-T(i,j)|}{min[T(i+1,j),T(i,j)]}\end{aligned} \tag{3.22}$$

or

$$\rightarrow b\frac{|T(i,j)-T(i-1,j)|}{min[T(i,j),T(i-1,j)]} \tag{3.23}$$

depending on which neighbor one regards. b is an appropriate constant.

In my algorithm I used a slightly different contrast measure, which simplifies the calculations, and eliminates $C_x = 0$ cases where we would have infinite conductance (3.24):

$$\begin{aligned}C_{x,mod} =& \quad b + C_x = b * \left\{1+\frac{|T(i+1,j)-T(i,j)|}{min[T(i+1,j),T(i,j)]}\right\} &=\\ =& \quad b*\frac{min[T(i+1,j),T(i,j)]+|T(i+1,j)-T(i,j)|}{min[T(i+1,j),T(i,j)]} &=\\ =& \quad b*\frac{max[T(i+1,j),T(i,j)]}{min[T(i+1,j),T(i,j)]}\end{aligned} \tag{3.24}$$

(3.24) comes because if we add the difference to the smaller of $T(i,j), T(i+1,j)$ then we get the larger value. Expression (3.24) computes the ratio of the larger number with the smaller one instead of dividing the difference with the minimum.

The method was simulated in discrete time. The discretized equation (3.25) is derived from (3.19) with applying (3.21) :

$$\begin{aligned}V_{t+1}(i,j)-V_t(i,j) = a\cdot\frac{1}{C_{i+1,j,i,j}}&[V_t(i+1,j)-V_t(i,j)]+\\ +a\cdot\frac{1}{C_{i,j+1,i,j}}&[V_t(i,j+1)-V_t(i,j)]...\end{aligned} \tag{3.25}$$

For the sake of better visibility the formulas for the remaining two directions $((i-1,j),(i,j-1))$ are not displayed. The contrast terms are calculated according to (3.24).

The calculation of this anisotropic diffusion with Weber contrast conductances is not easily feasible on the ACE-16k platform. It can be implemented on a resistive grid, where we can specify the conductances externally5, before the diffusion. It can be approximated with a neural-network, where we can mask specific cells of the resistive grid and have less conductance (EYE-RIS system). The computation of the contrast terms (3.24) can be implemented on digital architectures, because of the division operation. It should be calculated once in a cycle before the diffusion.

The method is source free because the way pixel $(i+1,j)$ influences pixel (i,j) (3.26) in the same manner (i,j) influences the pixel $(i+1,j)$ (3.27) in the opposite direction :

$$a \cdot \frac{1}{C_{i+1,j,i,j}}[V_t(i+1,j) - V_t(i,j)] + \tag{3.26}$$

$$+a \cdot \frac{1}{C_{i,j,i+1,j}}[V_t(i,j) - V_t(i+1,j)] = \tag{3.27}$$

$$= 0 \tag{3.28}$$

The results of the simulations with 50 iterations and $a = 5$ can be seen on Figure 3.15 c.).

Comparing Figure 3.15 a.), b.) and c.) we can see that in case of the anisotropic simulation we could reduce the diffusion along the borders of differently illuminated regions similarly to b.). On the other hand we have more texture than in the case of the small diffusion (b.).

We could also see that choosing the integration time as a criterion for the diffusion (3.20) we could stop the diffusion along the shadow borders and achieve similar results to [24], who used the original intensity, which requires the handling of a wider dynamic range.

3.10 Conclusions

The designed adaptive capturing algorithm is able to reduce the dynamic range of unequally illuminated scenes. The methods perform a retina-inspired reduction of the dynamic range, so the meaningful information is preserved. The less interesting low-pass component is reduced. This can be kept in a desired and reduced manner with the restoration of the DC component without causing saturation.

Anisotropic extension of the method avoids introduction of artifacts around edges.

My method captures high-dynamic range scenes without artifacts and by reducing less important information, which properties are similar to the impression of a human observer.

These results were published in [1, 3, 4]

The main possibility of progress in this field lies in the full hardware implementation of the project. The Q-EYE chip (EYE-RIS) is expected to give us an opportunity for the implementation of spatially varying exposure and anisotropic diffusion as well. Another possibility is the extension of the method for color images. In case of color images we have to deal with chromatic distortions, and a simple scaling of the luminance channel does not always yield sufficient results.

The next chapter shows a comparison of my algorithm to other high-dynamic range perception methods and sensors. The dynamic range of my method is also examined.

Chapter 4

Comparison of the Adaptive Image Sensing Algorithm with Alternative Methods

In this chapter the adaptive capturing method is compared to several high dynamic range sensors in sections 4.2 and 4.3. We also measure the intensity range that the method can perceive in section 4.4.

This chapter is organized as follows: section 4.1 gives a brief introduction. Section 4.2 shows a comparison to logarithmic sensors, section 4.3 to locally adaptive sensors and high dynamic range rendering methods. Section 4.4 discusses the method's dynamic range.

4.1 Introduction

In [18] high dynamic range sensing methods are compared. My method corresponds to the locally adjustable integrating methods (Multiple captures, Spatially varying exposure). These methods do not have locally adjustable sensors. They produce spatially variant exposure by taking several pictures with different exposure settings (Multiple captures), or they have several sensors with different exposure settings at a position. These methods yielded better signal-noise ratio (SNR) than the other non-integrating sensors. Non-integrating sensors measure the current evoked by the photodiodes, and they are not integrating it over a time to filter out noise (some logarithmic sensors). The Multiple Capturing method takes images of different integration time (similar way as in my experiments). The Spatially Varying Exposure method requires more sensors with different capturing parameter for one pixel. Compared to my method the first requires more time, the latter looses spatial resolution.

4.2 Comparison to Logarithmic Sensors

Compared to logarithmic sensors, based on [18] we can state that the integrating sensors have better SNR than the logarithmic sensors. Furthermore we can observe the contrast relationship of the two methods.

We may compute the sensitivity defined as the slope of the response curves [18]. This value is proportional to the response to a certain input difference. Assume an ideal logarithmic sensor, which spans the log-dynamic range linearly (see Figure 4.1). The adaptive method shifts the response curve to the local dynamic range by adjusting the integration time. Thus adapted to the local dynamic range it has a greater slope than the slope of the logarithmic sensors response curve.

This means that the same intensity difference around the local average generates a larger difference in the response, and yields a greater contrast using local adaptation (compare Figure 3.11 f. and g.).

The fact that the slope of the adaptive integration time sensor's response curve is larger than that of the logarithmic sensor can be deduced formally as well:

The response of the ideal logarithmic sensor versus the log-intensity (4.1):

$$V_{log}\left[log(I)\right] = \frac{V_{max}}{log(D)} log(I) \tag{4.1}$$

where D is the dynamic range spanned by the logarithmic sensor and $log(.)$ means the natural logarithm. The slope of the curve, sensitivity as used in [18] (4.2):

$$S_{log} = \frac{\partial V_{log}(I)}{\partial I}|_{I=I_0} = \frac{\partial V_{log}(I)}{\partial (log(I))} \frac{\partial log(I)}{\partial I}|_{I=I_0} = \frac{V_{max}}{log(D) I_0} \tag{4.2}$$

I compute the sensitivity at $I = I_0$, where I_0 is the local average, around which the sensitivity is compared to the adaptive integration time characteristics.

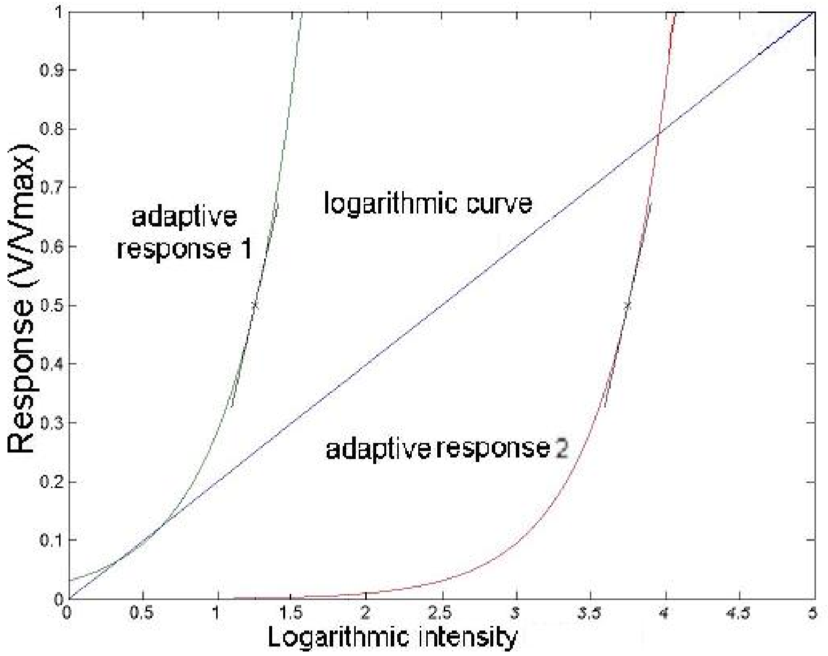


Figure 4.1: Comparison of the logarithmic and the variable integration time sensors. The ideal logarithmic sensor spans the whole intensity range and saturates at the end (diagonal line). The exponential curves show adaptive integration time responses at different local averages (indicated by x). At the local averages the slopes of the tangential are indicated.

Integrating sensors have linear response. Their response at $I = I_0$ is $V_{max}/2$ so it can be computed as follows (neglecting dark current) (4.3):

$$V_{lin}(I) = \frac{I \cdot V_{max}}{2 \cdot I_0} \tag{4.3}$$

The sensitivity computed as before (4.4):

$$S_{lin} = \frac{\partial V_{lin}(I)}{\partial I}|_{I=I_0} = \frac{V_{max}}{2 \cdot I_0} \tag{4.4}$$

Let us compute the ratio between the two sensitivities (4.5):

$$r_{i,l} = \frac{S_{lin}}{S_{log}} = \frac{log(D)}{2} = \frac{log(2) log_2(D)}{2} \tag{4.5}$$

This ratio specifies how much steeper are the adaptive integration time curves than an ideal logarithmic curve (see Figure 4.1). E.g. in case of a 16-bit scene

($D = 16$): $r_{i,l} = 5,54$, which means that using adaptive integration time method the same intensity difference evokes a 5,54 times greater response. So theoretically a logarithmic sensor needs 3-bit deeper accuracy to code the same information. Real logarithmic sensors do not certainly span their response range within the input range. Due to the fact that they would span it on a wider range (which makes the slope flatter) the LinLog characteristic is applied [28]. The initial part of such a response curve is linear and after a specified value it is logarithmic.

4.3 Comparison to High-Dynamic Range Sensors

In this section we deal with pixel level adaptive algorithms and implemented sensors which are capable of perceiving high dynamic range scenes. Some of these also use spatially varying exposure parameters.

Aizawa, [13] presents a hardware implementation using integrating type sensors, that charge capacitances. The capacitance of a pixel is reset if the pixel reaches a close to saturation region. The output of the method is the time required to reach this region, and the voltage of the capacitance within this region. The sensor's output needs post processing: the output voltage of the pixels where the integration was stopped is needed to be normalized with the integration time.

Boahen, [33] uses integration of the photocurrent, until the pixel intensity reaches a given threshold level. After reaching the level it sends an impulse (spike), resets the sensor and continues integrating. Thus the frequency of the spikes is proportional to the intensity on the scene. The method intends to mimic the spiking of the ganglion cells, but includes no processes corresponding to the retinal light adaptation. The whole dynamic range is coded in the duration of the intervals between subsequent spikes, so the sensor requires high temporal sensitivity.

In my method, the integration time of a pixel corresponds to the adaptation state of the cones. These two have a shifting effect along the log-intensity axis. I do not preserve the global intensity relationships of the high dynamic range scene as [13, 33], but perform an adaptation and a high-dynamic range compression already at the sensor level. In my method the intensity differences between the differently illuminated regions are eliminated, due to this fact the dynamic range is reduced. On the other hand the intensity differences to the neighboring pixels - the local differences - are maintained, and thus the texture information is not lost. Only the intensity of the spatial low-pass components is reduced, the object boundaries are enhanced, because they are mostly represented by higher frequency domains.

Of course this assumption is a limit of my method: it can be applied when the illumination varies over large scales, and the meaningful information is in the higher spatial frequency components.

On the other hand apart from saturated areas the original intensity can be retrieved by dividing the perceived intensity with the integration time. Hence the original high dynamic-range brightness of the scene can be computed if other dynamic-range compression method is desired.

The method presented by **Grossberg and Nayar**, [19] is a variant of the Multiple Captures method ([18]), it uses several pictures captured with different global exposure parameters (e.g. integration time). It builds the result image by summarizing the pictures. The key points are the values of the exposure parameters. These are set to maximize the slope of the resulting response curve compared to a desired curve (e.g. linear curve, see [19]). My method aims to maximize the slope of the response curve location dependently, which is achieved by adjusting the integration time. Only one image is taken of a scene to retrieve the result, so we do not loose time, for capturing several images. Furthermore in [19] and in [38] the Multiple Captures method requires a read-out of a CCD camera and assembling the image off-chip. The readout is a destructive process and the higher integration time images need to be re-captured again, so the total integration time is the sum of the individual times. In the architecture of this dissertation the chip would retrieve a perceivable output without the need of further post-processing or image fusion. Of course these ([19, 38]) solutions are cheaper, and does not require specific hardware, if the duration of the capture is not critical they can be well applied. My method has an advantage in case of dynamic scenes.

Another variant of the multiple captures method presented by **Gammal**, [39, 31, 41] uses a CMOS hardware with non-destructive readout. In these cases the time needed for the integration is the longest time as in my algorithm. The architecture integrates the pixels and reads them out at different time intervals. Because of the non-destructive nature the integration can continue with the partially charged capacitances. When a given pixel (i, j) reaches saturation at the k-th capture, the pixel value (V) of the previous, non-saturated capture is taken (V_{k-1}) together with its integration time length (T_{k-1}). Thus for a given pixel we have two values, the integration time and the captured value, and the ratio of the two specifies the high-dynamic range pixel value (V/T). Hence this HDR image needs post-processing again, to render to lower dynamic range. The pixel value V devoid of the integration time T is not sufficient, because it does not surely preserve the local relationship of the pixels. This is caused by the fact that neighboring pixels integration time does not influence each other (unlike in biology there is no local processing, [39] lacks this property). Hence a neighboring pixel $(i + 1, j)$ might have a lower value V than (i, j), even if $(i + 1, j)$ was lighter. This is because it could have been sampled at an earlier capture ($k - 2$ or before). The advantage of [39, 31, 41] is that it captures fast, and retrieves a saturation free image. However it is also a HDR perception method without rendering this range to lower intensity. My method perceives this dynamic range and in the meantime maps it to a lower dynamic range. This is a perceptually nice mapping because it keeps the information important for a human observer (spatial-temporal high-pass components) and reduces or discards the less important aspects (DC levels).

Mitsunaga and Nayar, [34, 37] use a type of spatially variant exposure ([18]), it gives another solution without the loss of temporal resolution. It uses regularly varying exposures, e.g. different exposure values in a 2x2 box, and repeats this box on the whole sensor array. This results in a structure, where an NxN sensor (e.g. 2x2) array belongs to every pixel of the result image. Thus the neighboring cells will have different settings. After taking an image, the pixels whose exposure parameter was not appropriate for the region are omitted (e.g. saturation, dim response). This is followed by a normalization and the method performs an interpolation on the usable pixels to obtain the result high-dynamic range image. In my method the integration time is adjusted according to the local average. Thus - with the exception of high spatial temporal brightness gradients

- we have the perceived intensity in the middle portion of the response curve. My method does not have to omit points as [34] and does not loose spatial resolution by interpolating the neighboring pixels to obtain the result.

Most of the methods mentioned so far retrieve a HDR image of a HDR scene. These methods require an additional step in which the scene is compressed to a lower dynamic range, so that it can be shown on conventional displays. My method does both of these tasks: perception of the high dynamic range and rendering it to a lower range. Hence we need to compare it to high dynamic range rendering methods, to see how it relates to a composed system of HDR perception and HDR rendering.

4.3.1 Comparison to High Dynamic Range Rendering Methods

High dynamic range (HDR) rendering methods are the post processing step of HDR perception. They receive an already captured HDR image, and the task is to compress it to a low dynamic range image without the loss of information and introduction of artifacts. Common artifacts, distortions are: ”halos” (see section 3.9), gradient reversal (local contrast relationships are reversed) and the loss of local contrast.

Pattanaik et al. ([42]) uses a special way of computing the local average. It deals with the problem that if the gain is computed based on the local average, then at high-contrast edges we can observe the ”halo” effect, similar as it was noted in section 3.9 and in [24]. The method in [42] solves this problem with a special form of anisotropic local averaging (or anisotropic diffusion). The weighted local average within a circular neighborhood of every pixel is calculated. The weights do not depend on the distance from the actual pixel, they express the effects of the sharp boundaries (4.6).

$$
\begin{aligned}
w(I) &= \begin{cases} 1 & 1/5 \leq |I/I_c| \leq 5 \\ 0 & otherwise \end{cases} && (4.6) \\
w(I) &= e^{-|log_5(I)-log_5(I_c)|^{25}} && (4.7)
\end{aligned}
$$

where I is the intensity of the pixel in the neighborhood and I_c is the intensity of the pixel for which the local average is computed. (4.6) punishes the pixels in the neighborhood, whose intensity is smaller or greater by factor 5 than the given pixels intensity. These pixels are omitted from the averaging (their weight is zero). (4.7) is a similar weight, but there is a smooth transition between the weights 0 and 1 at $|log_5(I/I_c)| \approx 1$. The method avoids the "halo" effect and yields perceptually nice results. The anisotropic variation of my method (section 3.9) yielded similar results, thus it can substitute the combination of this method and an HDR sensor. Comparative images can be seen on Figure 4.2.

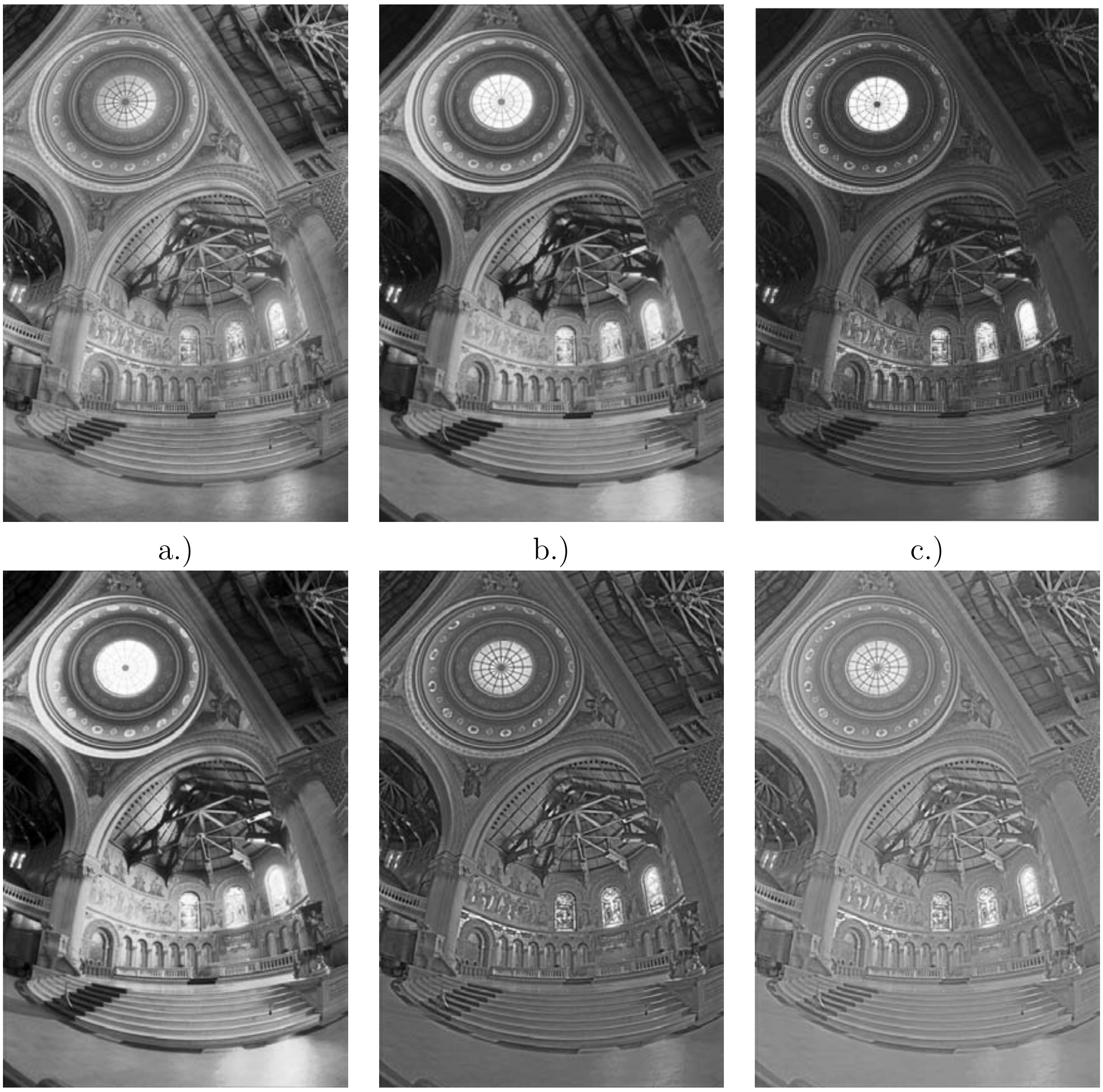
Reinhard's new operator [46] uses a method originating in traditional photography. In photography variable local enhancement was achieved during the development process. Using a paper light was masked at bright regions (called zones) or through a hole on the paper certain areas can be exposed to light for a longer duration [46, 48]. This is the so called dodging-burning procedure.

Reinhard's method first applies a global luminance correction, which corresponds to a logarithmic compression. This correction has a parameter (the so called key value), with which the user can define whether the result should be a light or a dark image.

Then this image is corrected locally based on the local average as shown in (4.8).

$$L_d(x,y) = \frac{L(x,y)}{1 + V(x,y,s_m(x,y))}, \tag{4.8}$$

where (x,y) is the coordinate of the given pixel. L_d is the locally corrected luminance. L is the luminance of the pixel and $V(x,y,s_m(x,y))$ is the local average at the given pixel. This local average is computed with a Gaussian kernel of radius (σ) $s_m(x,y)$. The important feature of the method is that the radius, s_m varies locally as well. Blurred images at several scales are computed. For each pixel they use the blurring result of the coarsest scale where no large changes occur within the region considered for the local averaging. This step avoids the "halo" effect, because differently illuminated pixels on the other side of a sharp jump are not taken into account. These differently illuminated pixels would cause an inadequate enhancement along the high-contrast edges as shown in section 3.9. The regions over which no large changes occur form the zones of the



a.) b.) c.)

d.) e.) f.)

Figure 4.2: Comparative images of the memorial church image [43]. a.) result of Fattal et al. [44], b.) the result of Reinhard's "new" operator [46], c.) the result of Pattanaik [42] and d.) Larson's result [47]. My method's results are on e.) f.). This is the simulated anisotropic version of my method (section 3.9) with iteration number 50 and difference scaling $a = 5$. The DC component had a ratio $c_{DC} = 0.75$ on e.). f.) shows the result when e.) is enhanced by the user.

system, which correspond to the differently illuminated zones of the photographic development process.

Using anisotropy my method can also realize sudden changes in the scene enhancement as between the zones of Reinhard. My method can also have smooth changes in the integration time, which is useful for eliminating smooth illumination changes as on Figure 3.6. Comparative results on Figure 4.2 show that as Reinhard's, my method achieves good visibility at bright and dark areas without introducing artifacts (halos). Reinhard's method yields higher overall luminance corresponding to the original scene, but it has also an additional parameter (key value) that can specify whether the result should be dark or white. My method yields similar results having a supplement enhancement by the user, at which the luminance was globally increased using a logarithmic transition curve (Figure 4.2 f.)).

Fattal et al. ([44]) uses a gradient based method. His method also assumes that large gradients are caused by illumination changes. This holds because the scene reflectance has usually a dynamic range of 100:1 and larger differences are caused by the illumination. In [44] first the gradients for each pixel of the logarithmically compressed image are computed as the difference to neighboring pixels (4.9).

$$\nabla H \approx (H(x+1,y) - H(x,y), H(x,y+1) - H(x,y)) \tag{4.9}$$

where H is the logarithm of the original image. Logarithmic compression is important, because this way the ratio of the intensities is taken, which corresponds to the human contrast sensitivity (Weber's law). My integration time adjustment also results in a Weber contrast (see section 3.7) and I used similar (proportional) contrast measures in the anisotropic extension of my method (3.22).

In Fattal's method after computing the gradients they are scaled: large values are attenuated thereby reducing the dynamic range. Because of the unequal scaling of the gradients the resulting field is usually not a gradient field of any image, and the resulting image can not be computed directly. The result image is obtained from this scaled gradient field with an optimization process. In this optimization an image is computed whose gradient field is the closest (in mean square difference) to the previously scaled gradient field. The result image I should minimize the term (4.10).

$$\int\int ||\nabla I - G)||^2 dxdy, \tag{4.10}$$

where G is the scaled version of H, where large gradients are attenuated.

The attenuation of large gradients is similar to the anisotropic version of my method. In my method the integration time is determined based on the anisotropically diffused image and thus the sudden large luminance changes are removed. Fattal does not remove them completely. Similarly my method can also restore them in a reduced manner by adding the DC component.

The difference in Fattal's method is that it uses gradients calculated at multiple image resolutions. Hence illumination transition edges, which occur at larger scales, are taken into account due to the different resolutions.

In my method I did not consider multiple resolutions, because wider illumination transitions change the conductances of the anisotropic method along a broader line. In case of a more sudden transition the conductances are more drastically reduced but the fall in a thinner line. A wider but smaller reduction of conductances can also stop the propagation of the luminance values between the strongly and poorly illuminated regions.

Images comparing my method and the algorithms of Pattanaik et al., Fattal et al., Reinhard et al and Larson et al. [42, 44, 46, 47] can be seen on Figure 4.2.

Larson's result shows a good subjective impression, but it has low contrast in bright (windows) areas. Fattal's, Reinhard's and Pattanaik's results show the illumination field of the scene more than my method. It depends on the application which might be better for the user. My method yields stronger contrast on bright and dark areas than any other method. It performs the processing together with the sensing and does not require an HDR sensor.

4.3.2 Comparison to Brendel's work

In [35] methods for image enhancement and image sensing are presented. The image enhancement algorithm computes a high-pass filtered image by subtracting the diffused image from the original image. It also computes a local contrast term: the difference of a given pixel to its neighbors. Then a weighted summation of the high-pass filtered image, the contrast image and the original image is computed

(4.11).

$$E = I + k_1 I(1 - k_2 D(I^2))^n + k_3 C(1 - k_4 D(C^2))^m \tag{4.11}$$

where I is the intensity of the perceived image, C is the contrast computed as the difference between a given pixel and its neighbors. The method requires six parameters to tune ($k_1, .., k_4, n, m$) beside the diffusion strength. The intensity is to be understood with CNN signal values which means -1 for white and 1 for black.

The first correction term (k_1) drives the intensity values toward their extreme 1 and -1 depending on whether they are below or above 0 (mean gray value). This expression is thought to abolish gray regions on the image. This is useful if we want to emphasize the intensity differences between the regions of the scene. However, in high dynamic range perception our target is to decrease these differences. The factor with k_2 reduces this correction if the region's intensity is near -1 or 1. First we have to state, that it is even problematic, that the dark value's enhancement is reduced.

The computed contrast term (k_3 term) is a differential-type of contrast: the difference is not scaled with the intensity of the given neighborhood. In my method we enhance the image according to the proportional (Weber) contrast (see section 3.7). By adjusting the integration time to be inversely proportional to the local brightness (3.15) the small differential contrast points on low-illuminated regions are scaled to the same order as those on high-illuminated regions (3.9).

In Brendel's work the differential contrast (the term k_3 in (4.11)) is scaled by a term which decreases by the square of the (average local) differential contrast ($-D(C^2)$). Even if the local average contrast is proportional with the local intensity, the square decrease is different from an inversely proportional relationship. The contrast of a region might be enhanced due to the first term, which increases the intensity depending on a factor which decreases with the square of the diffused intensity ($-D(I^2)$). The problem with this factor is that due to the square it punishes the dark areas in a similar way as the bright areas, and their intensity (and contrast) is less changed.

Furthermore, if a region has low contrast, its contrast is enhanced regardless of its brightness. This is useful in case of a contrast enhancement task, but not in case of high dynamic-range perception. In the latter case the goal is to retrieve the same contrast relationships that a human observer would perceive, and enhance the low-contrast at the poorly illuminated regions.

The method is extended to adaptive perception using local adjustment of the integration time. In this case the integration time is calculated based on the image enhancement method, but no exact formula is given for the calculation. Hence the statements about the enhancement method are valid.

My method requires only 2 parameters: the extent of the diffusion and the weighting of the DC component (section 3.8). The additive correction's iteration number can be fixed on 3. My method scales the intensity inversely proportional to the scene's local average. Contrast enhancement follows implicitly, if the intensity is increased in a region, then the intensity differences are enhanced in the same rate according to Weber's law.

4.4 Measuring the Method's Dynamic Range

One key measure of high dynamic range capturing methods is the sensor's dynamic range (4.12).

$$D = 10 \cdot log\left(\frac{I_{max}}{I_{min}}\right) [dB] \tag{4.12}$$

where D is the dynamic range in decibels. I_{max} is the maximal and I_{min} is the minimal perceivable intensity. Dynamic range can be calculated in bits, in this case the 2 base logarithm is taken and the multiplication with 10 is omitted.

This dynamic range is difficult to measure in the case of my adaptive algorithm. My method is not bound to a specific adaptive sensor, it can be implemented on any sensor with locally adjustable integration time. Thus the spanned dynamic range can be the dynamic range of the integration time parameter plus the sensor's dynamic range. There is only one constraint to this: the integration time can not have sudden changes. In case the illumination changes within the low-pass filter's radius the method results in the "halo" effect (section 3.9), unless anisotropic diffusion is used. Regarding the adaptive method the question is not the spanned dynamic range, but the range within which the method can change from highly illuminated region to a poorly illuminated one.

I carried out tests to measure the region within which the method can adapt to dynamic range differences. Artificial scenes were taken, where the right side was 30dB, 50dB and 100dB lighter than the left side, with a sharp intensity step between them. Captures with different integration time were simulated, so that with the lowest time the right side is black, and with the highest integration time, even the darker side is nearly saturated white. The size of the images was 600x40.

Three captures in the case of a 30dB scene can be seen on Figure 4.3.

Figure 4.3: The artificial scene of 30dB dynamic range captured with integration time $T = 10$ms (a.), $T = 511$ms (b.) and $T = 57,5$sec (c.)

Of course in the case of 50dB and 100dB these images are quite unnatural, however they are useful to test the method.

After adaptation we get a result where the interior parts of the dark and bright areas are perceived with medium gray intensity, but along the transition between them we get a white stripe on the bright side and a black stripe on the darker side (see Figure 4.4). This is a case of the "halo" effect mentioned earlier (section 3.9).

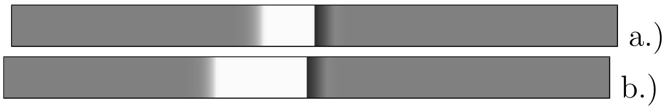


Figure 4.4: Results after adaptation. a.) shows the result in case of the 30 dB dynamic range, b.) in case of 50dB.

The white stripe is wider than the black stripe. This is due to the fact that in the algorithm we always make a spatial low-pass filtering on the shutter map (see Figure 3.5 and (3.10)). Thus the high and the small integration time values on the different sides of the intensity step influence each other. On Figure 4.5 a.) we can see that around the transition point of pixel 300 both the dark and the bright side are affected approximately in the same manner. Though this is illusory, because in linear scale there are tiny changes of the ideal integration time both on the dark and the bright side. These remain unnoticed because of the integration time's high dynamic range. However these changes are negligible to the large integration time values on the dark side, they mean a relatively big change for the bright side's short integration time.

If we plot the integration time on a log scale, then the changes in integration time, which are small relative to the high dynamic range but large with respect to the bright region's adapted short integration time are easy to recognize. On Figure 4.5 b.) the same integration time profile is shown on a log scale. We can see that the bright side reaches its adapted integration time value approximately 100 pixels from the intensity step.

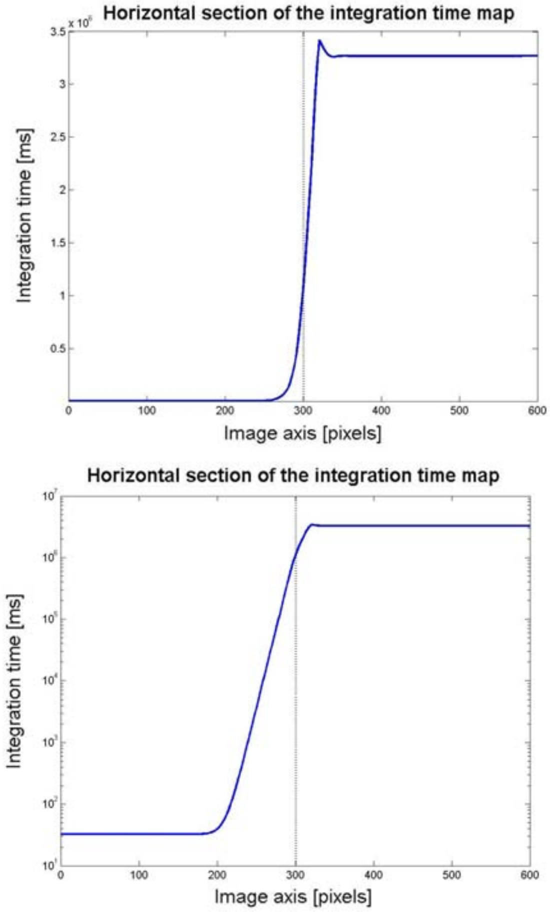
Both the bright and the dark parts are affected in a similar manner by the other side's values (see 4.5 a.)) because of the low-pass filtering of the integration-time map. However this is a relatively huge change for the small integration time values of the bright side it is a less perceivable change for the huge integration times of the dark side (see 4.5 b.)).

The width of the dark-influenced zone on the bright side depends on the dynamic range and the radius of the low-pass filtering of the shutter map (see (3.10) and (3.11)). The width of the zone increases linearly with the radius and nearly linearly with the dynamic range (see Table 4.1). A comparison between a linear relationship and the actual data can be seen on Tables 4.1 and 4.2.

The linear relationship between the transition zone width and the spanned dynamic range can be seen on Figure 4.5 b.) as well. The log integration time rises (approximately) linearly along the spatial axis. Thus within a given interval the integration time rises with a certain amount of decibel.

On the dark side of the intensity step, the width of the transition zone is far less (see Table 4.1). It is independent from the spanned dynamic range. The reason for these effects is that the subtle alterations of the dark parts integration time caused by the low-pass filtering are negligible relatively to the adapted integration time. Based on Table 4.1 the width of the transition zone on the dark part can be expressed as (4.13).

$$w_d = c_d \cdot \sigma \tag{4.13}$$



a.)

b.)

Figure 4.5: Horizontal sections of the integration time map after adaptation to the 50dB scene (see Figure 4.4). a.) shows the integration time values on linear scale, b.) shows it on log scale. The dotted line shows the location of the intensity step between the scene's bright and dark areas.

where w_d is the width of the transition on the dark side, σ is the parameter of the Gaussian kernel used for the diffusion in pixels, c_d is the constant which shows the connection between the width and the filtering radius and $c_d = 1.5$.

Based on these measurements we can approximate the width needed for a transition on the bright side at large dynamic range intensity steps:

$$w = c \cdot \sigma \cdot D, \tag{4.14}$$

where w is the width of the transition, σ is the parameter of the Gaussian kernel

radius / dB	5	10	20
30	16	31	63
50	26	51	102
100	51	99	200
dark tr	8	15	31

Table 4.1: Width of the transition zone on the bright side in pixels. The columns show the different filter radii, the rows show the dynamic range. The last row shows the width of the transition zone on the dark side of the intensity step.

radius / dB	5	10	20
30	16.00	32.00	64.00
50	26.67	53.33	106.67
100	53.33	106.67	213.33

Table 4.2: Linear approximation of the width of the transition zone on the bright side in pixels. The columns show the different filter radii, the rows show the dynamic range. The approximation was based on the transition width at $radius = 5$ and dynamic range: $30dB$

used for the diffusion in pixels and D is the spanned dynamic range in dB, c is the constant which shows the connection between the previous measures. $c = \frac{1}{10dB}$, which means 10 dB dynamic range can be spanned within σ distance.

The width of the white stripe on the bright side is due to the larger integration time values. As said earlier this is caused by the low-pass filtering of the integration time map. The stripe is wider on the bright side, because here even a distant and small effect of the diffused large integration time values of the dark part can mean huge changes. One can theoretically reduce this effect if the integration time is mapped to the log scale, then diffused and then mapped inversely (exponential function is taken). Such a result can be seen on Figure 4.6.

In case we take the logarithm of the integration time for the filtering and after the filtering we convert it back, then during the filtering the great differences in the integration times of the dark and bright pixels are reduced. As a result of this the large integration times (dark pixels) alter the small values less (bright pixels), but the small values influence the dark pixel's large integration time more, and

Figure 4.6: Adaptation results in case of the 100 dB test scene. The filter radius was 20. In case of a.) the integration time map was filtered according to the algorithm, on b.) the logarithm of the integration time.

thus we have a wider black stripe (Figure 4.6). This modification of the algorithm requires the accurate calculation of the logarithmic transfer curve and it's inverse or we might have immense errors. In simulation environment this calculation was possible, however its feasibility is questioned on an analog hardware architecture.

Finally we can declare that **the method can span huge intensity changes** within a usual image size (e.g. 640x480), but the transition zone can be the fifth of the image, unless we use anisotropic diffusion or modified shutter filtering.

4.5 Conclusions

In this chapter my method was compared to other high dynamic range perception methods, and it's dynamic range was also computed.

My method was compared to logarithmic sensors and local adaptive sensors: spatially varying sensors, multiple captures method, time-to saturation methods. I analyzed dynamic range compression methods in order to compare the method's performance to a combination of a HDR sensing and compression algorithm.

My methods performs similar results as cutting-edge dynamic range sensing or compression methods, but it combines sensing and compressing in the same algorithm.

Dynamic range depends on the integration time range of the applied sensor. The methods set a minimal limit to the width of a transition zone between bright and dark regions of a scene. The width of a transition zone for spanning of a given dynamic range was specified.

Chapter 5

Analysis of the Retinal ON and OFF Interactions

This chapter focuses on another application of the CNN-UM in visual processing. In the previous chapters biologically inspired visual systems were discussed, here I target the retina itself. CNN-UM had already some useful applications in the retinal modeling [14, 56], here I refine and apply these models for analyzing some specific retinal processes. My simulations revealed the role of some less understood retinal interactions.

5.1 Introduction

Simulation of biological functions can serve as a useful tool for understanding complex neural interactions. Simulations require accurate models, which are based on physiological measurements, morphological knowledge about the cells and the qualitative models made by neurologists. Recent measurements [15] revealed multichannel retinal dynamics, in which the visual input is represented through about a dozen different spatial-temporal channels. This gave rise to a new multichannel model of the retina. A short overview of the retinal structure can be found in Section 2.2. The CNN-UM (Universal Machine) paradigm has proved to be an effective tool for making a quantitative model of these retinal channels via multilayer CNN models [14, 56, 67]. This is due to the neuromorphic structure and cell-dynamics of the models. CNN also has a real-time physical implementation ([27]), so it provides a basis for implementation of retinal prosthetic devices as well.

In my work I further refined this model based on recent measurements. The model was applied for the analysis of nonlinear, retinal interactions. Simulation can be useful for clarifying biological function. A variety of retina manipulation scenarios can be tested in ways that are prohibitive in the "wet" biological preparation. My work was based on the effects of rectifying nonlinear synapses.

Synapses between retinal cells are rectifying, yet the retinal output is often nearly linear. My quantitative nonlinear spatial-temporal modeling shows how neural interactions correct for, and re-linearize signals as they pass through the neural structure of the retina.

Given a stimulus the center is the area which excites the retinal bipolar or ganglion cell, the surround (usually around the center) is the area which inhibits it. Stimulus in the surround area inhibits the cell, because the signal from the surround reaches the cell through inhibitory (horizontal) cells (See Section 2.2.3). At the outer plexiform layer the photoreceptor-horizontal cell interactions [57] create the antagonistic center-surround properties of bipolar cells [68]. This antagonism enhances the spatial-temporal high-pass component of the signal, like the spatial edges of a stimulus. Cells lying in the central part of a large stimulus give medium response because their inhibitory surround receives the stimulus as well. Cells lying outside the stimulus have negative response, because they receive only inhibitory surround (from cells lying under the stimulus). Cells lying along the edges of the stimulus have the highest response [1]: the center is excited, but some parts of the surround are lying outside the stimulus and do not inhibit the cell. The process of neural edge enhancement in the retina begins by the creation of an antagonistic (inverse sign) surround in retinal bipolar cells.

However, the edge-enhanced patterns are degraded as they course through the retina. The synapse from bipolar to ganglion cells is strongly rectifying (similar observation in: [51, 55, 58, 59]), so that only the depolarizing (positive) components of the neural images from the bipolar cells arrive at, and provide excitation to the ganglion cells. This is detailed in section 5.5. Because ganglion cells have dendrites with finite extent ([52]), these excitatory signals are diffused, thereby

---

[1] In case we consider a spot stimulus, then this spot has the highest response, because the surround is outside the stimulus.

blurring the neural image. Blurring is possible because the excitation is devoid of its antagonistic counterpart, which was negative.

In my simulations I showed that these degrading factors can be eliminated by a less known inhibition, the cross-inhibition which combines the information between the ON and OFF systems (see more details in section 5.5.2). These two systems respond to a light stimulus inversely: ON bipolars and ganglions give positive response to light increase, OFF cells give negative response. My simulations showed how the effects of rectification and blurring are eliminated.

Before I detail the actual simulations I review the underlying background information. In section 5.2 the model elements are qualitatively described. In section 5.3 the model elements are shown in detail with a quantitative description of the parameters. Section 5.4 shows a comparison between the multilayer CNN-UM retina model and a simple linear filter model. Section 5.5 deals with the new measurements that modified the earlier models. Section 5.6 contains the simulation results and section 5.7 interprets these results.

5.2 Methods

This chapter focuses on the results of the analytical modeling. The model is based on earlier basic findings [15] and recent measurements [54, 65]. The methods of the CNN modeling are briefly described in sections 5.2.1 and 5.2.2.

5.2.1 Analytical Modeling

CNN-like models of the rabbit retina were used to explore responses to fast flashes based on the work of Balya et al. ([14]).

The model was comprised of processing elements corresponding to the retinal cells ([66]) as shown in Figure 5.1. Retinal cell-dynamics' are modeled with CNN cells and a layer of retinal cells corresponds to a CNN layer (section 5.3.1).

Vertical synapses between different cells are modeled with inter-layer connections. The horizontal spread of the retinal signal due to dendritic tree width and gap-junctions is modeled with a diffusion template on the CNN architecture ([11]), the actual scaling of the diffusion template can be seen in section 5.3, equation ((5.5) and Table 5.1). The dendritic tree expansion has a diffusion-type

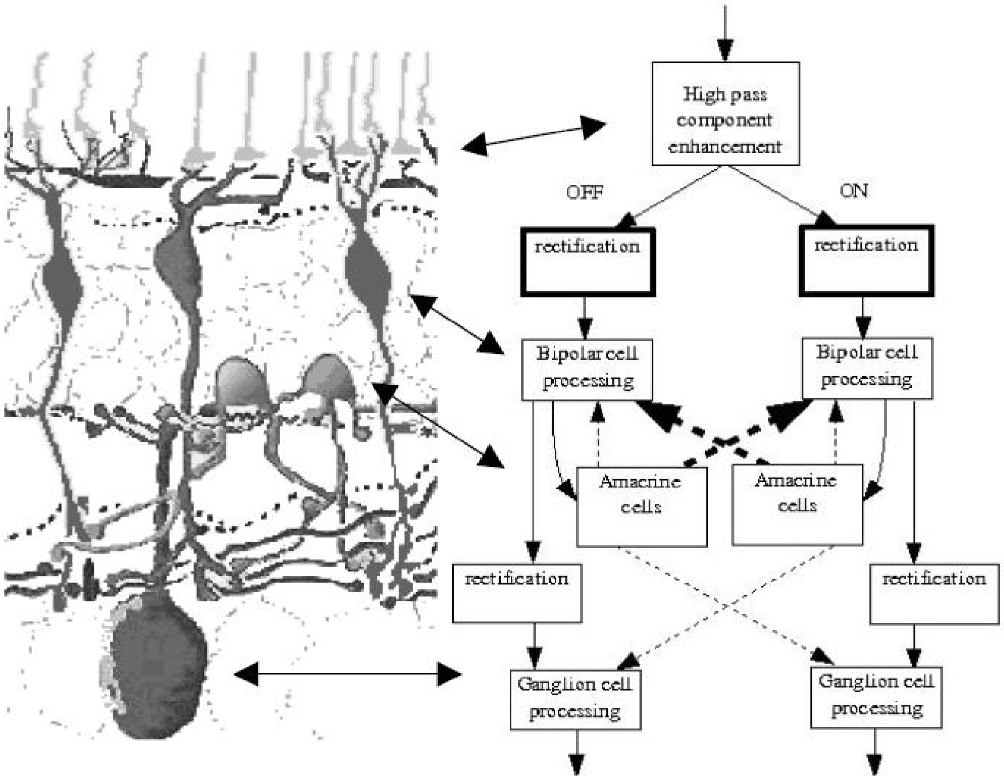


Figure 5.1: The main cell types of the retina and their interaction. On the left side we can see a drawing of the retinal cells. The right side shows a block diagram of the main elements of the retinal model. Double sided arrows show the corresponding elements of the model and the retinal picture. Solid arrows on the model show excitatory, dashed arrows show inhibitory interconnections. Bold boxes and arrows show the new elements of the model compared to [14].

(spatial low-pass filtering) effect, because due to it a cell collects the output from many presynaptic cells thereby blurring the image. The extent of this blurring depends on the dendritic tree radius (the quantitative expression is in section 5.3, equation (5.5)).

I was able to add and remove elements of the model by changing the parameters to perform modified-retina experiments that would have been impossible in the physiological retina (see section 5.3.3).

The spatial-temporal high-pass component enhancement shown in Figure 5.1 performs the processing of the outer retina (OPL, outer plexiform layer). The outer retina consists of cone photoreceptors and horizontal cells, which have feed-back to the cones. Spatial-temporal high-pass filtering is the result of the horizontal cell's feed-back. They perform a low-pass filtering of the cone signal. Then the horizontal cells make an inhibitory feed-back to the cones which reduces the low-pass component [57, 12].

The output from the OPL is rectified (Figure 5.1 rectification) at the cone-bipolar synapse [61] and fed forward to the bipolar cells which transmit the signal to the ganglion cells. At the cone-bipolar synapse, the ON and OFF pathways are created. The ON bipolar cells respond to an increase in the incoming light while the OFF bipolar cells respond to light decrease. This is due to the difference in sign between the cone-ON bipolar and the cone-OFF bipolar synapse [71, 72, 73, 74]. Bipolar cells excite amacrine cells, which make inhibitory feed-back connections to them. Bipolar cells also receive "crossover" feedback from amacrine cells, which are excited by the other system's bipolar cells ([63, 64]). This means that the ON bipolar cells receive OFF inhibition and the OFF bipolar cells receive ON inhibition (shown with bold arrows on Figure 5.1). This is the feed-back cross inhibition.

The bipolar-ganglion synapse has a nonlinear rectifying characteristic which is followed by spatial low-pass filtering corresponding to the physical extent of the ganglion cell's dendritic trees. The ganglion cells also receive crossover inhibition that is fed forward such that ON ganglion cells receive OFF inhibition and OFF ganglion cells receive ON inhibition mediated by amacrine cells. This is the feed-forward cross inhibition.

The difference between this work and earlier models ([14]) is in the existence of the cone-bipolar synapse' rectification and the use of feed-back cross inhibition via amacrine cells (Figure 5.1). These modifications were obtained directly or with parameter tuning (section 5.2.2) from the measurements (section 5.5).

5.2.2 Parameter Tuning

The spatial processing of individual cells was modeled as a low-pass filter, which is realized as a diffusion on CNN architecture according to the receptive field calculus (see [14] and section 5.3). The extent of the diffusion was calculated from the dendritic tree radius specified based on morphological data and previous measurements. A dendritic tree radius of $80\mu m$ was assigned to the ganglion cells and a $8\mu m$ radius to the bipolar cells ([53, 52]).

The temporal processing of the amacrine and ganglion cells was adjusted to fit ganglion excitation, inhibition and spiking measurements (as done earlier in [14]).

The temporal nature of the cone-bipolar pathway was modeled based on recent bipolar cell measurements ([54]). The processing of the cells and synapses were modeled as first and second order temporal filters followed by a memory-less nonlinear function. The response of the bipolar cells to a sinusoidal light stimulus was measured. The magnitude of this response varied according to the frequency of the sinusoid stimulus. The response contained not only this frequency but its higher harmonics as well, which were introduced by the nonlinear nature of the cone-bipolar synapse. The measurements were repeated at different input frequencies and the filter characteristics were fitted to the measured sinusoid responses (temporal sinusoid measurements on bipolar cells [54]). The parameters of the nonlinear characteristic were calculated from the second and third order harmonic components of the responses to sinusoidal inputs (similar to: [51]).

The nonlinear function was approximated with its Taylor polynomial:

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 \tag{5.1}$$

where $a_0..a_3$ are the coefficients and x is the input.

Substituting an input sinusoid of frequency $k\omega$ in (5.1) (input: $x = sin(k\omega t)$), the nonlinearity introduces the following upper harmonic components and alters the fundamental as it can be seen in (5.2).

$$\begin{aligned} f(x) &= \tfrac{a_2}{2} + a_0 + (a_1 + \tfrac{3}{4}a_3)sin(k\omega t) - \tfrac{a_2}{2}cos(2k\omega t) - \tfrac{a_3}{4}sin(3k\omega t) = \\ &= n_0 + n_1 sin(k\omega t) + n_2 cos(2k\omega t) + n_3 sin(3k\omega t) \end{aligned} \tag{5.2}$$

With the analysis of the 1st, 2nd and 3rd order harmonics (n_1, n_2, n_3) of the measured bipolar response we can specify the Taylor coefficients and this reveals us the nonlinearities of the cone-bipolar synapse.

A result of the cone-bipolar synapse modeling can be seen in Figure 5.2. The results show that a good fit to the measured data could be achieved. Results from other cells show that approximately 33% of the bipolar cells have nonlinear input synapses. A typical nonlinear rectifying synapse is shown on Figure 5.2. This modification was included in the model (Table 5.2 in section 5.3.2).

5.2.3 Limitations and Possible Sources of Error

My model of the cone-bipolar pathway approximates the cells and their synapses with distinct second order filters and static nonlinear functions. It does not model nonlinear filters or third (or higher) order filters. However the current amount of data was not sufficient to unambiguously define models with more free parameters. Noise sources existing the retina were not included in my model. The goal of this analysis was to explore the effect of nonlinearities and the function of cross-over inhibition. For about half of the measured cells, the degree of nonlinearity is smaller than in my model. However, the exaggerated nonlinearity helps to explore the effects of rectification and the ways that rectified inputs can interact.

5.3 Model Parameters

In this section the actual numerical values of the model parameters are reviewed.

First we show the parameters. The simplified CNN cell dynamics used in the model can be seen in (5.3). For a more detailed description of CNN state equation consult Section 2.1.

$$\begin{aligned}\tau_l \frac{dx_l(i,j)}{dt} &= -x_l(i,j) + \sum_{p=-1}^{1}\sum_{q=-1}^{1} D^{r_l}(p,q) y_l(i+p, j+q) + b_l u(i,j) + \\ &\quad + \sum_{k=1}^{L} f_{kl}[a_{kl} \cdot y_k(i,j)] \qquad (5.3) \\ y_l &= 1/2(|x_l+1| - |x_l-1|) \qquad (5.4)\end{aligned}$$

$x_l(i,j)$ is the state-variable of the cell with coordinates (i,j) in the l-th layer $l \in [1,L]$. If the image size is M x N then $i \in [1,M]$ and $j \in [1,N]$.

τ_l is the time constant of layer l.

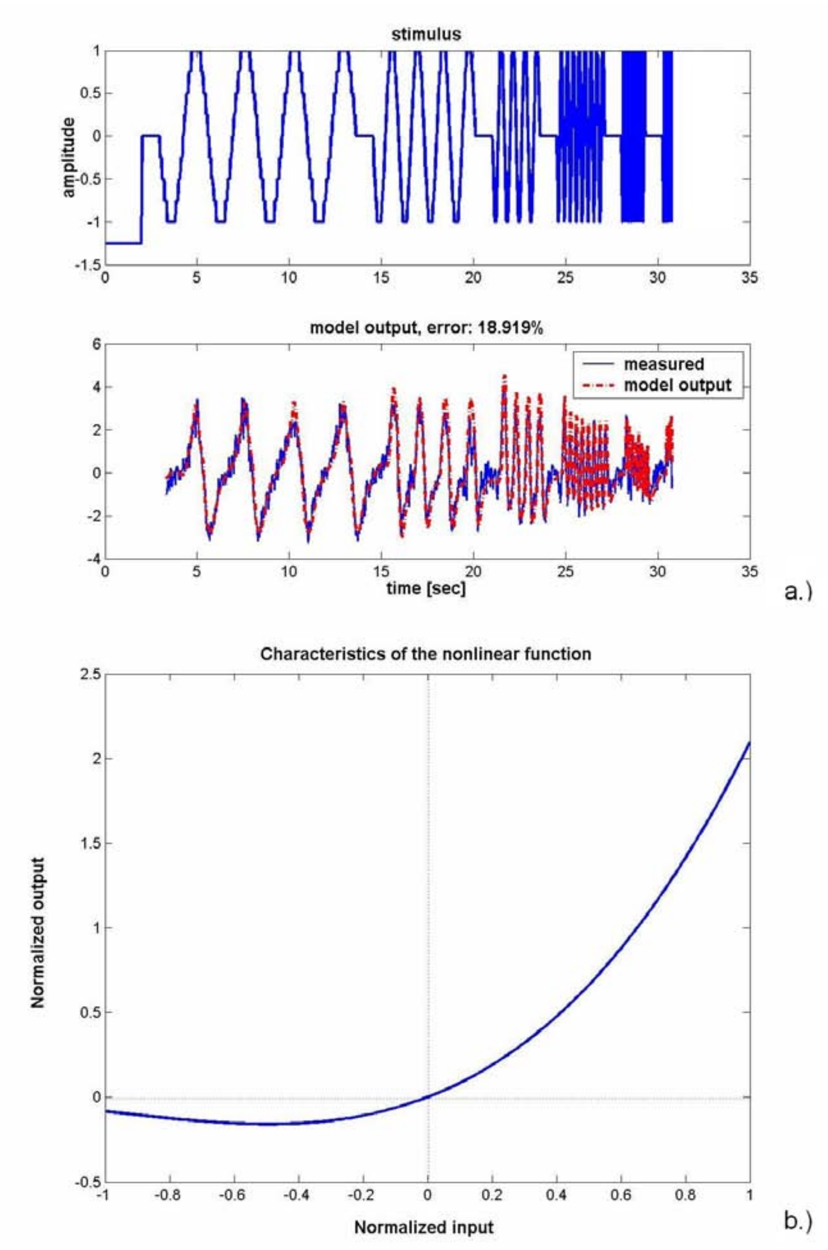


Figure 5.2: Results of bipolar cell simulation. a.) The top trace shows the input stimulus. The bottom traces show the measured response and the model output. The initial 3.29 seconds of the response is not shown because that is a response to a step stimulus, the last few seconds are omitted because of high frequency. These parts of the response were not the basis of the modeling. b.) The curve of nonlinear function. We can see that values between $[-1, 0]$ are almost totally cut off.

y_l is the output of the l-th layer, which is a piece-wise sigmoid nonlinear function of the state-variable x_l.

D^{r_l} is the diffusion template of the l-th layer, and r is the dendritic tree radius. The diffusion template collects the outputs of the neighboring pixels (expressed in the double sum on p, q).

b_l is the weighting of the input stimulus ($u(i, j)$) on layer l (it is non-zero only for the receptor layer).

a_{kl} expresses weight of the synapse between the k-th and the l-th layer, and $f_{k,l}[.]$ is the nonlinear characteristic of the synapse (for linear synapses: $f(y) = y$).

5.3.1 CNN Layers

The model had 11 layers, which are corresponding to the model-elements on Fig. 5.3.

- The spatial-temporal high-pass component enhancement corresponds to the cone and horizontal cells. Because the cones have second order temporal dynamics they were modeled with two layers (cone, cone2). In the majority of the simulations the feed-back temporal (cone2) and spatial (horizontal) high-pass filtering was approximated by feed-forward connections, because it needed smaller time constants and diffusion strength.

- The rectifications before and after the bipolar cells are realized with nonlinear receptor transfer-functions.

- The ON, OFF bipolar and ganglion cells on Fig. 5.3 have their corresponding CNN layers.

- The bipolar - amacrine - bipolar cell feed-back connections are modeled with two layers, the amacrine cells feeding to the ON bipolar cells (ON FB amacrine), and the amacrine cells feeding to the OFF bipolars (OFF FB amacrine).

- The bipolar - amacrine - ganglion cell feed-forward connections are modeled with two additional layers, the amacrine cells feeding to the ON ganglion

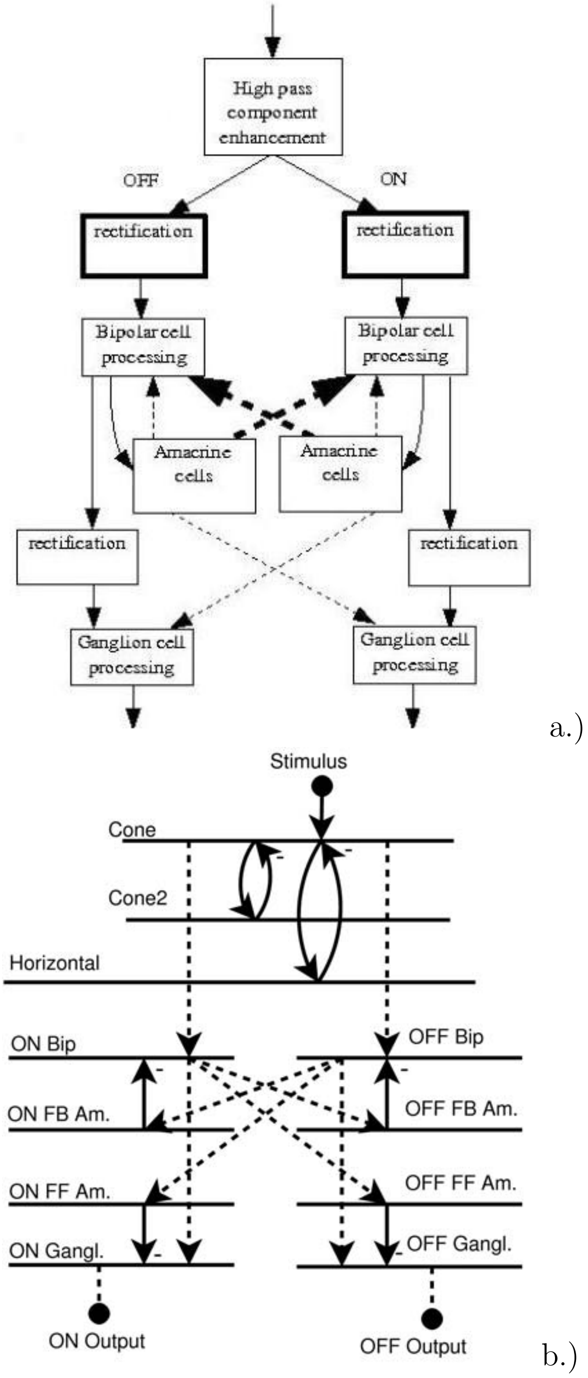


Figure 5.3: a.) The processing elements of the retina-model (part of Fig. 5.1) and the associated CNN layers b.) Dashed lines denote rectifying synapses. Inhibitory synapses are marked with a "-" sign.

cells (ON FF amacrine), and the amacrine cells feeding to the OFF ganglions (OFF FF amacrine).

5.3.2 Layer Parameters

In the followings the actual values of the time constants τ_l, the diffusion template matrices D^{r_l}, and the weights b_l and a_{kl} are shown.

The diffusion template matrix can be seen in (5.5) according to [14].

$$D^r = \frac{(r/4)^2}{10} \begin{bmatrix} 1/2 & 1 & 1/2 \\ 1 & -6 & 1 \\ 1/2 & 1 & 1/2 \end{bmatrix} \tag{5.5}$$

where r is the dendritic tree radius in μm and it is 4 times the space constant (λ) in [14], ($r = 4\lambda$). The diffusion matrix of the layers is thus specified by r_l.

The dendritic tree radii (r_l), the time constants (τ_l) and input weights (b_l) can be seen in Table 5.1.

	$r_l[\mu m]$	$\tau_l[ms]$	b_l
cone	0	20	-0.75 (-1)
cone2	8	60 (20)	0 (-1)
horizontal	280 (80)	20	0
ON bipolar	8	5	0
OFF bipolar	8	5	0
ON FB amacrine	8	5	0
OFF FB amacrine	8	5	0
ON FF amacrine	8	5	0
OFF FF amacrine	8	5	0
ON ganglion	80	50	0
OFF ganglion	80	50	0

Table 5.1: Dendritic tree radii (r), time constants (τ_l) and the input weights (b_l) of the different cells. The values in parentheses are in case of feed-forward OPL model.

Inter-layer connection weights can be seen in Table 5.2.

Rectification is applied after the eventual sign inversion of the synapse. Cone-bipolar synapses and bipolar outputs are rectified. Where no rectification is applied the transfer-function is linear.

Post-synaptic (TO) Pre-synaptic (FROM)	cone	cone2	horizontal	ON bipolar	OFF bipolar	ON FB amacrine	OFF FB amacrine	ON FF amacrine	OFF FF amacrine	ON ganglion	OFF ganglion
cone	X	-1	-3								
cone2	1	X									
horizontal	1		X								
ON bipolar	-4^+			X		-1					
OFF bipolar	4^+				X		-1				
ON FB amacrine					1^+	X					
OFF FB amacrine				1^+			X				
ON FF amacrine					1^+			X			
OFF FF amacrine				1^+					X		
ON ganglion				1^+				-1		X	
OFF ganglion					1^+				-1		X

Table 5.2: Weights of inter-layer connections (a_{kl}). Columns show the pre-synaptic cell, which transmits the signal to the post-synaptic cell (rows). $^+$ signs show the locations of the rectifications ($f_{kl}(y) = max(y, 0)$).

In case of the feed-forward OPL model the connection weights of the first 5 layers are in Table 5.3.

5.3.3 Retina-Modification Experiments

The switching off of different elements occurs with setting the weight to the post-synaptic cell to zero (e.g. switching off feed-forward cross inhibition).

In case of switching off the bipolar nonlinear output and the cone-bipolar nonlinearity the cross inhibitory networks were also switched off. This was necessary, because in this case both the ON and OFF bipolar had the linear signal, and their interaction would have reinforced the already existing signal.

When the OPL's processing was switched off, the cone-bipolar synapse weight was reduced because the cone signal was larger due to the lack of inhibitory horizontal cell feed-back.

Pre-synaptic (FROM) Post-synaptic (TO)	cone	cone2	horizontal	ON bipolar	OFF bipolar	ON FB amacrine	OFF FB amacrine
cone	X	-0.67					
cone2	0	X					
horizontal	1		X				
ON bipolar	[-4		3.2]$^+$	X		-1	
OFF bipolar	[4		-3.2]$^+$		X		-1

Table 5.3: Weights of inter-layer connections (a_{kl}) for the OPL and the bipolar cells in case of the feed-forward model. Columns show the pre-synaptic cell, which transmits the signal to the post-synaptic cell (rows). $^+$ signs show the locations of the rectifications. In case of the bipolar cell input synapse the rectification is applied after the weighted summation of the cone and horizontal inputs.

5.4 Comparison of CNN and Band-pass Filter Retina Models

Before I discuss the measurements and the simulations, I shortly review other modeling approaches, compared to which the neuromorphic model suits better for retinal analysis. Beside the multilayer CNN-UM model there are other modeling approaches. I compared the CNN-UM model to a black-box model which was built from spatial and temporal filters. Such a model can be seen in [17] based on which an input-output model was made in Berkeley for the multichannel retina and later a bit more sophisticated model in [69]. These models are different from the multilayer retina model in their building elements and in the way they construct the model. I compared the models regarding both of these aspects. In section 5.4.1 we compare the building elements, in section 5.4.2 the two distinct approaches, and in section 5.4.3 we show some comparative results.

5.4.1 Building Elements of the Two Models

The CNN-UM model was already described in section 5.2 and 5.3, here we review the black-box filter model ([17]).

The model contains distinct elements for the temporal and the spatial processing. The temporal function is implemented with simulating a simple temporal differential equation (5.6). The spatial function is implemented with a Gaussian convolution the sigma parameter of which equals the dendritic spread (width of the dendritic tree). The model contains non-linearities which are specified with their Taylor series or are a simple rectification cutting off positive or negative values.

Considering a CNN layer without spatial coupling, it has a similar function as a temporal low-pass filter, which is straightforward from the differential equations of the two models (5.6). This holds although the CNN model has a piece-wise sigmoid nonlinear output, because the model operates mostly in the linear range.

$$\tau \frac{dx}{dt} = -x + \sum_{k \in L} a_k \cdot I_k \tag{5.6}$$

where x is the state of the given layer and τ is its time constant, L is the set of layers that synapse to the given layer (which might contain the stimulus as well), a_k is the weight of each synapse.

Nonlinear functions are realized in both models. Delay element is currently implemented in the black-box model, it can be a distinct temporal element. The delay element in the CNN-UM model is approximated with a CNN layer without coupling [14], which is a low-pass filter, as it can be seen in (5.6).

The large difference of the two models' building elements lies in the realization of spatial processes. The filter model realizes spatial processes separated from temporal processing. There are temporal filters and in a distinct element a convolution with a Gaussian kernel. In the CNN model the spatial blurring is caused by connections to the neighboring elements through a diffusion template (see (5.5)).

Through proper weighting both models take into account that distant cells have smaller effect on a given cell than proximal ones. Hence both models realize a blurring equivalent to a Gaussian blurring. The difference between the two approaches is that in the filter model the spatial summation occurs in one step over the whole dendritic tree. It might contain delay elements (Figure 5.4) but the delay is not depending on the spatial distance, because the temporal process is separated from the spatial. In the CNN-UM model the signal from distant elements arrives later than from proximal elements (see Figure 5.4).

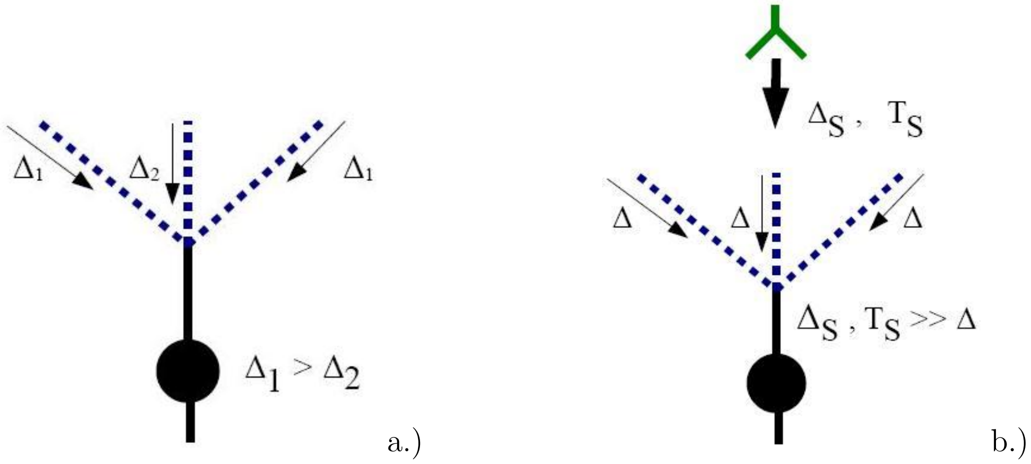


Figure 5.4: Figures showing the spatial and spatio-temporal processes in the two models. a.) shows the spatial integration in the CNN model. The distant elements have a larger spread (Δ_1) duration than the center ones (Δ_2). b.) shows the filter model's blurring. The spatial spread duration is constant over the whole dendritic field (Δ). Synapse delays and time constant are shown with Δ_S and T_S respectively.

The question arises which of the two is biologically relevant, and how important is this difference. The spread of the signal in the dendritic tree is usually negligible compared to the duration of synaptic processes. The first in a membrane potential, while the later operates with chemical transmitters. Furthermore it is only the difference between the signal spread duration of the centrally located and distally located dendrites that requires spatio-temporal process, a universal delay can be modeled with separate temporal delay in the filter model.

A summary of the comparison of the different spatial and temporal processes can be seen on Table 5.4.

Mathematically the elements of the two models are very similar. The difference lies in the fact that the CNN model has a spatio-temporal low-pass filtering while the other model carries out a spatial convolution.

	CNN model	filter model
temporal filtering	+	+
delay	-	+
Gaussian spatial blurring	+	+
spatio-temporal blurring	+	-

Table 5.4: Comparison of the elements of the CNN and the filter model.

The CNN-UM has efficient hardware realization [70]. On simulation environment the filter model's building elements can be simulated faster, because they can be realized with digital temporal filters and spatial convolution, while the CNN cells require the computation of a differential equation.

5.4.2 Two Approaches of Modeling

The differences in the building elements of the two models are not significant regarding the current stage of retinal modeling. The greater difference between the neuromorphic retina model and the black-box filter model lies in the structure and the construction of the two models.

In the model presented in [14] the modeling elements can be assigned to the retinal cells and layers. The model is built on morphological knowledge, which defines the number of layers, the connections between them and their diffusion strength based on the dendritic tree radius. The strength of the excitatory and inhibitory connections, the layer time constants and the fine-tuning of the diffusion strength are the result of fitting the parameters to ganglion cell excitation, inhibition and voltage measurements (see sections 5.2 and 5.3).

The black box model is built in a deductive way, the parameters are derived from the measurements. The goal is to produce a minimal model that retrieves the retinal output. This model consists of a second order temporal filter followed by a difference of Gaussians operation. The first Gaussian has a smaller blurring, the latter has a larger. The output is rectified in order to be able to produce the nonlinear ganglion output. The model structure can be seen on Fig 5.5.

The model has six free variables:

- Temporal parameters: the second order filter consists of two first order (low-pass) filters. The two time constants of these filters are the first two parameters (τ_1, τ_2). They are subtracted from each other, the third parameter is the ratio of the weights (r_t).

- Spatial parameters: width of the two Gaussian blurrings, which is the σ parameter of the Gaussian (σ_1, σ_2). The ratio of the weights of the two Gaussians (they are subtracted from each other) ($\frac{g_1}{g_2}$ on Fig 5.5).

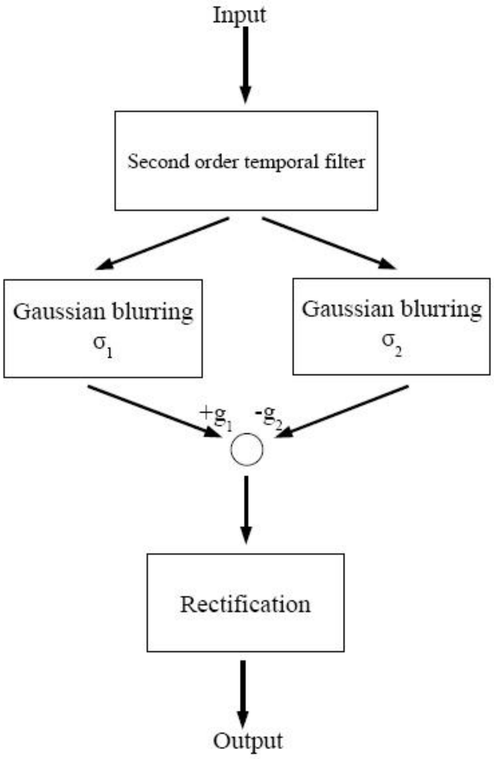


Figure 5.5: Structure of the filter model. The model consists of a second order temporal filter, a difference of Gaussian calculations and a rectification.

After an initial setting the model parameters are optimized using a gradient method to minimize the mean square error of the measurements.

5.4.3 Comparison

The neuromorphic model has the advantage that it produces not only the ganglion cell output but the intermediate cell outputs as well. Due to the neuromorphic model construction it allows us to make a good estimation of the processing of cells which were not measured directly. The neuromorphic model enables the investigation of the role of different neural elements, because we can clearly see the function of the cells in the whole network not only the final output. We can see how a given neural element processes its input, and we can shut off chosen elements in order to analyze their effect on the final response.

On the other hand the flashed square (when the stimulus is a square flashed for 1 sec.) measurements did not have enough information to tune all the parameters of the neuromorphic models, it allowed multiple solutions. This was due to the existence of several inhibitory mechanisms (horizontal, various amacrine cells) and the existence of nonlinear synapses, the effect of which is only partially revealed in the ganglion cell flashed square stimulus measurements.

The black-box model can easily fit the flashed-square measurements, which were shown quantitatively. The normalized mean square error (MSE) for the outputs of the models was calculated. This was computed the following way: the square of the difference of the model and the measured ganglion output was taken. Then I summarized it over the raster points of the flashed square response and divided it with the square of the measurements (5.7). One axis of the raster is space, the other is time and it shows the time development of a central section.

$$NMSE = \frac{\sum_{i,j}(M(i,j) - Y(i,j))^2}{\sum_{i,j} M(i,j)^2}, \tag{5.7}$$

where M is the measured output, Y is the model output and i, j are the spatial and temporal coordinates.

The output optimal gain, temporal delay and spatial offset were computed for both models to minimize the MSE term. Hence the calculated error represented difference in the shape of the responses; scaling, delay differences were abolished. The MSE of the models for some channels of [15] can be seen on Table 5.5.

On the other hand more complicated stimuli or measurements of internal retinal cells (bipolar, cone, inhibitory cells) can clear the effect of the previously mentioned internal parameters. In this case the black-box model consisting of linear spatial and temporal filters followed by a rectification can not fit the measurements. Such an input-output model has to be extended with other processing elements for obtaining internal cells outputs.

cell type	CNN model	simple filter model
on beta	6.72%	4.89%
on parasol	11.08%	11.80%
on delta	28.15%	25.14%
on bistratified	21.51%	21.00%
off parasol	25.35%	23.31%
off delta	21.25%	12.68%
LED	35.27%	33.41%

Table 5.5: Model output mean square error for the flashed square stimulus for the neuromorphic CNN and the filter model. All the responses were normalized to the stimulus.

.

The current black-box can violate even the ganglion output measurements. This occurs when we need the existence of inner retinal nonlinearities. For example if we need a model with (the outer retina's) spatio-temporal high-pass component enhancement then rectification, and this rectification should be followed by a blurring. Such a measurement can be seen on Figure 5.6.

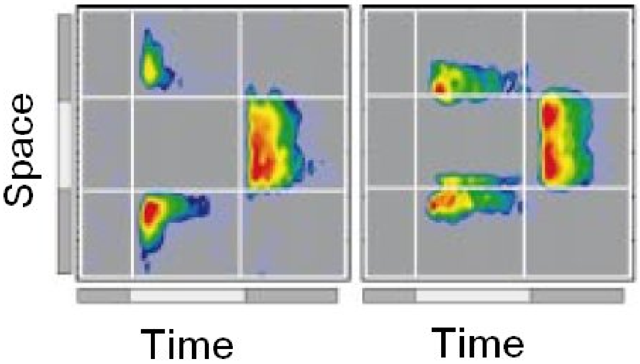


Figure 5.6: Spiking patterns (output) of two ganglion cells (adopted from [15]). White bars show the spatial and temporal locations of the stimulus. The measurement in the right column shows inner-retinal nonlinear behavior.

The ganglion output measurements on Figure 5.6 can not be achieved without a model where the rectification is followed by a spatial blurring. The stimulus onset and offset activities occupy complementary spatial regions, which is due to the effect of the spatial and temporal high-pass filtering. Due to the temporal high-pass filter, the transmitted signal will be the mirrored signal after the light offset (plus some scaling). This means that the final output should be also mirrored, and spatially complementary after the light offset. However we can see a little overlap in the regions, which is only possible if we rectify the output of the linear filter and then blur it. Hence the black-box filter model needs extension with further processing elements. This would result in a structure, where we have spatio-temporal linear filters (corresponding to the OPL) and then a rectification and a blurring (corresponding to the ganglion processing). This would be a significant step toward a model with all the retinal layers, even if they were only rough approximations.

When the effect of all the retinal cells is measured, then a suitable "black-box" model can end up in a model of similar complexity than a neuromorphic model, which contained the necessary elements from the beginning.

In the following section we show some measurements that support the existence of inner-retinal nonlinearities, which can be modeled only in our neuromorphic model. I also show the usefulness of the neuromorphic model in analyzing the role of the retinal cells in obtaining the final retinal output. Even if the final output is approximately linear for some channels, which might be modeled with a black-box model, the neuromorphic model is a useful tool for analyzing the behavior of the retinal cells.

It is obvious that for the task of retinal analysis a neuromorphic model is better because it enables the observation of the different layers and their effects. Between the filter model elements and the CNN layers the latter was chosen. This might cause longer simulation times, but my retina analysis task was less focused on the execution speed. My task was the accurate modelling of the retina and its analysis.

With the CNN-UM model I had the advantage, that earlier models and modelling tools were already established ([14]). This retina simulator used the performance of the underlying Aladdin simulator, where the simulation of the differential equations is optimized. This means beside the applied approximative formulas (first order Euler and Runge-Kutta), that the timestep increases if no changes occur in the input flow and the state variable (adaptive Euler). This reduces the simulation duration in case of image flows with long constant sections.

Furthermore I didn't have to make an equivalent model of the neuromorphic model from filter elements.

In case a lager number of simulation is needed to be executed, a possible future development is the merging of the advantages of the two models: 1, the differential equation formulas from the CNN models (Aladdin system), and 2, the convolution matrices from the filter model. The application of the latter can enhance the speed in the computation of blurring of large dendritic trees.

5.5 Measurements

In sections 5.2 and 5.3 the retinal model was described, in this section I discuss the measurements which support the existence of rectifying synapses and cross inhibition.

5.5.1 Rectifying Synapses

The main focus of this subsection is the relationship between the neural interactions and the effects of rectifying synapses. There are two stages where rectifying synapses were modeled (see Figure 5.1): the cone-bipolar and the bipolar-ganglion synapse. The newly included cone-bipolar rectification is discussed in section 5.2.2. Here the measurements supporting the bipolar-ganglion rectification are presented.

The temporal profiles on Figure 5.7 illustrate the non-linear rectifying properties of the synapse from bipolar to ganglion cells. We can also see the space-time plots of ganglion cell responses to a $600\mu m$ square flashed for 1 second. The bipolar responses (legend: "Bipolar") have both depolarizing (positive) and hyperpolarizing (negative) components at the location of the square. Bipolar cells excite ganglion cells. The effect of the bipolar cells on the ganglion cells is shown on the ganglion excitation measurements. Excitation to the ganglion cells (Ganglion), however is monophasic: it exists only where the bipolar cells were depolarized (positive). In regions of bipolar hyper-polarization, excitation to the ganglion cell is lost (Figure 5.7).

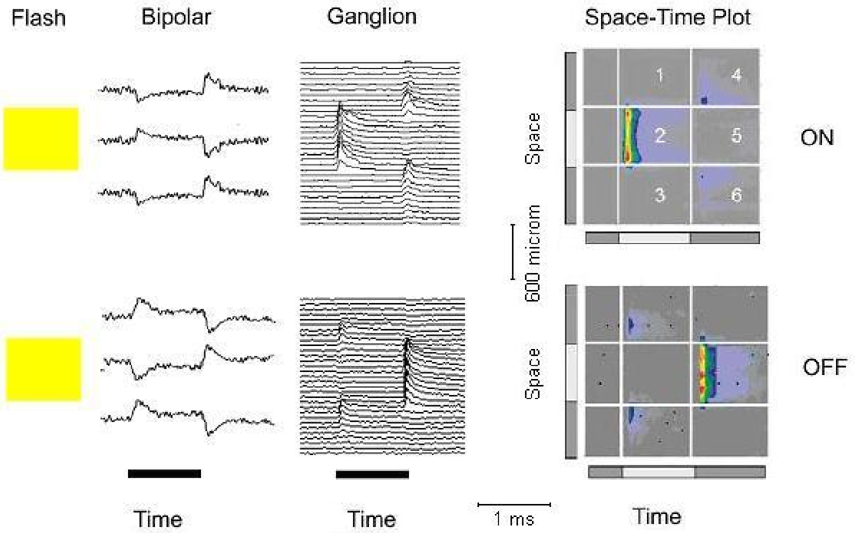


Figure 5.7: Time course and rectification of bipolar to ganglion cell activity. Black bars in the middle and left column, and white bars in the right column show the location of the stimulus. Upper row: ON activity, lower row: OFF activity. Left column: bipolar cell responses at center and regions surrounding the flashed square. Here, the center bipolar response is taken from patch recording made in a retinal slice in response to a 1 sec to a diffuse flash, showing transient depolarizing peak at light ON and transient hyperpolarizing peak at light OFF. Bipolar responses in regions surrounding the flash were not recorded, but are well approximated by the inverted versions of the center responses shown above and below the center responses in the Bipolar column. Center column: ganglion cell space-time patterns. Responses are rectified versions of the bipolar activity. The ON ganglion cell follows the upward (depolarizing) components of the bipolar activity but remains silent when the bipolar cells hyperpolarize in both the center and surround regions. Right column: color-coded space time maps of ganglion cell (ON Beta and OFF Parasol) excitation ([15]).

These recordings make it clear that the bipolar to ganglion cell synapse is strongly rectifying and that portions of the antagonistic surround, initially generated via feedback to cones and read out by the bipolar cells, are lost at the excitatory input to the ganglion cells. Rectification can also be seen to eliminate the after-hyperpolarization associated with the offset of the sustained stimulus. Thus, rectification eliminates the hyperpolarizing components associated with edges in both space and time.

5.5.2 Cross-Inhibition

Crossover inhibition acts to compensate for the degrading effects of rectifying nonlinearities. In such crossover pathways, ON activity is used to inhibit the OFF system, and OFF activity is used to inhibit the ON system. In this interaction the excitation and inhibition are out-of-phase. Thus the subtracted inhibition is out-of-phase with the excitation and reinforces it.

Figure 5.1 shows the locations of cross-inhibition:

1. bipolar cells: bipolar cells receive excitation from the cones. They receive inhibition from the bipolar cells of the opposite system via amacrine cells. This is a bipolar-bipolar interconnection and therefore it is called feed-back cross-inhibition.

2. ganglion cells: ganglion cells receive excitation from bipolar cells of the same system and inhibition from the other system's amacrine cells. This is called feed-forward cross-inhibition.

Figure 5.8 a.) shows recordings from ganglion cells receiving cross-inhibition ([15]). Similar measurements were made on bipolar cells (Figure 5.8 b.)) [54].

Figure 5.8 shows that the inhibitory input to ganglion cells falls at regions where the original antagonistic surround generated by horizontal cells would have been. For the ON cells, the antagonistic surround is derived from the OFF bipolar cells (via amacrine cell interneurons). The OFF bipolar cells depolarize in exactly the regions of the antagonistic surround for the ON bipolar cells. The two inputs, excitation and inhibition, are out-of-phase, and reinforcing, reconstructing the original signal as shown in section 5.5.3.

5.5.3 Retinal Signals

Figure 5.9 gives a brief sketch about the likely function of cross-inhibition and how this can counteract the information loss due to nonlinearities.

If a square stimulus is flashed, a response is evoked in the cones belonging to the ON and OFF systems. The horizontal cells perform spatial low-pass filtering and subtract it from the cones, so the OPL's output (the cone's output) is the high-pass enhanced signal (see Figure 5.9 "bipolar dendrites", and for the

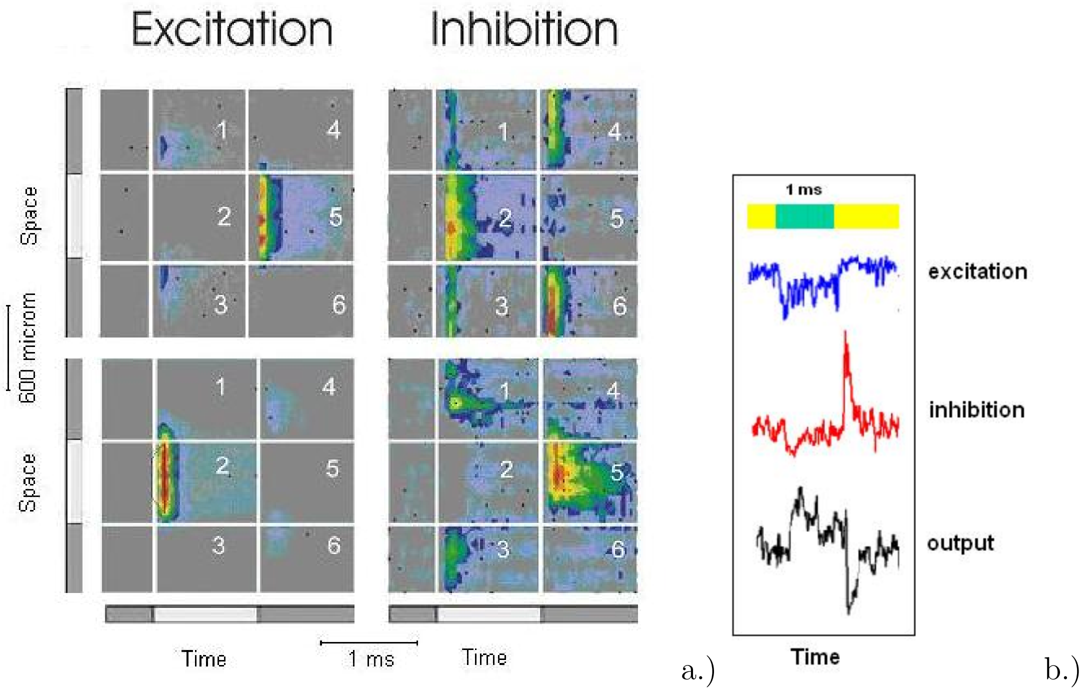


Figure 5.8: Excitation and cross inhibition. a.) Pairs of excitation and inhibition responses for 2 different ganglion cell types [15]. Horizontal and vertical white bars show the locations of the stimulus. Upper row: for the OFF cell, excitation falls mostly in regions 1, 3 and 5 while inhibition falls in regions 2, 4 and 6. Lower row: for the ON cell excitation falls in region 2 while inhibition falls in regions 1, 3 and 5. b.) Excitation, inhibition and output voltage recordings for bipolar cell [54]. The excitation curve is inverted (-excitation) because of the nature of the measurements. We can see that excitation and inhibition are not overlapping, and their interaction reconstructs the biphasic output.

underlying elements: Figure 5.1). This is the input of the bipolar cells as can be seen in Figure 5.9 ("bipolar dendrites"). The ON and OFF system responses are mirror images of each other. The OFF system shows a positive response outside the rectangle frame, which is the surround response. After rectification both channels loose information (negative values). These responses are restored at the ganglion level by cross inhibition, which carries the eliminated signal. Here the excitation (ON) and the inhibition (OFF) are recombined. The same is true with opposite excitation and inhibition for the OFF system (not shown). In section 5.6 this phenomenon is investigated quantitatively in detail. The ganglion re-

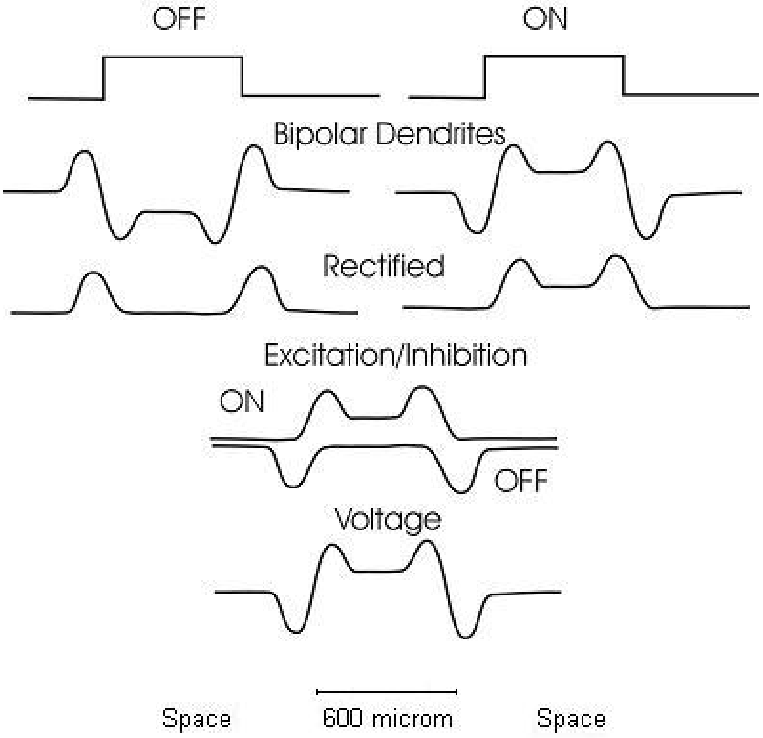


Figure 5.9: Reconstruction of the ON signal through crossover inhibition. Sketches of the spatial pattern of the ON and OFF signals at the peak of the light response to a bright flash at different sites of the retina. Only the depolarizing regions of the signal (rectified) are expressed in the ganglion inputs (excitation/inhibition). The row labeled Excitation/Inhibition shows the merger of excitation from the ON pathway and inhibition from the OFF pathway. These two signals reconstruct the original ON signal shown in the upper right.

sponse is rectified again at the stage where the output voltage is converted to spiking. This raises the following question: what is the function of an inhibition that seemingly does not overlap with the excitation? It does not modify positive values driven by excitation but regenerates negative values in the output voltage, which are simply eliminated by rectification at the next stage.

Section 5.6 presents an explanation for the function of this cross-over inhibition by showing the connection between rectification, cross-inhibition and dendritic-tree spatial blurring. This question is also addressed in section 5.7 together with the benefits of transmitting linear signals instead of nonlinearly distorted ones.

5.6 Simulation Results

In the simulations the signal processing of the retina was investigated. The model was tuned to fit the ganglion and bipolar measurements and the other internal elements of the model were adjusted according to morphological data (section 5.3). Hence the signals in the different model elements provide a reasonable approximation of the real biological signals in the retina.

My first aim was to observe how the input signal is processed by the different retinal cells. Then I examined how the output signal (ganglion voltage) changes if the cross inhibition or rectification is abolished. The spatial (section 5.6.1) and the temporal (section 5.6.2) profiles of the responses are also shown.

5.6.1 Spatial Profiles

In most of the simulations the same stimulus as in [14, 15] was used: a 600 micrometer square flashed for 1 second (see Figure 5.10). In Figure 5.10 the membrane voltage of the bipolar cells can be seen. The cone output (not shown) is a qualitatively similar signal to the OFF bipolar signal. Then at the cone-bipolar synapse, the ON bipolar signal is reversed, the OFF bipolar keeps its sign (see Section 2.2.2 and [73]). After this at both pathways the negative values are cut off by the cone-bipolar rectification. However the final bipolar output is a biphasic signal because the amacrine cross feed-back restores the signal loss due to the cone-bipolar rectification. This recombination is not detailed here because it occurs in the same way as the later discussed cross-inhibition at the ganglion level.

The bipolar output is an edge-enhanced signal. Negative values outside the edges and high positive values inside them (ON bipolar) are the result of the OPL's spatio-temporal high-pass component enhancement (Figure 5.9). This process enhances the edges and sharpens the image, which is exposed to blurring at later processing stages. The OFF bipolar cells have an opposite sign, which means they have negative values at the site of the flash.

After the bipolar-ganglion rectification the negative values are lost in both systems. The ON system's lost negative values along the edges are expressed in the positive values of the OFF system and vice versa (Figure 5.10). Loss of either system would mean some signal would be lost without the possibility of a restoration.

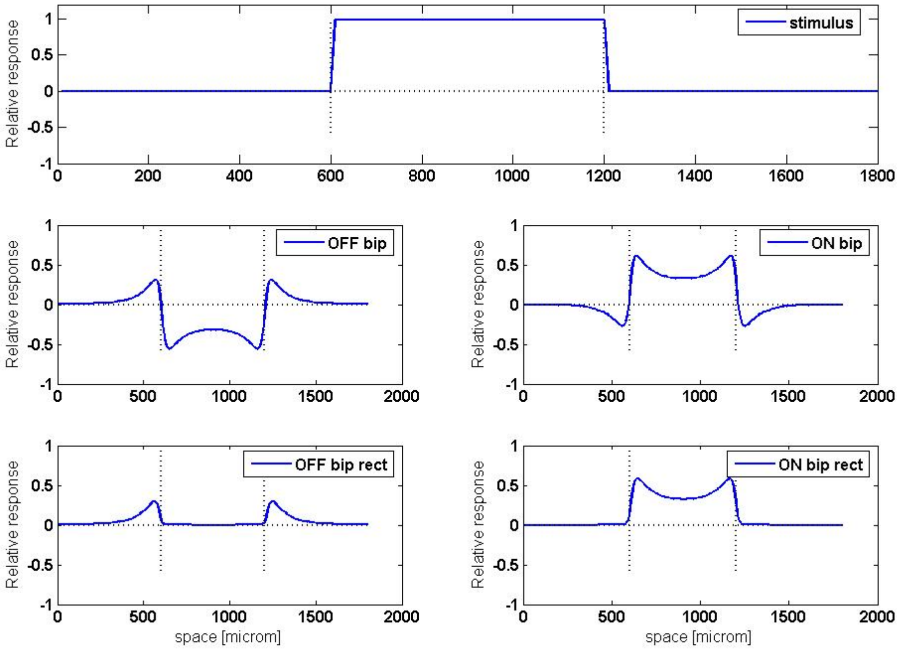


Figure 5.10: Simulated rectifying responses. The flashed square stimulus was used. The spatial profiles of the responses and the stimulus 200ms after the light onset are shown. Legends on each plot show the simulated cell. Top trace shows the stimulus. Vertical dashed gray lines show the frames of the flashed stimulus. Vertical axis shows the relative response with respect to the maximal response. Left column: top: OFF bipolar activity; Bottom OFF bipolar activity rectified as it enters the OFF ganglion cell. Right column: top: ON bipolar activity; Bottom ON bipolar activity rectified as it enters the ON ganglion cell.

At the ganglion layer this rectified signal is altered by the ganglion dendritic tree blurring (Figure 5.11). The excitation to the ganglion cells comes from the bipolar cells. It is summed up over the region covered by the ganglion cells dendrites and thereby blurs the signal. Inhibition comes from the opposite class of bipolar cells to ON from OFF, or to OFF from ON, and passes through amacrine cells. Inhibition is also blurred by the dendrites of the ganglion cells (see Figure 5.11 and 5.12).

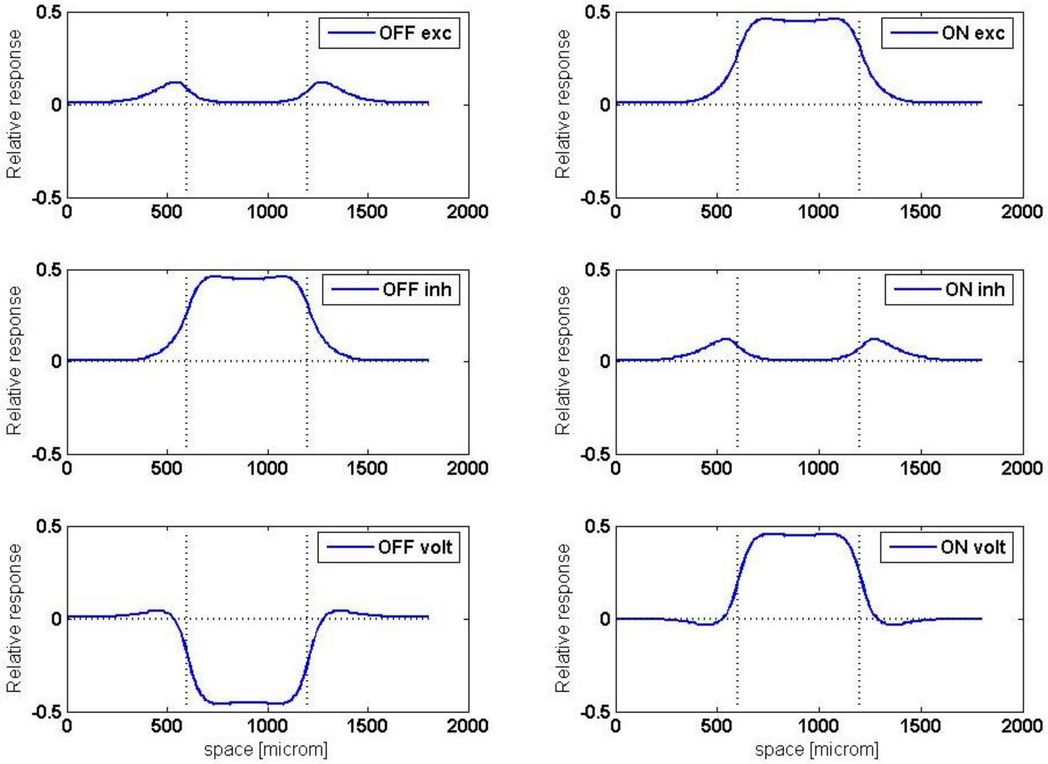


Figure 5.11: Spatial profiles of the responses 200ms after the light onset. Reconstruction of the spatial dimensions of the representation of the stimulus through integration of the rectified excitation and inhibition. Vertical axis shows the relative response with respect to the maximal response. Top row excitation represents the stimulus but spreads beyond the stimulus dimensions. Middle row: inhibition represents the stimulus, but spreads beyond the dimension of the stimulus (dashed vertical lines). Bottom row: membrane voltage, the integration of excitation and inhibition, this is more constrained to the dimensions of the stimulus.

The effect of bipolar rectification and the interaction between ganglion excitation and inhibition is illustrated with modeled responses in Figure 5.11. The rectified ON bipolar excitation responds in the interior of the square, but is diffused by the ganglion cell dendrites (top row right column), the OFF system represents the exterior of the square but spreads into the interior of the stimulus (vertical dashed lines) (left column). The model shows how the ganglion response is reconstructed from the summation of excitation and the inhibition. In the ON output voltage pattern, inhibition restricts the spatial spread of the ganglion cell representation so that it is confined to the dimensions of the original stimulus. In the opposite OFF system due to the inhibition the activity is driven to the outside of the stimulus location and the stimulus frames are kept.

To emphasize the importance of cross-inhibition we can compare the ganglion excitation and the ganglion voltage (output) signals. If the ganglion cells did not receive inhibition, ganglion excitation would be the only source of voltage output of the cell. In this case the resulting output pattern would be more blurred and less tightly confined to the extent of the stimulus.

To make the recombination of ON and OFF signals through cross inhibition more clear, the order of blurring and cross inhibition can be changed in the model (see Figure 5.12). This transformation assumes that the (CNN and the retinal) cells operate in the linear range of their sigmoid response curve. Under the model's conditions both the cross inhibition and the blurring are simulated as linear operators, so changing the order of operations is possible and the output voltage remains the same (5.8).

$$D(Exc) - D(Inh) = D(Exc - Inh), \tag{5.8}$$

where D is the diffusion operator, Exc is the excitation and Inh is the inhibition.

If the rectified signals were recombined before blurring (Figure 5.12) then the bipolar cell's original biphasic signal would be reconstructed (Figure 5.10). For the ON system this can be seen in (5.9) for the OFF in (5.10).

$$B \overset{rectification}{\longrightarrow} B_+ \overset{OFFcrossinh.}{\longrightarrow} B_+ - (-B_-) = B \tag{5.9}$$

$$-B \overset{rectification}{\longrightarrow} -B_- \overset{ONcrossinh.}{\longrightarrow} -B_- - (B_+) = -B, \tag{5.10}$$

where B is the bipolar cell's non-rectified linear signal, B_+ is the rectified (positive) part of B, and B_- is the negative, hence $B_+ + B_- = B$.

This linear signal has enhanced edges and the blurring affects it less. In the case of ON ganglion cells the positive values of the square's interior can spread less because of the negative values outside the edges, and vice versa. The same applies to the OFF system with sign inversion. In the actual model the recombination is done after the blurring. Because of the reversible operations (5.8) this has an equivalent effect: it restores the high-pass component enhancement of the OPL.

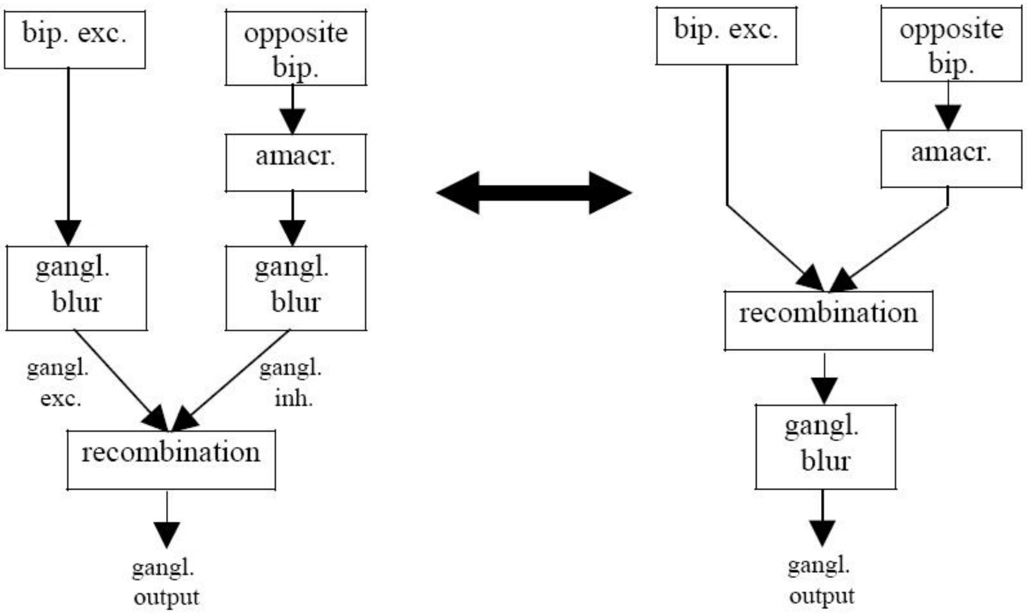


Figure 5.12: Recombination before or after blurring: the recombination of the excitation (ON or OFF) and cross inhibitory (opposite system) signals at the ganglion level. The left side shows the neuromorphic computation of the voltage, the right side shows the change of the recombination and the ganglion blurring.

5.6.2 Temporal Profiles

Temporal curves show similar effect of the OPL's high-pass component enhancement, rectification, blurring and recombination as in the case of the spatial plots.

Figure 5.13 and 5.14 shows the effect of rectification and out-of-phase inhibition in the temporal domain. We can see that the bipolar cell response is biphasic (Figure 5.13 ON/OFF bip) because of the outer retina's temporal high-pass component enhancement. This is due to the cone temporal second order dynamics and horizontal cell feed-back. Cone-bipolar nonlinearity was eliminated by the amacrine feed-back originating from the opposite system analogous to the spatial

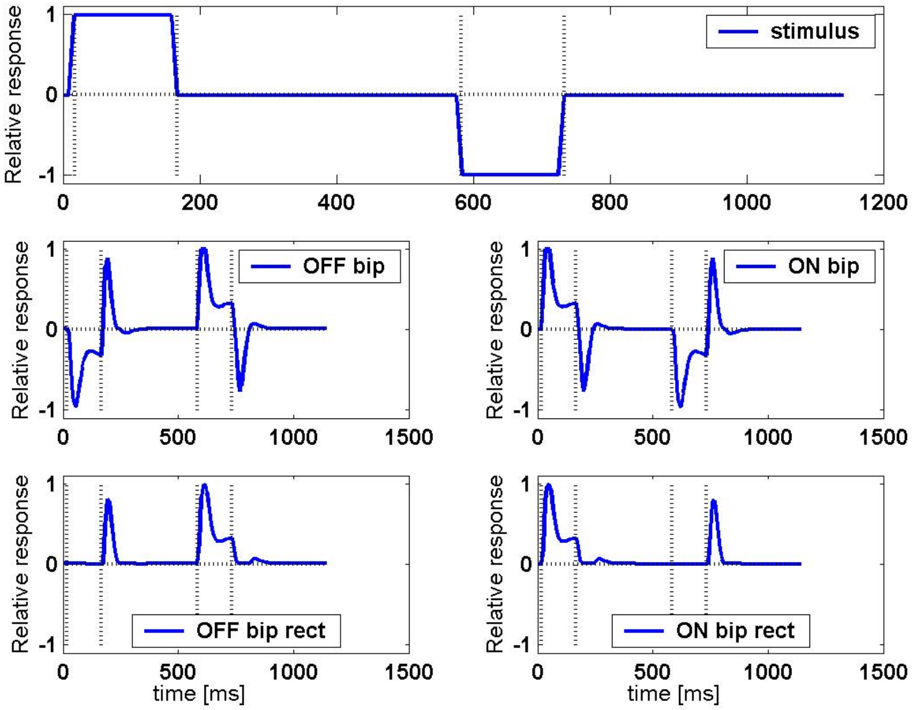


Figure 5.13: Time course of bipolar cell response to a stepped increment and decrement in intensity. The traces show the time evolution of the cells at the center of the flashed square. Top row: stimulus, middle row: bipolar cell input, bottom row: rectified bipolar cell output.

domain (section 5.6.1). ON cells respond positively to the light onset and negatively to the offset. OFF cells behave in the opposite (sign inversed) way (Figure 5.13). After rectification the biphasic nature of the responses is abolished (Figure 5.13 ON/OFF rect bip) and ganglion excitation spreads out in time (Figure 5.14 ON/OFF exc). Inhibition from the opposite system restores the biphasic response and response duration is restricted.

Figures 5.11 and 5.14 illustrate how crossover inhibition sharpens the representation of the images in both space and time. Both effects were based on the following phenomena:

1. The OPL's high-pass component enhancement introduced negative values even for my purely positive stimulus.

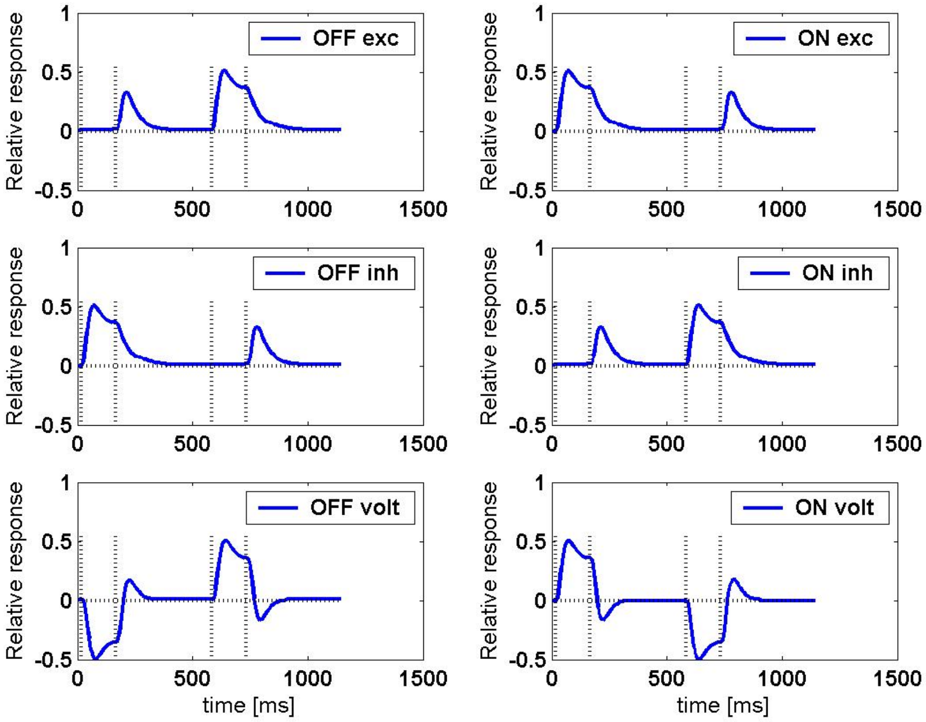


Figure 5.14: Integration of excitation and inhibition in the time course of the ganglion cell response. The traces show the time evolution of the cells at the center of the flashed square. Top row: excitation to the ganglion cell spreads out in time. Middle row: inhibition to the ganglion cell falls in complementary regions and is also diffused in time. Bottom row: integration of excitation and inhibition results in a more discrete, temporally less diffused response in the ganglion cells.

2. The cone-bipolar and the bipolar-ganglion synapses eliminated these negative portions of the OPL/bipolar signals.

3. Ganglion excitation deprived from the negative components blurs both in space and time.

4. The eliminated parts could be recovered by the effect of the cross-inhibition, which feeds the missing information from the opposite system.

In the next section the role of the cross-inhibition is discussed by comparing the normal output with blocked inhibition outputs.

5.7 Role of the Cross-Inhibition

The crossover inhibition described above serves a number of functions. The illustrations above indicate that cross inhibition provides the component of the signal which had been lost through rectification. Therefore it appears to sharpen the representation of the stimulus and to re-linearize the overall transfer function. Similar push pull mechanisms have been described in retina and at higher visual centers in earlier studies [60].

Figure 5.15 shows again that cross inhibition restricts the signal to its original frame. It also illustrates that this inhibited output is the same as in the case of a linear system without rectification and cross-inhibition. As seen in section 5.6.1 cross inhibition restores the non-rectified bipolar signal (see (5.9) and (5.10)). Cross-inhibition corrects the distortion caused by nonlinearity. The resulting signal is more accurate and keeps the frames of the stimulus.

To be more coherent with the retinal processes we have to state the following: under real retinal conditions this recombination is not that clear, because nonlinearities are not exact rectifications and there are noise sources. Though cross-inhibition shapes the signal toward its non-distorted linear form.

Earlier studies [57] showed that inhibition acts to restrict the response to the stimulus boundaries. However the self feed-back inhibition (either in the OPL or by amacrine cells) is not always sufficient to maintain this restriction. Using measurements I showed that their beneficial spatio-temporal high-pass component enhancement is destroyed by rectification. Cross-over inhibition counteracts this degradation and improves the accuracy of the representation of the visual stimuli by the ganglion cells.

Figure 5.15 gives another hint as to why the non-rectified linear signal is more resistant to blurring. In case of the linear/recombined signal the negative values introduced by the OPL's high-pass function are kept. Hence the edge enhancement is not destroyed. To illustrate the importance of the OPL's high pass component enhancement we may observe the linear system's response and the response without the OPL (Figure 5.15 curves: ”linear system” and ”no OPL”). They both lack cross-inhibition and rectification. They have a difference: the model of the latter does not contain the OPL's high pass component enhancement.

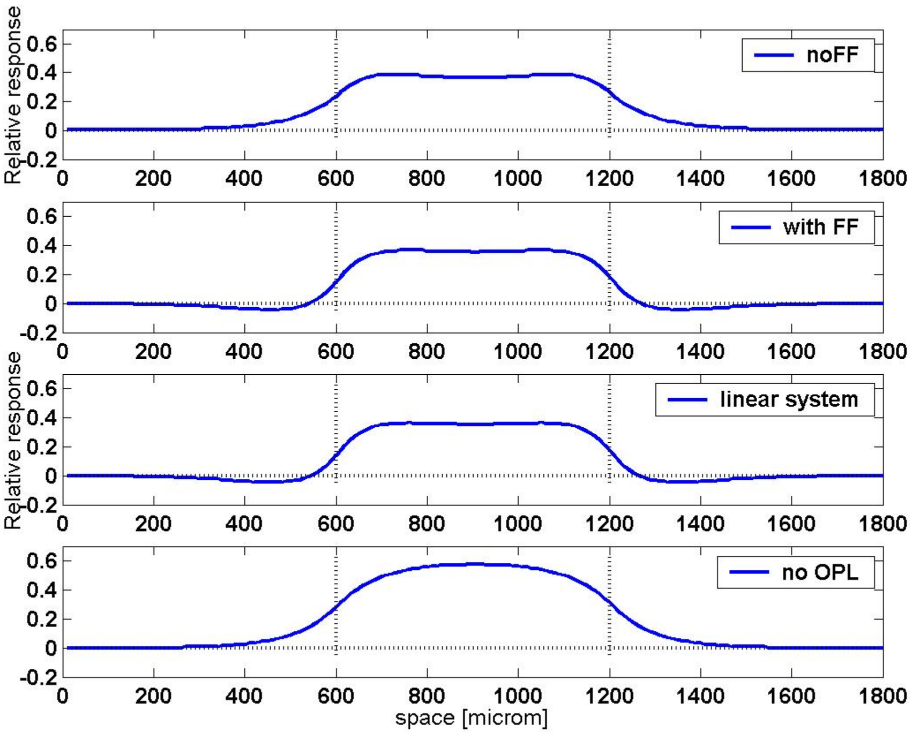


Figure 5.15: Spatial profiles of ON ganglion responses. Top row: the output of the model with rectification but without cross-inhibition 200ms after light onset. Rectified version of the original pattern spreads out across the boundaries of the stimulus. Second row: reconstructed pattern that includes inhibition is constrained to lie within the boundaries of the stimulus. Third row: the model's output without rectification and cross inhibition (linear system). This is the original pattern that would have been generated if the input excitation had not been rectified. It illustrates that the reconstruction compensates for the distortion introduced by rectification. Lower row: the output of the system without rectification and cross inhibition, the OPL's high pass component enhancement was also abolished.

We can see that the signal dramatically spreads out, showing the great importance of the OPL's enhancement. Figure 5.15 shows that without the OPL's high-pass component enhancement even a linear system is blurred. Having the OPL and rectification this enhancement is reduced and the output is blurred (Figure 5.15 ”no FF”), and the cross inhibition reconstructs the OPL enhanced signal's results (Figure 5.15 ”with FF” and ”linear system”).

5.7.1 Frequency Components

Figure 5.16 illustrates the role of crossover inhibition in re-linearizing signals in another way. The input stimulus (blue) is a fast sine wave (in time) amplitude-modulated by a slow sine-wave. The response is shown in green. Because of the rectifying nonlinearity, excitation (purple) and inhibition (black) are distorted such that their mean value increases with the increase in amplitude of the fast sine wave, even though the mean value of the input signal remains at zero. When excitation and inhibition are re-combined in the reconstructed signal, it once again becomes a modulated sine wave with zero mean.

The effect of rectification and recombination can be understood from the frequency response of the signals in Figure 5.16. The rectified elements of excitation and inhibition shown in Figure 5.17 as dashed red and solid black, show a strong low frequency component. This component arises because of the non-zero mean value that was introduced by rectification. This low-frequency component was not present in the original stimulus, it is the component caused by the rectification.

When excitation and inhibition are once again combined into the final voltage response, low frequency terms are removed, as shown in Figure 5.17. They are removed because they are in phase (see Figure 5.16) and subtraction removes them. High frequency components (representing the input stimulus) are enhanced because they are in the opposite phase (see Figure 5.16) and reinforce each other.

5.8 Conclusions

Synapses are inherently rectifying because of the exponential nature of the calcium activation curve that underlies transmitter release. In spiking systems this does not interfere with the magnitude of the signals because although rectification may degrade the magnitude of the individual spiking events, signal magnitude is coded in spike frequency, not magnitude. But at graded potential synapses, rectification can dramatically affect the form of transmission because rectification directly affects the transmission of signal magnitude. At the photoreceptor to bipolar synapse rectification exists, and distorts transmission. As a result, bipolar cells retain some of the antagonistic surround established by feedback from

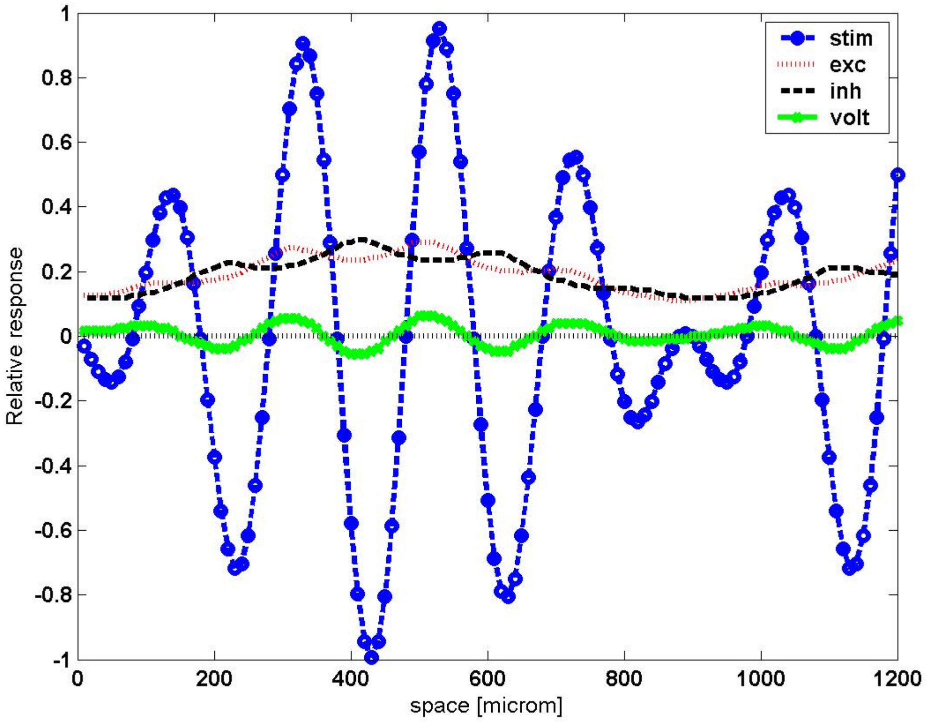


Figure 5.16: Reconstruction of the rectified excitation and inhibition generated by a sine wave-modulated sine wave results in a linear output. Vertical axis shows the relative response with respect to the maximal response. Stimulus: dashed line with circles, Excitation: dotted line. Inhibition: dashed line (without circles). Output voltage: solid line. The excitation and inhibition are distorted by rectification, creating a non-zero mean in those signals. When excitation and inhibition are re-combined, the resulting voltage output, green, is once again sinusoidal with zero mean.

horizontal cells to cones. A form of crossover inhibition restores the rest of the antagonistic surround there. Then, at the bipolar-to-ganglion cell synapse the signal is rectified again: the negative parts of the signals originating in the outer retina are lost from the excitatory input to the ganglion cells (see Figures 5.7-5.8 for the physiological results). For the ON pathway, the signals arriving at the ganglion cell, representing the stimulus, now devoid of its antagonistic surround, are subject to diffusion across the dendrites of the ganglion cell as shown in Figs 5.11 and 5.14.

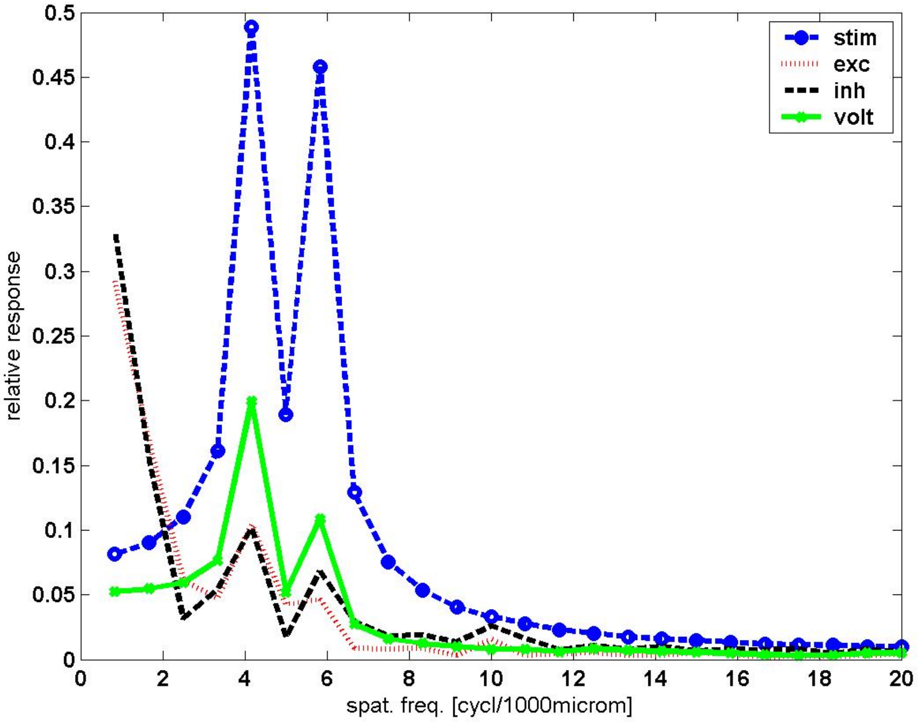


Figure 5.17: Spatial frequency components of the signals shown in Figure 5.16. Stimulus: dashed line with circles, Excitation: dotted line. Inhibition: dashed line (without circles). Output voltage: solid line. Excitation, inhibition and the voltage values are 5 times enlarged for the better visibility. Here excitation and inhibition acquire a strong low frequency component. This is the non-zero mean generated by rectification. This low frequency component is eliminated when the excitation and inhibition are integrated into the final voltage response.

I extended the already existing model with the rectification and cross inhibitory interconnections. I showed that the OPL's high-pass component enhancement reduces the signal blurring caused by ganglion cell dendrites. I also showed that this enhancement and reduction of blurring is damaged because of internal nonlinearities. Finally I showed that cross-inhibition reduces these distorting effects.

These results are based on a series of biological measurements. These prove the existence of cross-inhibition, rectification and ganglion blurring. My simulations showed the compensating interaction between these phenomenons. These are predictive simulations, the effects of the interactions are totally proven only when measured in the retina. These effects are highly probable up to the previously described limits of the measurements.

The question arises, how much are these results in agreement with previous measurements. First we have to state again, that the existence of nonlinearities and cross inhibition is based on measurements:

- The existence of the cone-bipolar nonlinearity is based on modelling of bipolar excitation signals [54]. It is also in agreement with results in [51].
- The bipolar-ganglion nonlinearity is based on measurements from [15], and unpublished measurements shown in this dissertation. Measurements in [55] show also the existence of bipolar nonlinearity. Although it shows it for Y bipolars.
- Cross inhibition (or out-of-phase) is a well known type of inhibition [54, 65, 60, 79, 15]

The fact fact that inhibition reduces blurring in the retina, is already shown in [57]. Here measurements with pharmaceutical blockers show, that ganglion response spreads out in space and time, if the inhibition is blocked. The special type of inhibition, the cross-inhibition is examined in [79], for thalamocortical interactions. Cross-inhibition is found to sharpen the response.

My dissertation added the simulation results that cross-inhibition might compensate for the effects of nonlinearities and thereby reduces blurring. Furthermore it is because of nonlinearities that cross-inhibition has a signal sharpening role. Future simulations should incorporate the effects of noise sources, which might show, that cross inhibition (as a different signal path) has a role in reducing the effects of noise sources.

Cross-over inhibition is necessary for maintaining the sharpness of visual representations in both space and time provided that: 1) the signals have been rectified, and 2) the ganglion cell dendrites act to diffuse the excitatory input in both space and time. This diffusion causes the excitatory and inhibitory signals, now both rectified, to interact. The intersection of this interaction between excitation and inhibition lies near the edge of the representation of the stimulus.

Finally it should be mentioned that cross inhibition has a parallelism in engineering. The application of differential circuits has the same benefits. A signal x is split into two different signals x and $-x$. The difference of the two carries the valid signal. Using the difference as information the effects of correlated noise sources and even order nonlinearities (e.g. rectification) on the two signals can be reduced or even eliminated. In case of ON and OFF channels, the separation and recombination of them plays the same role.

The results about the role of ON-OFF interactions were published in [2].

Chapter 6

Conclusions

In my two theses I solved two problems in the field of visual perception using CNN-UM.

The adaptive capturing algorithm is able to reduce the dynamic range of unequally illuminated scenes. With the restoration of the DC component it is able to maintain the illumination differences without causing saturation. The method compresses the dynamic range without the loss of local contrast and reversing the intensity relationships of the different regions. The anisotropic extension of the method avoids the introduction of artifacts around edges. Hence it performs similar results as cutting-edge dynamic range compression methods, but it combines sensing and compressing with the same algorithm.

The main possibility of progress in this field lies in the full hardware implementation of the project. The EYE-RIS chip is expected to give us an opportunity for both the implementation of spatially varying exposure and anisotropic diffusion. Another possibility is the extension of the method for color images. In case of color images we have to deal with chromatic distortions, and a simple scaling of the luminance channel does not always yield sufficient results.

My retina modeling works resulted in the clarification of a less understood form of inhibition. The effects of nonlinearities cannot be eliminated using self feed-back inhibition. These processes can well enhance the spatial and temporal edges, but the distorted information by the nonlinearities can be restored only using an inhibition originating from the opposite system. This analysis showed the power of quantitative modeling in the understanding of the retinal function.

There are several ways to improve the modeling. These improvements require new measurements, with more complex stimuli than the flashed square or with recordings from inner retinal cells. Such measurements can disclose the nature of nonlinearities in the retina and clarify the validity of multiple solutions. Under these solutions both different structures and different parameter settings of the same model are meant.

Summary

6.1 Summary of the main results

1. Thesis: *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) (6.1).

$$T_{n+1}(i,j) = \overline{T_n}(i,j)\frac{V_{max}/2}{\overline{V_n}(i,j)}, \tag{6.1}$$

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} are 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 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.

The methods adjust the integration time continuously, hence it adapts temporally to changed illumination conditions.

The adjustment of capturing parameter was chosen because in 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 created. 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 (3.16)). As a result an image is obtained, on which the human perceivable intensity differences are visible, but the image has no saturated areas.

As illumination often changes suddenly along shadow borders, anisotropic processes were applied for the computation of the diffusion. Here the integration time difference between neighboring pixels sets a constraint on the spread of diffusion. (3.19) 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.

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.

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.

2. Thesis: *Qualitative and quantitative analysis of the retinal func-*

tion 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-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 the 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.

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 amacrine cells.

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 .

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.

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 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, electrophysiologists 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.

References

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.

List of other references

[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.

[18] A. E. Gamal, “High dynamic range image sensors,” *Tutorial at International Solid-State Circuits Conference*, February 2002.

[19] M. D. Grossberg and S. K. Nayar, “High Dynamic Range from Multiple Images: Which Exposures to Combine?,” in *Proc. of ICCV Workshop on Color and Photometric Methods in Computer Vision (CPMCV)*, (Nice, France), October 2003.

[20] R. A. Norman and F. S. Werblin, “Control of retinal sensitivity: I. Light and dark adaptation of vertebrate rods and cones,” *The Journal of General Physiology*, vol. 63, pp. 37–61, January 1974.

[21] C. A. Mead and M. A. Mahowald, “A silicon model of early visual processing,” *Neural Networks*, vol. 1, pp. 91–97, 1988.

[22] L. Spillmann and J. S. Werner, *Visual Perception - The Neurophysiological Foundations*, ch. 5 / III. San Diego, California: Academic Press, Inc., 1990.

[23] J. E. Dowling, *Neurons and Networks, An Introduction to Neuroscience*, ch. Retinal Processing of Visual Information. Cambridge: The Belknap Press of Harvard University Press, 1992.

[24] V. Brajovic, "A model for reflectance perception in vision," in *Bioengineered and Bioinspired Systems, Proceedings of SPIE*, 2003.

[25] C. Torras, *Computer Vision: Theory and Industrial Applications*, ch. 5.2. Berlin Heidelberg: Springer Verlag, 1992.

[26] M. S. Landy and J. A. Movshon, *Computational Models of Visual Processing*, ch. 14. Chambridge, Massachusetts: MIT Press, 1991.

[27] G. Linan, R. Dominguez-Castro, S. Espejo, and A. Rodriguez-Vazquez, "Ace16k: A programmable focal plane vision processor with 128x128 resolution," in *15th IEEE European Conference on Circuit Theory and Design (ECCTD'01)*, (Helsinki), pp. 345–348, August 2001.

[28] "Dalstar 1m28-sa, cmos area scan cameras - camera user's manual." `http://www.ctc-g.co.jp/~hts/dalsa/pdf/1M28_DALSTAR-SA.pdf`.

[29] W. H. A. Beaudot, "Sensory coding in the vertebrate retina: towards an adaptive control of visual sensitivity," *Network: Computation in Neural Systems*, vol. 7, pp. 317–323, May 1996.

[30] "Image flows of the dynamic adjustment are available at the following sites." `http://digitus.itk.ppke.hu/~wagner/dynres.html`.

[31] D. Yang, A. E. Gamal, B. Fowler, and H. Tian, "A 640x512 CMOS image sensor with ultra-wide dynamic range floating-point pixel-level ADC," in *IEEE Journal of Solid State Circuits* (12, ed.), (December), 1821–1834, 1999 34.

[32] F. Guichard, L. Moisan, and J.-M. Morel, "A review of P.D.E. models in image processing and image analysis," *Journal de Physique IV*, vol. 12, pp. 137–154, 2002.

[33] E. Culurciello, R. Etienne-Cummings, and K. Boahen, "A biomorphic digital image sensor," *IEEE Journal of Solid-StateCircuits*, vol. 38, pp. 281– 294, February 2003.

[34] S. K. Nayar and T. Mitsunaga, "High Dynamic Range Imaging: Spatially Varying Pixel Exposures," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, (Hilton Head Island, South Carolina), pp. 472–479, June 2000.

[35] M. Brendel and T. Roska, "Adaptive image sensing and enhancement using cellular neural network universal machine," *International Journal of Circuit Theory and Applications*, vol. 30, pp. 287–312, mar-jun 2002.

[36] J. L. Dannemiller and B. R. Stephens, "Asymmetries in contrast polarity processing in young human infants," *Journal of Vision*, vol. 1, pp. 112–125, 2001.

[37] R. Street, "High dynamic range segmented pixel sensor array." U.S. Patent 5 789 737, 1998.

[38] K. Takahashi, T. Hieda, C. Satoh, T. Masui, T. Kobayashi, and K. Yoshimura, "Image sensing device with diverse storage times used in picture composition." U.S. Patent 5 638 118, 1995.

[39] B. Wandell, A. E. Gamal, and B. Girod, "Common principles of image acquisition systems and biological vision," *Proceeding of the IEEE*, vol. 90, pp. 5–17, January 2002.

[40] H. B. Barlow, "Optic nerve impulses and Weber's law," *Cold Spring. Harbor Symposia on Quantitative Biology*, vol. 30, pp. 539–546, 1965.

[41] X. C. Liu and A. E. Gamal, "Synthesis of high dynamic range motion blur free image from multiple captures," *IEEE Transactions on Circuits and Systems I: fundamental theory and applications*, vol. 50, no. 4, pp. 530–539, 2003.

[42] S. Pattanaik and H. Yee, "Adaptive gain control for high dynamic range image display," in *Proceedings of 18th Spring Conference in Computer Graphics (SCCG2002)*, pp. 83–87, ACM Press, 2002.

[43] P. E. Debevec and J. Malik, "Recovering high dynamic range radiance maps from photographs," in *SIGGRAPH 97*, August 1997.

[44] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," in *Proceedings of the 29th annual conference on Computer graphics and interactive techniques.*, pp. 249–256, ACM Press, 2002.

[45] T. Jyrinki, "Seminar on computer graphics: High dynamic range images." `http://www.tml.tkk.fi/Opinnot/Tik-111.500/2003/paperit/TimoJyrinki.pdf`, Spring 2003.

[46] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," in *Proceedings of the 29th annual conference on Computer graphics and interactive techniques.*, pp. 267–276, ACM Press, 2002.

[47] W. G. Larson, H. Rushmeier, and P. C., "A visibility matching tone reproduction operator for high dynamic range scenes.," *IEEE Transactions on Visualization and Computer Graphics*, vol. 3, no. 4, pp. 291–306, 1997.

[48] M. White, R. Zakia, and P. Lorenz, *The new zone system manual.* Morgan & Morgan Inc., 1984.

[49] S. N. Pattanaik, "High dynamic range imaging," in *Tutorial notes (speakers: Paul Debevec, Greg Ward, Erik Reinhard, S. N.Pattanaik), ACM-SIGGRAPH 2005*, (Los Angeles), August 2005.

---

[50] Á. Zarándy, C. Rekeczky, I. Szatmári, and P. Földesy, "The new framework of applications: The aladdin system," *IEEE Journal on Circuits, Systems and Computers*, vol. 12, no. 6, pp. 764–781, 2003.

[51] M. H. Hennig, K. Funke, and F. Wörgötter, "The influence of different retinal subcircuits on the nonlinearity of ganglion cell behavior," *The Journal of Neuroscience*, vol. 22, pp. 8726–8738, October 2002.

[52] R. L. Rockhill, F. J. Daly, M. A. MacNeil, S. P. Brown, and R. H. Masland, "The diversity of ganglion cells in a mammalian retina," *The Journal of Neuroscience*, vol. 22, pp. 3831–3843, May 2002.

[53] M. A. MacNeil, J. K. Heussy, R. F. Dacheux, E. Raviola, and R. H. Masland, "The population of bipolar cells in the rabbit retina," *The Journal of Comparative Neurology*, vol. 472, pp. 73–86, 2004.

[54] A. C. Molnar and F. S. Werblin, "Amacrine-mediated feedback between bipolar cells enhances temporal diversity in the mammalian retina," in *Proceedings of The Association for Research in Vision and Ophthalmology 2005 Annual Meeting*, (Fort Lauderdale, Florida), May 1-5 2005.

[55] J. B. Demb, K. Zaghloul, L. Haarsma, and P. Sterling, "Bipolar cells contribute to nonlinear spatial summation in the brisk-transient (y) ganglion cell in mammalian retina," *The Journal of Neuroscience*, vol. 21, pp. 7447–7454, October 2001.

[56] C. Rekeczky, B. Roska, E. Nemeth, and F. S. Werblin, "The network behind spatio-temporal patterns: building low-complexity retinal models in CNN based on morphology, pharmacology and physiology," *International Journal of Circuit Theory and Applications*, vol. 29, pp. 197–239, 2001.

[57] B. Roska, E. Nemeth, L. Orzo, and F. S. Werblin, "Three levels of lateral inhibition: A space-time study of the retina of the tiger salamander," *The Journal of Neuroscience*, vol. 20, pp. 1941–1951, March 1 2000.

[58] M. S. Arkin and R. F. Miller, “Bipolar origin of synaptic inputs to sustained off-ganglion cells in the mudpuppy retina,” *Journal of Neurophysiology*, vol. 60, pp. 1122–1142, 1988.

[59] M. S. Arkin and R. F. Miller, “Synaptic inputs and morphology of sustained on-ganglion cells in the mudpuppy retina.,” *Journal of Neurophysiology*, vol. 60, pp. 1143–1159, 1988.

[60] J. H. Belgum, D. R. Dvorak, J. S. McReynolds, and E. Miyachi, “Push-pull effect of surround illumination on excitatory and inhibitory inputs to mudpuppy retinal ganglion cells.,” *The Journal of Physiology*, vol. 388, pp. 233–243, 1987.

[61] G. D. Field and F. Rieke, “Nonlinear signal transfer from mouse rods to bipolar cells and implications for visual sensitivity.,” *Neuron*, vol. 34, pp. 773–785, 2002.

[62] J. A. Hirsch, “Synaptic physiology and receptive field structure in the early visual pathway of the cat,” *Cerebral Cortex*, vol. 13, pp. 63–69, January 2003.

[63] M. A. MacNeil, J. K. Heussy, R. F. Dacheux, E. Raviola, and R. H. Masland, “The shapes and numbers of amacrine cells: matching of photofilled with golgi-stained cells in the rabbit retina and comparison with other mammalian species,” *Journal of Computational Neurology*, vol. 413, pp. 305–326, 1999.

[64] M. A. MacNeil and R. H. Masland, “Extreme diversity among amacrine cells: implications for functionthe ganglion cells,” *Neuron*, vol. 20, pp. 971–982, 1998.

[65] B. Roska, A. Molnar, and F. S. Werblin, “Parallel processing in retinal ganglion cells: how integration of space-time patterns of excitation and inhibition form the spiking output,” *Journal of Neurophysiology*, vol. 95, pp. 3810–3822, March 1 2006.

[66] F. S. Werblin and J. Dowling, "Organization of the retina of the mudpuppy, necturus maculosus. II. intracellular recording," *Journal of Neurophysiology*, vol. 32, pp. 339–355, 1969.

[67] F. S. Werblin, T. Roska, and L. O. Chua, "The analogic cellular neural network as a bionic eye," *International Journal of Circuit Theory and Applications*, vol. 23, pp. 541–569, 1995.

[68] E. R. Kandel, J. H. Schwartz, and T. M. Jessel, *Principles of Neural Science*, ch. 26, p. 516. McGraw-Hill, fourth edition ed., 2000.

[69] D. Bálya and T. Roska, "Retina model with real time implementation," in *Proceedings of IEEE international symposium on circuits and systems (ISCAS 2005)*, (Kobe), 2005.

[70] D. Bálya, I. Petrás, T. Roska, R. Carmona, and A. Rodríguez-Vázquez, "Implementing the multi-layer retinal model on the complex-cell CNN-UM chip prototype," *International Journal of Bifurcation and Chaos*, vol. 14, no. 2, pp. 427–451, 2004.

[71] M. M. Slaughter and R. F. Miller, "Characterization of an extended glutamate receptor of the on bipolar neuron in the vertebrate retina," *The Journal of Neuroscience*, vol. 5, pp. 224–233, January 1985.

[72] M. M. Slaughter and R. F. Miller, "An excitatory amino acid antagonist blocks cone input to sign-conserving second-order retinal neurons.," *Science*, vol. 219, pp. 1230–1232, 1983.

[73] "Cell types and neural pathways in the salamander retina (Mc Reynolds)." `http://www.physiology.med.umich.edu/mcreynolds/retina1.htm`.

[74] M. J. Zigmond, F. E. Bloom, S. C. Laudis, J. L. Roberts, and L. R. Squire, *Fundamental Neuroscience*, ch. 28. San Diego, California: Academic Press, 1999.

[75] "Ocular anatomy lecture - retina." `http://eyelearn.med.utoronto.ca/Lectures05-06/Anatomy/08Retina.htm`.

[76] P. Wikovsky and A. Dearry, “Functional roles of dopamine in the vertebrate retina,” *Progress in Retinal Research*, vol. 11, pp. 247–292, 1992.

[77] F. S. Werblin and D. R. Copenhagen, “III. lateral interactions at the inner plexiform layer,” *The Journal of General Physiology*, vol. 63, pp. 88–110, 1974.

[78] B. Roska and F. S. Werblin, “Rapid global shifts in natural scenes block spiking in specific ganglion cell types,” *Nature Neuroscience*, vol. 6, no. 6, pp. 600–608, 2003.

[79] T. W. Troyer, A. E. Kurkowski, N. J. Priebe, and K. D. Miller, “Contrast-Invariant Orientation Tuning in Cat Visual Cortex: Thalamocortical Input Tuning and Correlation-Based Intracortical Connectivity,” *The Journal of Neuroscience*, vol. 18, pp. 5908–5927, August 1998.

[80] S. P. H., “The ON and OFF channels of the Visual System,” *Trends in Neurosciences*, vol. 15, pp. 86–92, 1992.

[81] L. O. Chua and L. Yang, “Cellular neural networks: Theory and applications,” *IEEE Transactions on Circuits and Systems*, vol. 35, pp. 1257–1290, 1988.

[82] L. O. Chua and T. Roska, “The CNN paradigm,” *IEEE Transactions on Circuits and Systems*, vol. 40, pp. 147–156, 1993.

[83] L. O. Chua, T. Roska, and P. L. Venetianer, “The CNN is universal as the Turing Machine,” *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, vol. 40, no. 3, pp. 289–291, 1993.

[84] T. Roska, L. Kék, L. Nemes, Á. Zarándy, M. Brendel, and P. Szolgay, “CNN software library (templates and algorithms), version 7.2.,” Tech. Rep. DNS-CADET-15, Analogical and Neural Computing Laboratory, Computer and Automation Research Institute, Hungarian Academy of Sciences (MTA SzTAKI), 1998.