A Tremor-Based Retina Model for Robust Contour Detection

Ádám Csapó, András Róka, Barna Reskó, Péter Baranyi

Cognitive Vision Research Group,
Computer and Automation Research Institute, Hungarian Academy of Sciences
baranyi@sztaki.hu

Abstract: The biological relevance of traditional and widely adapted edge-detection algorithms with overlapping receptive field architectures has been disproved in very recent papers of cognitive researchers. In the biological vision system, cells responsible for visual acuity do not overlap in their receptive fields. In order to reconcile differences between the biological vision system and the overlapping architecture used in classical convolution-based image filtering algorithms, this paper introduces a novel tremor-based retina model. The model takes into consideration data convergence, as well as temporal properties of ganglion cells. Based on the evaluation of the information-theoretical implications of the model, a hypothesis is formulated, according to which involuntary tremors are used by the biological vision system to compensate for the lack of overlaps through time. From this follows a second hypothesis, which formulates that from an information processing point of view, the functional role of involuntary eye-movements extends to more than just the maintenance of action potentials. In support of these hypotheses, the article discusses the model's biological relevance. Numerical simulations are also presented.

1 Introduction

The most prevalent image-filtering artificial receptive fields used today, such as Laplace and Sobel operators, work reliably only when assuming a convolution-based receptive field architecture with overlaps. Using these edge-filtering without any overlap in receptive fields results in a vast number of undesirable blind spots.

Recent findings in the research of retinal physiology suggest that receptive fields of the same types of mammalian ganglion cells barely overlap in the fovea [1, 6]. The model proposed in this paper uses a non-overlapping receptive field architecture enhanced with artificial tremors (rapid, involuntary eye movements) in order to make up for losses in information. The image is moved around randomly at a steady rate, thus an overlap is achieved between receptive fields through time. Aggregating responses of such artificial ganglion cells at given time intervals produces an effect similar to that observed when using overlapping receptive fields, but computation time and the amount of handled information are greatly reduced. The positive effects of the model can be used efficiently in real-time systems with demanding time constraints.

The paper is structured as follows: in an introductory section, an overview is given on the biological elements which served as an inspiration for this model. Further sections treat the details of the model. Test results and applications are also presented.

2 Biological Overview

The state of a nerve cell in the biological vision system is affected by every photoreceptor cell that provides it with input (either directly or indirectly). This set of cells providing input is referred to as the receptive field of the nerve cell. The structure of receptive fields can be observed from the firing patterns of cells when they are stimulated in an artificial environment.

The structure of the most common kinds of receptive fields in the fovea (the central area of the retina responsible for image contouring) is center-surround. The intensity of incoming light is reflected by the membrane potentials of photoreceptor cells (rods and cones) [7]. Photoreceptor cells providing input to the center of receptive fields provide excitatory (or inhibitory) input, while those providing input to the periphery are responsible for inhibitory (or excitatory) input.

Such a structure leads to higher-order ganglion cells being able to be depolarized only when the light intensity reaching the center of the receptive field differs from the light intensity transmitted to the periphery.

2.1 Non-Overlapping Receptive Fields

For a long time, experiments have shown extensive overlaps between receptive fields on the retina. However, when comparing the number of axons of photoreceptor cells to the number of axons in the optic nerve, it was discovered that the 130 million axons of rods and cones are condensed into 1.2 million axons in the optic nerve. Because of this fact, it was assumed that the retina performs some kind of information compression [4].

With the evolution of experimental methods, it was possible to make a distinction between many kinds of ganglion cells. Devries and Baylor [2] were able to distinguish between 11 kinds of ganglion cells, based on their receptive fields and response characteristics. Different kinds of ganglion cells provide different kinds of information, such as contour information, intensity information, motion information, as well as information on uniformly lighted image segments. It was also shown that receptive fields of ganglion cells of the same type do not overlap in the central fovea; the center of these receptive fields are located at a distance of one diameter. These measurements were confirmed by Packer and Dacey [6].

Because of the lack of overlaps between receptive fields in the fovea, only intensity transitions that fall in the center of the receptive fields can be detected at any given time. The computational model proposed in this paper provides a possible explanation to how humans are nevertheless capable of detecting contours.

2.2 Role of Involuntary Eye Movements

The three major kinds of involuntary eye movements that occur during fixation are microsaccades, drifts and tremors (also sometimes referred to as nystagmuses) [5, 3].

Of the three eye movements, later sections of this paper concentrate on tremors. Tremors are involuntary, rhythmic oscillations of the eye that have frequencies of about 90 Hz and amplitudes of roughly the diameter of a cone on the fovea (therefore the diameter of the smallest of photoreceptor cells). There is currently no hypothesis on the functional role of tremors, however, artificially eliminating tremors, researchers have found that vision faded away.

The model for edge-detection proposed in this paper uses non-overlapping receptive fields, but also incorporates tremors in order to achieve the effects of overlapping receptive fields through time. It will be shown that besides following the structure of human visual perception, the model accounts for the 130:1 information reduction ratio characteristic to the pathway between photoreceptor cells of the retina and ganglion cells [8].

3 Tremor-Based Retina Model

In order to achieve its goal, the proposed edge-filtering model uses artificial receptive fields that are structurally similar to those found on the retina. However, unlike the human vision system, the current implementation of the model uses constant-sized, 3-by-3 artificial receptive fields. This approximation is justified because in this case only foveal receptive fields are modeled, as the model is used for the filtering of edges in still images, not for detecting abrupt temporal changes in moving images.

3.1 Receptive Field Structure

Receptive fields are represented by two-dimensional matrices (as in many previous models), each matrix value representing a weight with which the corresponding stimulus is multiplied. The configuration of weights depends on the type of receptive field being modeled; on-centered fields have positive values in the central area, surrounded by

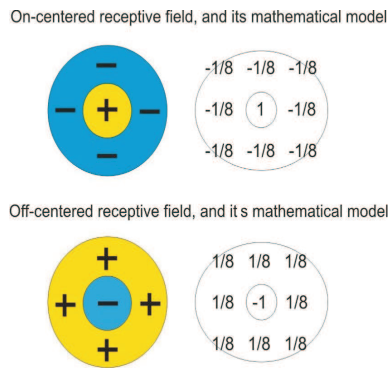

Figure 1: Mathematical approximation of biological receptive fields used in the model

all negative weights, while off-centered fields contain inverted values with respect to its on-centered counterpart.
The weighting used in this model approximates an operator that calculates the second derivative of the image, and in this respect, resembles the Laplace-operator (Figure 1). To demonstrate the necessity of overlaps in receptive fields, figure 2 shows an edge-filtered image using the approximated Laplacian operator, as well as the case where receptive fields are used in a non-overlapping manner. It is clear that the lack of overlaps results in a deteriorated image; certain sections of line segments are either extremely blurred, or are completely missing. While the model takes into account the convergence between photoreceptor cells and ganglion cells by yielding an output of 1/9-th the size of the input (there are no overlaps between the receptive fields used, therefore each output corresponds to a disjoint 3-by-3 portion of the image), it does not incorporate temporal properties of ganglion cells. Involuntary eye-movements are also disregarded.

Figure 2: The image on the top was obtained with overlapping receptive fields, while the one on the bottom was obtained using non-overlapping receptive fields. The losses in information are clear.

3.2 Temporal Model of Ganglion Cell Responses Used in the Tremor Based Retina Model

Ganglion cells can be categorized not only by their receptive fields, but also based on the temporal properties of their responses.

Ganglion cells in the fovea generally produce increased activity when stimulated, but when the stimulus disappears (or even if the stimulus is kept alive), their activity only gradually decreases until it reaches 0. However, if for some reason, stimulating effects increase, the output of ganglion cells increases even more, and only declines to 0 after a certain amount of time has passed.

The stochastic response properties of ganglion cells are approximated using a sliding window function. This solution conserves the essence of functionality, but at the same time ensures that unnecessarily complex computations are not brought into the model. The modeled ganglion cells produce the sum of their inputs, and hold their outputs for T time units. If, during this time, their inputs are stimulated even

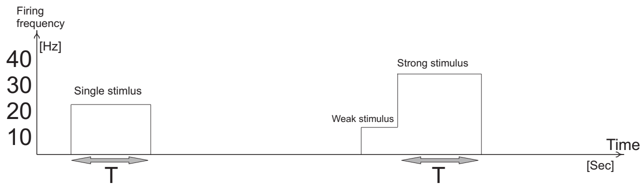

Figure 3: Temporal response characteristics of the modeled ganglion cells.

more, then their output will increase. Otherwise, after T time units, the output level returns to 0 (figure 3).

3.3 Model Incorporating Tremors

The weighting of the 3-by-3 matrices used to model receptive fields, can be seen in figure 1. The sets of pixels seen by each receptive field are disjoint, and their union covers the whole image. Each ganglion cell computes the sum of its inputs, and produces its output according to the window function described in 3.2.

The input image is in constant motion, compared to the stationary receptive fields. This motion is a sequence of random jerks, with a maximum amplitude of 2-3 pixels. In this way, receptive fields are made to overlap through time.

At any given moment, the output of each ganglion cell is determined as the maximum of the sum of its inputs throughout the previous T time units. Hence, the output of each cell is characterized by the largest change in contrast during T units.
In mathematical terms, the response of each artificial ganglion cell is:

$$ED(Im, Rs, A, T, t) = \max_{t-T<i\leq t} \{\Delta * f_i(Im, Rs, A)\} \tag{1}$$

where Im is the intensity image, Rs is a random seed, A is the maximal amplitude of tremors, and T is the parameter characteristic to the temporal function describing the response of the ganglion cell. Δ is the on-centered receptive field shown in figure 1, and f_i is a function that translates the image both horizontally and vertically by a random number of pixels calculated using random seed Rs.

Typically, the maximum firing frequency of nerve cells is 40-50 Hz. Such frequencies can only be perceived if firing is present for at least two periods. In case of a firing rate of 40 Hz, this doubled period time is 50 ms. This gives tremors with 90 Hz frequencies time for 4-5 periods. Therefore, in the model, the minimum value of T needs to be 4-5 times the period of tremors.

The maximum value of T can be approximated using different considerations. The visual system needs about 200-250 ms for the recognition of a simple object at a subconscious level. During this time, the amelioration of edge-detection performance is still worthwhile. 250 ms are enough for 22-23 tremor periods. Therefore, the maximum value of T is about 22-23 times the period of tremors.

From a biological point of view, the factor between T and the frequency of tremors is between 4 and 22. Figure 4 shows the edge-filtered image with respect to T. It can be clearly seen that the growth in image quality slows down as T increases (figure 5).

It should not be considered as certainty that the most valid way of modeling tremors is by using stochastic movements. If tremors were not random, but composed of the superposition of several vibrations having different frequencies, high-quality edge-filtered images could be obtained even for lower values of T. In literature, there are relatively small amounts of data on the composition of tremors, precisely because no functionalities were attributed to tremors thus far.

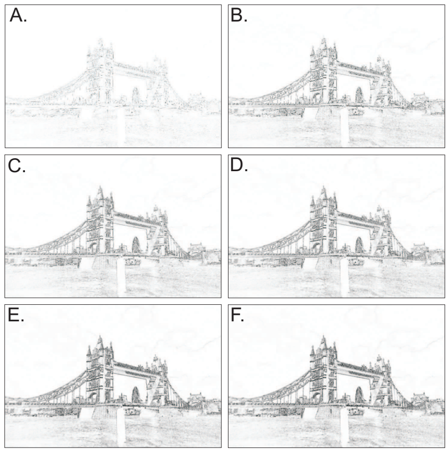

Figure 4: The edge-filtered image with respect to T. (A:T=1Tr, B:T=5Tr, C:T=10Tr, D:T=20Tr, E:T=40Tr, F:T=100Tr, where Tr is the number of tremor periods.)

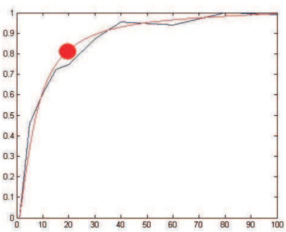

Figure 5: The edge-filtered image quality with respect to the number of tremor periods. By making the assumption that the model does not introduce false edges to the result, image quality can be empirically characterized by the percentage of non-white pixels compared to all pixels in the image.

Figure 6: Convergence / compression has a 9 : 1 ratio using 3-by-3 receptive fields.

4 Evaluation of the Model in Terms of Spatial and Temporal Resolution

The model proposed in this paper can be of use in the early functional stages of distributed intelligent systems with very simple computational units, primarily because of its accuracy and its low exigence of memory. The output image provided by the model, while perfectly capable of representing all relevant contour information, is reduced in size by a ratio of 1:9 compared to the original input image (figure 6). In a possible future implementation modeling the variety of receptive field sizes on the retina, even greater size reductions could be reached.

Due to the lack of overlaps in receptive fields, implementations of the model are also faster than previous models. This is especially true when analyzing moving images. A moving image can be sampled, and in one possible scenario, each output image - provided at every sampling interval - could contain the maximum of the last T samples. Because of the adaptivity of this method, stable elements of the image would be stable on the output (despite the fact that each sample was only filtered once instead of T times), while rarely occurring sudden changes in the input images would yield blurred image portions similar to those perceived when a rapidly moving object crosses the visual areas linked with peripheral receptive fields. In this case, even though each sampled image would be only filtered once (instead of T times), the same effect could be achieved through time, because abrupt changes are very rare at high sampling rates. Hence, computation times can be reduced by an order of magnitude.

5 Conclusion

A novel method for edge-detection was proposed. The model used was based upon previous edge-filtering methods as well as recent discoveries in retinal physiology.

Because of the non-overlapping receptive field architecture used, the obtained edge-filtered image is reduced in size compared to the original image. The model does not take into account the fact that receptive fields in the biological vision system have varying sizes. An enlargement of peripheral receptive field areas in terms of pixels would approximately account for the 130:1 reduction ratio present in the human vision system.

In the evaluation of the model, two hypotheses were formulated. According to the first one, involuntary tremors play a crucial role in causing overlaps in otherwise non-overlapping receptive fields through time. The second hypothesis states that even if tremors would have a role in the maintenance of action potentials (as is the case with microsaccades), their functional role extends to much more from an information processing point of view.

Through test results, the sensitivity of obtained results in terms of the model's variable parameters was treated. The obtained results compete with the best filtering techniques used today. Because of its robustness, low memory exigence and its time-wise optimized implementations mentioned earlier, the model could be efficiently used in distributed systems with cost-efficient computational units.

References

[1] Steven H. DeVries and Denis A. Baylor. Mosaic arrangement of ganglion cell receptive fields in rabbit retina. *The Journal of Neurophysiology*, 78(4), 1997.

[2] Baylor DA Devries SH. Mosaic arrangement of ganglion cell receptive fields in rabbit retina. *J. Neurphysiol.*, 78(4):2048–2060, 1997 Oct.

[3] Hennig and Worgotter. The effect of fixational eye movements on hyperacuity. Computational Neuroscience, Dept. of Psychology, University of Stirling.

[4] Meister M. Multineural codes in retinal signaling. *Proc Natl Acad Sci USA*, 93(2):609–614, 1996 Jan.

[5] Susana Martinez-Conde, Stephen L. Macknik, and David H. Hubel. The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience*, 5(3):229–240, 2004.

[6] Dacey DM Packer OS. Receptive field structure of h1 horizontal cells in macaque monkey retina. *J. Vis.*, 2(4):279–292, 2002.

[7] D M Schneeweis and J L Schnapf. The photovoltage of macaque cone photoreceptors: adaptation, noise, and kinetics. *J. Neurosci.*, 19(4):1203–1216, 1999 Feb.

[8] R. Sekuler and R. Blake. *Perception*. Mc Graw Hill, 1994.