Edge Detection Model based on Involuntary Eye Movements of the Eye-retina System

Barna Reskó, András Róka, Ádám Csapó, Péter Baranyi

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

Abstract: Recent results in retinal research have shown that ganglion receptive fields cover the mammalian retina in a mosaic arrangement with very small overlap. Classical biology inspired models of the retina performing image processing operations such as edge-detection, in contrast, by using convolution of a filter and the image, map overlapping input image regions to output signals. Without such an overlap, the output becomes a sparse sampling of the filtered output, which leads to the loss of important pieces of visual information. This paper proposes a retinal model with a non-overlapped receptive field structure. The model implements involuntary eye movements, tremors and drifts. Artificial eye-tremors, small vibrations of the eyes, achieve receptive field overlaps through time. Artificial eye-drifts are also implemented and used for orientation selective edge detection. Based on the experimental results, the following three hypotheses are claimed in this paper: 1) convolution-based retinal models are biologically inadequate, 2) eye tremors may play an important role in retinal contour perception, and 3) artificial eye drifts may also be applied in artificial vision systems to achieve orientation selectivity.

1 Introduction

Soft computing techniques, biology inspired approach to computation became more and more popular in the last decade, with a lot of related research interest. Artificial vision and image processing has also tried to exploit the knowledge accumulated about biological vision systems, from the retina to the visual cortex. This lead to biologically inspired computational models of different parts or functionalities of the visual system, from edge detection to more complex operations [1] [2].

There is however one thing in common between the majority of the so called biologically inspired models of image processing: their output image is computed as the convolution of an input image and an image filter [3] [1]. This has two important implications on the output image:

1 Input and output images have the same spatial resolution,

2 A certain pixel value on the input image influences several pixels values on the output image.

According to findings in retinal research, the number of retinal cones (photoreceptors sensitive to color) is about 6 million, the number of rods (photoreceptors sensitive to light regardless of its color) is about 125 million. Meanwhile, the number of ganglion cells sending the visual information to the brain is about 1.2 million. This suggests a heavy retinal image processing and compression (also referred to as convergence), first described in [4, 5]. There is thus a difference of a magnitude between the spatial resolutions of photoreceptors and ganglion cells. Furthermore, receptive fields of mammalian ganglion cells barely overlap each other, they tend to cover the foveal areas of the retina in a mosaic arrangement [6]. An important implication is that a certain photoreceptor on the retina is very likely to influence the output of only one ganglion cell.

The above findings pose a contradiction between the biology inspired, convolution-based models of image processing, and the receptive field structure of the retina. To solve this contradiction, this paper proposes an image processing model for edge-detection, which is based on the non-overlapping arrangement of center-surround structured retinal ganglion-cell receptive fields. The proposed model also performs the information compression (convergence) experienced in the retina.

It has long been known that our eyes are never still, even during fixation. Scientists today agree on the existence of three main involuntary eye movements during visual fixation in humans: tremor, drifts and microsaccades ([7] [8]. The most interesting eye movements concerning the proposed model are tremor and drifts. Tremor, the smallest of all eye-movements, is an aperiodic, wave-like motion of the eyes [9], with a frequency of roughly 90 Hz and amplitude of about the diameter of a cone. Drifts are slow wandering motions of the eyes that take place between microsaccades, and can span through a few dozens of photoreceptors [10].

Experiments with the proposed model revealed that the output image quality becomes much better if a high frequency, low amplitude vibration (or we can call it an artificial tremor) is introduced to the input image. If artificial drifts are also included, and they effect the input image directly, the output image tends to contain oriented contour segments corresponding to drift orientations.

Based on the contradiction between convolutional techniques and new findings in biology, as well as on the experimental results published in this paper, the following three hypotheses are claimed:

1 Convolution-based image processing models are biologically inadequate,

2 Eye tremors may play an important role in contour perception,

3 Artificial eye drifts may support orientation selectivity in artificial vision systems.

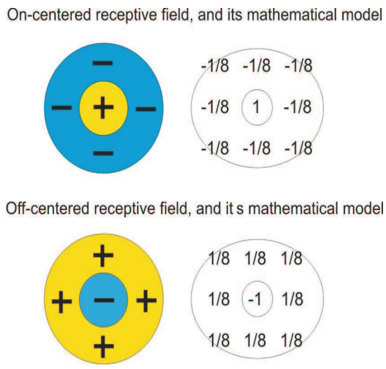

Figure 1
Approximation of biological receptive fields by matrix F used in the model

The rest of this paper is organized as follows: In Section 2 the model based on the non-overlapped receptive field structure of ganglion receptive fields is presented. In Section 3 this is followed by the application of tremors and drifts to the model, along with experimental results. In the following Section, three hypotheses are claimed, supported by the results of this paper and novel findings of biological research. Finally, Section 4 concludes the paper.

2 The proposed eye-inspired model for edge detection

2.1 Filters used for edge detection

The model that we propose in this paper performs edge detection based on the on-center off-surround and off-center on-surround receptive fields reported to be present in the retina by Hubel in [11]. On the retina, the size of the receptive fields increase from the fovea towards the peripherals, causing blurred vision on these areas. The model we propose considers only the foveal areas of the retina, where the size of receptive fields is relatively small and uniform. This constrains the size of the receptive fields of the model to be constant.

Receptive fields are represented by a 3×3 pixel-size, two-dimensional filter matrix F (as in many previous models), each matrix value representing input weight with which the corresponding stimulus is weighed. 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 all negative weights, while off-centered fields contain a central negative value, surrounded by all positive weights.

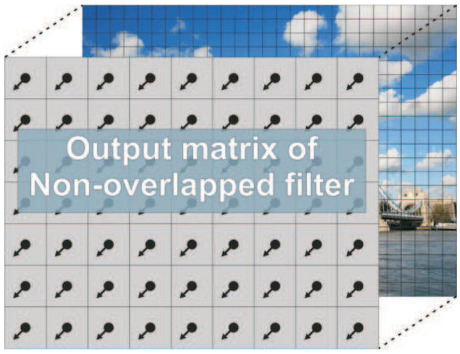

Figure 2
The mosaic-like filter matrix arrangement over the input image

The weights of F used in this model resembles the Laplace-operator, which calculates the second derivative of the image. The Laplace operator is approximated by an operator that contains the exponentials of 2 in order to facilitate binary representation and computerized implementation of the model (Figure 1).

2.2 Non-overlapped image filtering

Similarly to the non-overlapping receptive fields of the foveal areas of the retina, the proposed model also lacks overlaps between filter matrices. The input image is tiled using a mosaic arrangement of the filter matrix F in a non-overlapped manner. The center of each F is at a distance of three pixels from the four closest filter matrix centers. The image pixel values are multiplied by the corresponding F values and a weighed sum is calculated for each filter matrix center. The matrix of the filter centers provide the output matrix of the non-overlapped filtering operation. Figure 2 shows this processing structure.

Note that the output of this processing structure is identical to the 3-by-3 subsampling of the convolution of the input image with the same filter matrix F. The output thus contains 9 times less pixels than the output obtained using convolution. In order to make the two results comparable, the size of the input image used in convolution-based filtering is also reduced by a factor of $1/3$ in each dimension. Figure 3 shows the difference between the two approach, using the approximated laplace operator in both cases.

The non-overlapped image filtering can be formalized as follows. Denote the input image I, the filter mask F, the output images J and K obtained using classical convolution-based and the non-overlapped filtering respectively. In classical convolution based filtering, the output can be expressed as the convolution of I and F over

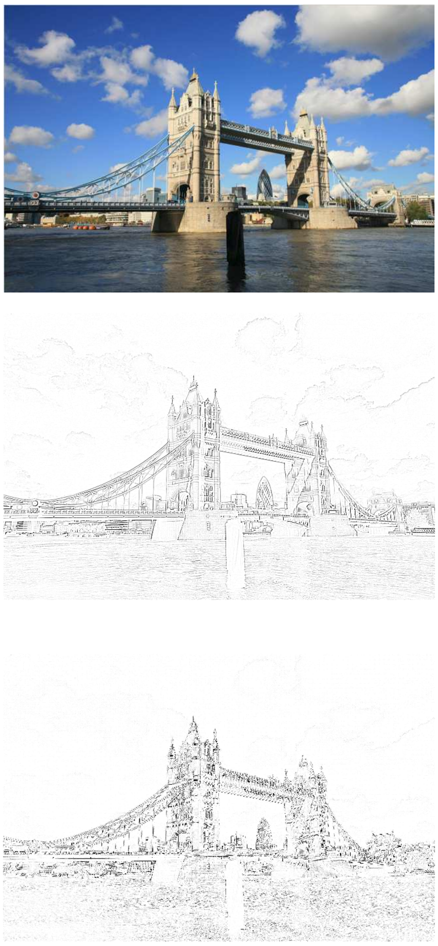

Figure 3
Input image (top), convolution (middle) and non-overlapped filtering-based (bottom) edge-detection using the approximated laplace operator

a discrete 2D space:

$$J[x, y] = F \otimes I \equiv \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} F(j, k) \cdot I(x - j, y - k). \tag{1}$$

The result of the non-overlapped filtering K can be obtained from the result of the overlapped filtering J, as follows:

$$K_{i,j} = J_{m \cdot i, n \cdot j}, \tag{2}$$

where n and m indicate the size of matrix F, in most of the cases $n = m = 3$. It can be seen from equation 2 that the number of elements in K is $m \cdot n$ smaller than those in J. The bad quality of the non-overlapped result K shown in Figure 3 is obviously caused by the exclusion of 8 pixels out of 9, when derived from the convolution-based output J.

It is to note that the non-overlapped filtering gives similar results with any filter operator matrix F, such as the sobel operator. In the proposed model a Laplacian-like operator was used to achieve a maximal similarity to the retina.

In the next section we present how to overcome the poor quality of the output image K obtained using non-overlapped filtering.

3 Artificial eye movements, results

The model described in the previous section resembles the eyes only in the arrangement and connectivity of light sensitive cells (photoreceptors) and visual information processing cells (ganglion cells) of the retina. While the eyes perform three different motions (microsaccades, drifts and tremors), the proposed model is mechanically static.

In this section first artificial tremors are introduced to the model. The artificial tremors can be simulated by software, and can also be implemented by adding mechanical vibrations to the image sensor (camera). Later in the section artificial drifts will also be introduced to the model.

3.1 Tremors

In order to create an artificial tremor that is maximally similar to its biological counterpart, the physical properties of tremors have to be considered. According to [12] the eye-tremors have a frequency of about 90Hz, and an amplitude of about one cone. The critical flicker frequency of the eyes in foveal regions is about 10Hz [13].

Based on these measurement data, an artificial tremor can be designed for the non-overlapped filtering model. The parameters are chosen to obtain an artificial tremor with maximal similarity to the tremor of the eye-retina system. The parameters of the artificial tremor are:

Amplitude:	$A = 2\ pixels$
Frequency:	$f = 9\ frame^{-1}$
Direction:	random

The simulation of the tremors is done by the shifting of the input image I to a random direction by one $A/2$ pixel, which yields the image I_{s_1}. The shifted image I_{s_1} is then filtered in a non-overlapped manner using the filter F, resulting in K_{s_1}. This process is done as many times as defined by the frequency parameter, resulting in images $K_{s_i}, i = 1 \ldots f$. The f images are then integrated into one image K. The integration can be pixel-wise weighed summation:

$$K = \sum_{i=1}^{f} w_i \cdot K_{s_i}, \tag{3}$$

where w_i is the weight for each output image K_{s_i} or maximization:

$$\forall x, y : K[x, y] = \max_i K_{s_i}[x, y]. \tag{4}$$

The resulting image K is one output frame of the non-overlapped filtering system.

Using this approach, each output pixel aggregates edge information from its neighborhood defined by the properties of the artificial tremor. Unlike in the static (tremorless), non-overlapped case, edge information contained in the input image will not be lost, but aggregated into the output pixel. A similar information aggregation or compression takes place in the retina, called convergence, which accounts for the 1:125 ratio between the ganglion cells and the photoreceptors.

The results of using the non-overlapped filtering with and without artificial tremors are compared on Figure 4.

3.2 Drifts

Drifts are continuous motions of the eye occurring between microsaccades, and sweep through about a dozen of photoreceptors. The role of drifts in the biological vision system is to compensate microsaccades in maintaining accurate visual fixation [12].

Based on the results obtained using artificial tremors, the question arose: what results would the non-overlapped filtering technique yield with the introduction of

Figure 4
The output of the non-overlapped filtering, without (left) and with (right) using artificial tremors. The laplace operator was us in both cases

Figure 5
Non-overlapped filtering after applying a horizontal or a vertical artificial drift respectively

artificial drifts? To answer the question, the parameters of artificial drifts have to be given. Again, we tried to design artificial drifts with a maximal similarity to the parameters of eye-drifts found in biology. The result is a motion that goes through 15 – 20 pixels, in a given direction, following a straight line.

During an artificial drift, the output frames obtained by the non-overlapped filtering with artificial tremors are used as input. The pixels of these frames are averaged along the path of the drift, which is identical to applying a motion blur operator on the frames.

The results of the application of artificial drifts on the images are shown in Figure 5.

For the first look the results are quite blurred and the small details became invisible. A closer look however reveals that the pixels corresponding to an oriented contour (horizontal or vertical in the case of the two sample images on Figure 5) are stronger on the resulting image. This suggests that the application of oriented artificial drifts on the edge detected image acts as an oriented edge enhancement operation. The

Figure 6
Orientation selective edge detection based on oriented motion blur and non-overlapped filtering

edge enhancement is however rather poor, which is not surprising since edge information on the pixel level is not orientation selective, and the artificial drift integrates contours perpendicular to its direction. Obviously this is something undesirable. The result is also in accordance with results of Hubel [11], who showed that orientation selectivity appears in the visual cortex, and not in the retina.

Going one step from the biological model of the eye gives much better results. It is much better to apply oriented artificial drifts on the original image instead of the edge detected image to obtain directed motion blur, with the purpose of orientation selective contour detection. The motion blur operation is a directional noise filter that smoothes noise (and any discontinuity in intensity level, including edges) in one direction, but does not modify the discontinuities in a perpendicular direction. The resulting image keeps the original contours that correspond to the blur direction, but blurs other contours as their orientation diverges from the blur direction. Applying the non-overlapped filter with artificial tremors on such an image yields a much better oriented edge detection, as shown in Figure 6.

Note that orientation selective edge detection using directed motion blur also performs well with classical, convolution based edge detection.

3.3 Evaluation of the non-overlapped filtering in terms of spatial and temporal resolution

The tremor-based non-overlapped filtering method proposed in this paper can be useful in early functional stages of visual processing, primarily because of its accuracy and its low exigence of memory. The output image provided, 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. This is because every artificial neuron receiving output corresponds to a 3-by-3 portion of the input. In a possible future implementation modeling the variety of receptive field sizes on the retina, even greater size reductions could be reached (the optic nerve in the human brain, for example - which is the continuation of the output of ganglion cells at the back of the retina - carry one thirteenth of the information amount that the ganglion cells previously receive).

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 - arriving at every sampling interval - could contain the average of the last f 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 f 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 f times), the same effect could be achieved through time, because abrupt changes are very rare at high sampling rates. This way, spatial attributes could be transformed into temporal attributes, and because results would arrive at the same frequencies as previously, computation time could be reduced by an order of magnitude.

4 Hypotheses

Supported by new findings in biology and experimental results presented in this paper, the following three hypotheses are claimed:

Hypothesis 1: Convolution-based image processing models are biologically inadequate.

This hypothesis is supported by the recent research findings in biology, claiming that receptive fields tile the surface of the retina in a mosaic arrangement with minimal overlap. Such an arrangement also accounts for the high level of image compression found in the retina but not found in the case of convolution-based filtering. If the same compression method (exclusion of 8 9th of the pixels) is applied to convolution-based methods, the result lacks most of the information and is very poor.

Hypothesis 2: Eye tremors may play an important role in contour perception.

This hypothesis is supported by the experimental results of section 3 that show the difference between the edge detected images obtained by using non-overlapped filtering with and without artificial tremors. The experiments show that the application of tremors dramatically enhance image quality. This is explained by the neighborhood integrating compression principle of tremors, in contrast with the neighborhood excluding compression principle of the tremorless, non-overlapped filtering method.

The hypothesis may extend to the biological eye-retina system.

Hypothesis 3: Artificial eye drifts may support orientation selective contour detection in artificial vision systems.

The results in section 3.2 show that biology equivalent implementation of eye-drifts do not produce orientation selective contours. On the other hand, the application of artificial drifts on the input image causing motion blur tends to keep only the oriented contours. Using the non-overlapped filtering can extract these contours, which tend to have a specific orientation equivalent to the orientation of the drift.

Conclusions

A novel, biologically adequate image filtering method was proposed. The method implements the non-overlapping mosaic arrangement of retinal receptive fields, as well as the non-voluntary eye-motions, the tremors and drifts.

The characteristics of the proposed method are different from those of convolution-based image filtering methods. Similarly to the retina, the non-overlapped filtering also implements image information compression, as the result of non-overlapped receptive fields and artificial tremors. In convolution-based filtering no compression is performed. The first hypothesis claimed in this paper is based on the above mentioned differences. It is to note that for human observers, the non-overlapped edge detected image looks more expressive than the convolution-based one, which might indicate that out vision system prefers the biology-close way of producing contour images.

The experimental results done with the proposed non-overlapped method using a Laplace operator show, that small, pixel-sized vibrations dramatically improve the quality of edge detected images. This is generalized in the second hypothesis, which claims that similar phenomena may exist in the eye-retina system, which would explain the yet unknown role of eye-tremors. This hypothesis requires further support from biological research and computer-based experiments.

The paper also presented experiments where well defined motions passing through a dozen of pixels were applied to the images, in analogy to eye-drifts. The role of the motion was to cause oriented blur between image pixels. It was proved by simulations that if the motion has parameters similar to eye-drifts, where the ganglion cells are fast enough to sample many times during the motion, orientation selectivity does not occur. However, if the same motion blur is applied to the input image, which would happen if ganglion cells and photoreceptors were too slow to sample many times during the motion, orientation selectivity occurs, as claimed by hypothesis three, supported by experiments presented above.

All the three hypotheses claimed by this paper need further support, from both biology and computer-based experiments.

References

[1] C. Grigorescu, N. Petkov, and M. A. Westenberg, "Contour and boundary detection improved by surround suppression of texture edges," *Image and Vision Computing*, vol. 22, no. 8, pp. 609–622, 2004.

[2] V. Navalpakkam and L. Itti, "An integrated model of top-down and bottom-up attention for optimal object detection," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (New York, NY), pp. 2049–2056, Jun 2006.

[3] J. Thiem, C. Wolff, and G. Hartmann, "Biology-inspired early vision system for a spike processing neurocomputer," in *Biologically Motivated Computer Vision*, pp. 387–396, 2000.

[4] H. B. Barlow, "Summation and inhibition of the frog's retina," *J. Physiology*, vol. 119, pp. 69–88, 1953.

[5] S. W. Kuffler, "Discharge patterns and functional organization of mammalian retina," *J. Neurophysiology*, vol. 16, pp. 37–68, 1953.

[6] S. H. DeVries and D. A. Baylor, "Mosaic arrangement of ganglion cell receptive fields in rabbit retina," *The Journal of Neurophysiology*, vol. 78, no. 4, 1997.

[7] A. L. Yarbus, *Eye Movements and Vision*. Plenum, 1967.

[8] R. H. S. Carpenter, *Movements of the Eyes*. Pion, 1988.

[9] L. A. Riggs, F. Ratliffa, J. C. Cornsweet, and T. N. Cornsweet, "The disappearance of steadily fixated visual test objects," *J. Opt. Soc. Am.*, vol. 43, pp. 495–501, 1953.

[10] F. Ratliff and L. A. Riggs, "Involuntary motions of the eye during monocular fixation.," *J. Exp. Psychol.*, vol. 40, pp. 687–701, 1950.

[11] D. Hubel, *Eye, Brain and Vision*. W.H. Freeman & Company, 1995.

[12] S. Martinez-Conde, S. L. Macknik, and D. H. Hubel, "The role of fixational eye movements in visual perception," *Nature Reviews Neuroscience*, vol. 5, no. 3, 2004.

[13] J. L. Brown, *Vision and Visual Perception.*, ch. Flicker an Intermittent Stimulation. New York: John Wiley and Sons, Inc., 1965.