Two feature extraction methods in the construction of a Visual Feature Array

Barna Reskó, Zoltán Petres, Péter Baranyi

Computer and Automation Research Institute
Hungarian Academy of Sciences

Abstract — This paper presents a cognitive model based on the visual cortex, which is able to perform image contouring by the means of extracting contour line segments as abstract objects. Two methods are presented to extract the line segments from the input image. These line segments are later organized according to their orientation and length in a three-dimensional array, the Visual Feature Array, which allows image processing and transformations along new dimensions, such as orientation or length. Negative Filtering is the process when the Visual Feature Array is used to reconstruct the original contour image by redrawing the stored line segments, thus removing noise from the image. The presented approach is strongly based on cognitive psychology and neurobiology. The processing model has a strictly parallel architecture in order to mimic its biological inspirator, and to allow constant time processing on a parallel computational hardware.

1 Introduction

In order to show why cognitive models can give the necessary boost in computation, consider the example where a person has to decide whether there is a cat or something else on an image. Such a task is impossible for a computer to perform today, yet a human can do it reliably in half a second or less. This result is very interesting when considering that the "processing time" of a typical neuron is in the range of milliseconds while that of a logic gate of a modern silicon-based computer is in the range of nanoseconds. The computational capacity of the brain thus has to lie in its special architecture and particular information representation and processing, rather than in the speed of its processing elements. It is our belief that in order to step beyond the borders of today's computer systems' architectures the basic way of information representation and processing has to be changed. For new ideas we turn to existing cognitive systems in biological architectures to study them, because they already bear the solutions that we are seeking for. Hubel and Wiesel first described the visual system [1], and suggested that iso-orientation domains are packed in essentially linear parallel stripes, which Hubel [2] subsequently referred to as the "ice-cube" model. The model of Hubel, and later V1 models [3] suggest that cells in the visual cortex are organized in a 3D structure, where a location on the visual field and an input stimulus preference (*e.g.* orientation preference) can be assigned to each cell. A cognitive system is implemented in a biological neural network, where simple units of computation are connected in a very complex structure. Our research goal is to turn the cognitive information processing system into engineering models which can later be organized into a cognitive psychology inspired model running on a biology related computational architecture.

This paper introduces a model strongly based on the cognitive functions of the visual cortex for image contour detection. The model was elaborated on the analogy of the mammalian visual system. Each phase from the retina to the visual cortex is represented in the model by imitating the biological structures and cognitive functions in order to perform similar image transformations and operations. In classical image processing algorithms, such as edge detection using a sobel filter, both the input and the output are matrices containing pixels. These algorithms thus represent a pixel-to-pixel transformation between two matrices. Similarly to the neural networks in the cerebral cortex, the model proposed in this paper implements a pixel-to-feature transformation, where *feature* refers to a more abstract visual object, such as a line segment of a certain length and orientation, or a line crossing. The result of the transformation is thus a feature-level abstraction of the input image. There are two different ways to do the pixel-to-feature transformation. In the first case a binary mask matrix is applied on the edge detected input image, as described in [4][5]. The other possibility is to use a mask matrix generated using a Gabor function, which approximates the response characteristics of cortical cells performing feature abstraction in the visual cortex [6].

The extracted abstract features can also be re-transformed into the pixel level by a feature-to-pixel inverse transformation, allowing a visual representation of the feature-level abstraction. The re-transformation of features into pixels will exclude noise from the result, thus it can be used as a filtering technique, described later in this paper.

The rest of the paper is organized as follows. Section 2 describes the proposed architecture of the model for high speed image processing. Section 3 is devoted to the model evaluation and experimental results. Finally, Section 4 concludes the paper.

2 Cognitive model of the visual pathway

A scene projected to the retina becomes a two-dimensional image, which is transferred to the brain for further processing. Such an image is composed of image features like regions of a certain color and texture, their boundaries as segments of different orientation and length. The image features make part of more abstract features like simple shapes, curves, circles.

The main goal of the present model is to *understand* the basic primitives of an image, on the analogy of the cerebral cortex as a complex cognitive system. The *understanding* of a feature in biology is defined as the firing of a set of neurons, which tend to fire when that particular feature is presented on the input as a stimulus. In the proposed model a feature is represented by the activation of a single neuron instead of a set, and it is considered *understood* when the corresponding neuron fires. The neurons of the understood features can project their outputs to higher and lower levels in the neural hierarchy. Projecting the output further up allows the neurons in higher levels to understand more abstract features as the composition of lower level features. On the other hand, a neuron that projects its output to lower levels in the neural hierarchy can be considered as an *expectation* from above, and will help the low-level neurons to understand the lower level features.

This paper concentrates on how primitive image features are understood, and how they can be used as an expectation in lower levels.

The proposed model in this paper receives an image on its input, which is immediately subjected to an edge detection filter. This filter is based on the receptive field characteristics of the retinal ganglion cells. In the small region of the visual field which is centered around the position of the ganglion cell the afferent connections have a relatively high positive weight, while in the surrounding regions the synapse weights are inhibitory. The receptive field is modeled with a 3×3 matrix M_1 with higher positive input weight values in the middle and small negative values in the surrounding regions.

$$M_1 = \begin{pmatrix} -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} \\ -\frac{1}{8} & 1 & -\frac{1}{8} \\ -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} \end{pmatrix} \tag{1}$$

The output pattern of the cells with input weights of M_1 will be an edge detected image of the original image. It is to note that at this level of neural processing the image features understood (or represented by neural activation) are pixels of an edge detected image, edge elements.

Going further on the visual pathway we find that the receptive fields of the neurons in LGN are also circular like those in the retina. This suggests that the LGN does not add any extra image processing functionality to the visual pathway. It rather has an important role in modulating the input to the cortex by attention, but the exact functionality is still a subject of research.

For the above reason we consider the retinal and LGN-neurons as primary edge detectors, and their overall functionality in the aspects of image processing is covered by the M_1 matrix in the model. The input from the cells of such receptive fields project into the visual cortex, where further image processing takes place.

The image representation in the visual cortex is retinotopic, which means that neighboring regions of the visual field are projected to neighboring regions in the cortex. The neurons of such a region are tuned to respond to a variety of input stimuli described by different receptive fields characteristics. This implies that a vast variety of receptive fields belong to one small region of the visual cortex, and thus to a small region of the visual field. The variety of receptive fields representing different visual features (*e.g.* line orientations) can be organized along new dimensions.

After an edge detection discussed above, an edge detected image is available in the matrix I where

$$I \in \mathbb{R}^{n \times m}, \tag{2}$$

n and m representing the image dimensions.

According to the visual cortex, several different features can be extracted from the edge detected image I. The extraction of the features begins with those having the largest number of pixels, *i.e.* the longest lines. When the first feature is extracted from the edge detected image I, the feature pixels are removed from I, resulting a new matrix that we refer to as $I^{(1)}$. After extracting and removing the k^{th} feature from $I^{(k-1)}$ the matrix $I^{(k)}$ remains. Using this notation the original edge detected image is denoted $I^{(0)}$. This step is necessary to ensure that only one of many possible similar features is extracted from the edge detected image $I^{(0)}$. The k^{th} feature is removed from $I^{(k-1)}$ and added to a two-dimensional matrix F_k, such that

$$\forall i, j, k : (F_k)_{i,j} \in \{0; 1\}, \tag{3}$$

and the value $(F_k)_{i,j}$ indicates if any pixel of the detected feature k is present in the edge detected image at the position $I_{i,j}^{(k-1)}$.

It is important to note that the features to extract are ordered by the number of pixels they contain in order to ensure that

$$\mathcal{F}_k \supseteq \mathcal{F}_l, k < l, \tag{4}$$

where $\mathcal{F}_k$ is the set of pixels contained by the k^{th} feature. Since there are several image features to be extracted from the image, there will be a matrix F for each of these features. We define the three-dimensional array with the F matrices overlapped along a third dimension as follows:

$$\mathcal{V} \in \mathbb{R}^{n \times m \times r} \tag{5}$$

For the three-dimensional matrix $\mathcal{V}$ we introduce the notion of *Visual Feature Array* or *VFA*, where r represents the total number of visual features. By construction, the element $\mathcal{V}_{i,j,k}$ of the VFA represents if an edge pixel $I_{i,j}^{(k-1)}$ belongs to the k^{th} visual feature.

In the VFA each element corresponds to the response of a cortical neuron tuned to a certain feature in a certain location. In the VFA the features are organized along a third dimension, orthogonal to the other two dimensions. Such a system of visual features yields a 3-dimensional neural array model of the primary visual cortex.

In the visual cortex there are neurons tuned to a whole variety of visual features. The present model includes the orientation selective cortical cells with end-inhibition characteristics. Each feature in the VFA can thus be described by an orientation angle and an optimal length. The possible orientations are equally distributed with a specified angular resolution. The angles represented in the VFA are defined with the angle α and angular resolution θ, such that

$$\alpha \in [0 \ldots \pi], \alpha = k \cdot \theta, k \in \mathbb{N}, \tag{6}$$

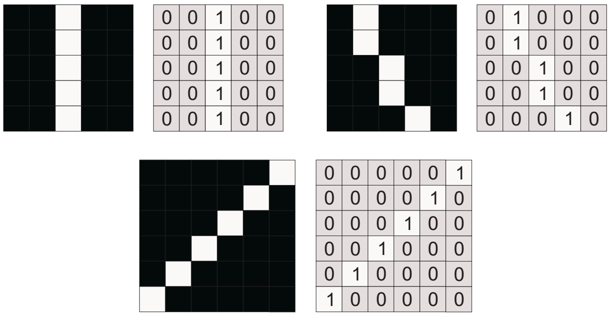


Figure 1: The matrices in the model that represent the receptive fields of cortical orientation tuned end-inhibited cells.

and thus the matrix elements $(F_{\alpha=\pi/5})_{i,j}$ will be values of 1 where an edge line segment with an orientation close to $\pi/5$ is found in the edge detected image at $I_{i,j}$.

The end-inhibition property of the neurons is also formalized in the model. An optimal length l of a neuron is a length to which it gives a maximal response. The different lengths are distributed between the shortest length and the longest length, and their number is h. Since the line lengths are measured in pixels, the shortest possible line segment is 3 pixel long. The maximal length can be chosen taking the requirements of the input image and the available computational capacity into consideration. Normally this value is between 20 and 30 pixels.

Given an angular resolution of θ and the number of different length values h, the number of possible visual features r can be assessed as follows:

$$r = \frac{\pi}{\theta} \cdot h. \tag{7}$$

A visual feature k is thus characterized by two values, an orientation α and length l. The matrix elements $(F_k)_{i,j}$ will thus have a value of 1 if the edge pixel on the edge detected image $I_{i,j}$ belongs a feature with the characteristics of k.

In the visual cortex there are receptive field characteristics that actually define the visual feature the particular neuron is responsive to. In order to extract the desired features from an edge detected image, for each feature k a mask matrix R_k obtained from a corresponding receptive field has to be defined. In the proposed model the visual features are extracted by a convolution of the edge detected image and a matrix R_k. In the present case the receptive fields are modeled by binary matrices instead of matrices with real values. These matrices contain the sought feature as it may appear on the binary edge detected image. We can choose to use binary matrices to detect visual features because it is possible to well approximate the sought features, and binary operations are easier to implement in a hardware. A series of mask matrices are shown in Figure 1.

Consider a grouping transformation on the VFA, which simply groups all the layers into one final layer containing all the extracted features. This transformation

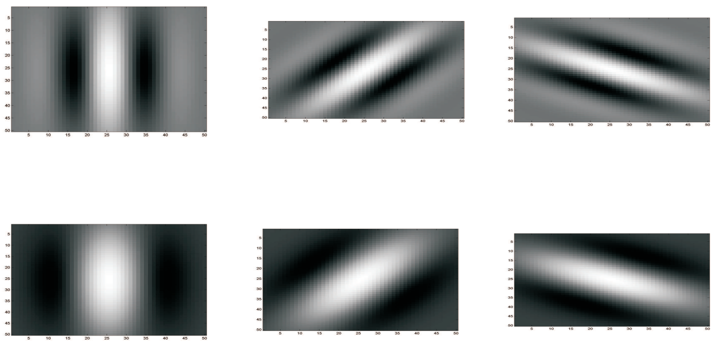


Figure 2: Several differently oriented non-binary RF matrices.

equals sending the output of the VFA neurons *down* in the neural hierarchy, and can be used to reconstruct an image by redrawing the detected visual features. This reconstruction will include only the features that were extracted from the original image. This implies that the noise (pixels not considered as the part of any feature) will not be present in the reconstructed edge detected image.

2.1 Feature extraction using Gabor functions

It is also possible to use a Gabor filter-like mask matrix to extract the image features from the image. Gabor functions can be determined as the product of a two-dimensional cosine, and a two dimensional Gaussian-function:

$$G_{\lambda,\sigma,\Theta,\alpha}(x,y) = \exp(-\frac{\tilde{x}^2+\gamma^2\tilde{y}^2}{2\sigma^2})\cdot\cos(2\pi\frac{\tilde{x}}{\lambda}+\alpha) \tag{8}$$

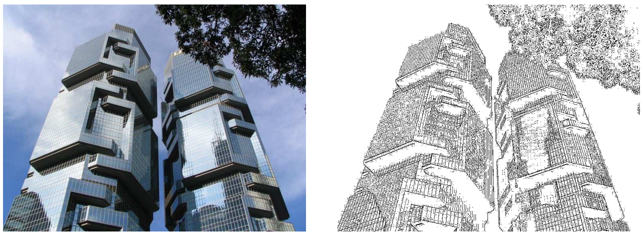
$$\tilde{x} = x\cos\Theta + y\sin\Theta, \tag{9}$$

$$\tilde{y} = -x\sin\Theta + y\cos\Theta \tag{10}$$

where γ is a constant, called the spatial aspect ratio, that determines the ellipticity, σ determines the size of the receptive field, $\frac{1}{\lambda}$ is the spatial frequency, Θ is the preferred orientation, and φ defines symmetry. Figure 2 shows different sized and oriented RF matrices.

Using these receptive fields, more sophisticated and cleaner results are obtained, because of the inhibitory areas of RF matrices (black areas in figure).

We have introduced the notion of *negative filtering* as the process of understanding image primitives and reconstructing the image from them. The notion arose from the fact that on contrary to a filtering process, the above defined process adds useful information to the image, instead of subtracting it.



(a) (b)

Figure 3: Original test image (a) and the result of the primary edge detection (b)

3 Model evaluation, results

The proposed model has two important advantages compared to classical solutions. By virtue of the simple but numerous computational units (neurons) that work parallel on the solution, the model can perform the proper activation of the VFA and the negative filtering in constant time. This, however requires a parallel hardware implementation of the model.

In our case only a computer simulation of the model was available, which allowed to evaluate the functionalities of the model, but not its performance.

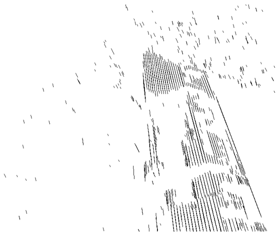
The input test image used to evaluate the model is shown in Figure 3a. This image is subjected to a primary edge detection according to the model. The result is a binary image of edge elements, with white dots representing high-contrast points on the original image. This edge-detected image is shown in Figure 3b.

The edge-detected image in our model corresponds to the image that is projected to the visual cortex. In the model, this image is used as the input to the neurons in the VFA. In the present implementation 5 different line lengths were used with the possible orientations to calculate the values of the VFA. These lengths were 3, 5, 9, 17, and 33 pixels.

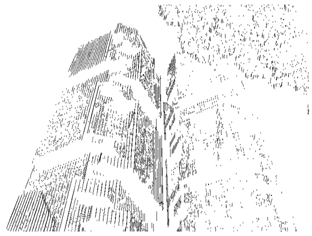
The VFA layers after the grouping the 3, 9 and 33 pixel-long segments are shown in Figure 4.

The union of the VFA layers yields the top-down reconstruction of the edge detected image from the detected line segments. The reconstruction will exclude the edge elements detected as noise noise, which was not recognized as a visual feature (a line segment of certain length and orientation). The final, fully reconstructed, negative filtered image composed from the five layers of $\mathcal{V}^{(l)}$ is shown in Figure 4d.

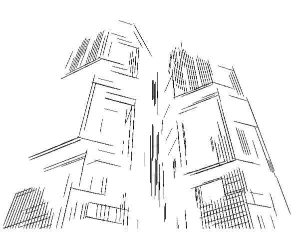
If we look at the case where the Gabor filter is applied, a much smoother edge representation is obtained.



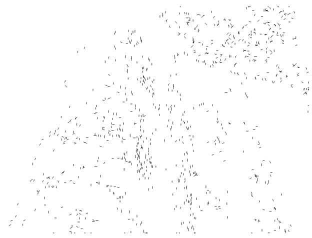
(a)



(b)



(c)



(d)

Figure 4: The segments of two different orientations (a)-(b) and the segments of two different line lengths (c)-(d).

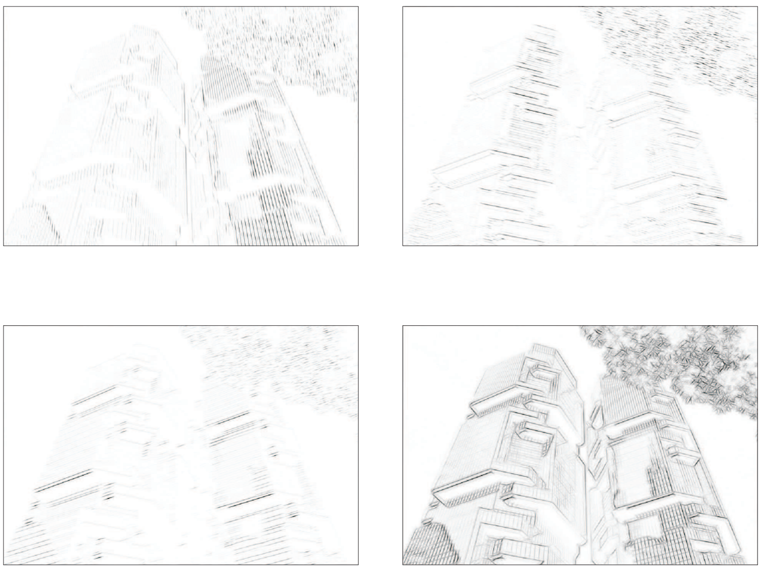


Figure 5: Results from the VFA module, using RF matrices containing Gabor-filters. The sub-figures from top to bottom: Edge-detected image, two sheets of VFA, and all of the orientations in one image.

4 Conclusion

A model for intelligent contour detection was presented in this paper. The basic structure and functionality of the model is based on the mammalian primary visual cortex, which can perform edge contour extraction on an edge detected image. The extracted contour pixels are clustered into visual features which are more abstract representations of the visual information. The features are organized into a three-dimensional orthogonal array (the VFA) according to their properties. The extracted features are used in two ways: further abstraction or top-down image reconstruction. This latest adds an augmented information space to the original edge detected image, which we refer to as negative filtering.

The two different feature extraction methods have several advantages and disadvantages when compared to each other. The application of binary masks allow a very fast filtering technique, however, the obtainable angular resolution is rather poor. The binary matrix has a very important advantage: it can be implemented using digital circuits.

The Gabor filter-like matrices have a great advantage on the quality of the feature extraction. They can be adjusted by several parameters, thus a very dense angular resolution can be achieved. The output of this extraction method is also different from that of the binary matrices. In the Gabor function case a membership degree of a certain point to a certain feature is obtained. This solution is much closer to the biological system, and further processing can be applied taking the feature strengths into consideration. However, a very important drawback of this method it it's computational complexity compared to the binary mask method. Using a special parallel processing computational tool can however solve this problem, making the application of Gabor-function a desirable solution.

The VFA containing different features can be submitted to grouping transformations, that merge layers of the VFA according to certain rules, such as similar line length or orientation. The grouping transformations are necessary for further transformations, such as line crossing and vertex detection.

The model and especially the VFA has been designed to operate in a fully parallel manner. In the present system binary array values were used for the sake of easy hardware implementation. An FPGA or other parallel implementation of the model yields a constant time contour detection and visual feature extraction.

References

[1] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiology*, 160:106–154, 1962.

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

[3] A. Grinvald, D. Malonek, A. Shmuel, D. Glaser, I. Vanzetta, E. Shtoyerman, D. Shoham, and A. Arieli. *Imaging of Neuronal Activity*. Cold Spring Harbor Laboratory, 1999.

[4] Zoltán Petres, Barna Reskó, Péter Baranyi, and Hideki Hashimoto. Biology Inspired Intelligent Contouring Vision Device in Intelligent Space. In *Proceedings of the 6th International Symposium on Advanced Intelligent Systems (ISIS 2005)*, pages 865–870, Yeosu, Korea, Sept 28–Oct 1 2005.

[5] B. Reskó, Z. Petres, A. Róka, and P.Baranyi. Visual cortex inspired intelligent contouring. In *Proceedings of Intelligent Engineering Systems (INES 2005)*, Athens, Greece, September 16–19 2005.

[6] John G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Opt. Soc. Am. A*, 2(7):1160–1169, 1985.