Production Systems and Information Engineering
Volume 4 (2006), pp. 33–51

A MODEL FOR VISUAL FEATURE EXTRACTION BASED ON THE MAMMALIAN VISUAL CORTEX

Barna Reskó
Computer and Automation Research Institute, Hungary
Cognitive Systems Research Group
resko@sztaki.hu

Zoltán Petres
Computer and Automation Research Institute, Hungary, Hungary
Cognitive Systems Research Group
petres@tmit.bme.hu

Péter Baranyi
Computer and Automation Research Institute, Hungary
Cognitive Systems Research Group
baranyi@sztaki.hu

[Received November 2005 and Accepted May 2006]

Abstract. The present paper proposes a model for intelligent image contour detection. The model is strongly based on the architecture and functionality of the mammalian visual cortex. A pixel-to-feature transformation is performed on the input image as the afferent visual information. The result of the transformation is a three-dimensional array of data representing abstract image features (contour objects), instead of another array of pixels. The contour feature recognition is performed by a vast and complex network of simple units of computation that work together in a parallel way. The use of a large number of such simple units allows a clear structure that can be implemented on a special hardware to allow fast, constant time feature recognition.

Keywords: Visual Feature Array, negative filtering, contour detection

1. Introduction

The main goal of this paper is to present a neurobiological and cognitive psychological analogy based cognitive framework. The framework is based on the biological architecture and cognitive functionalities of the mammalian visual cortex, which is able to perform image contouring in an intelligent way.

Besides the possibilities of practical applications of the framework, it also aims to extend the limits of classical computation.

In order to show why cognitive models can give the necessary boost, consider the example where a test person has to determine whether there is a cat or something else in the shown image, and press a button according to the decision. 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 becomes more shocking if we know that the "processing time" of the basic processing unit of the brain (a typical neuron) is in the range of milliseconds, while the basic processing unit (a logic gate) of a modern silicon-based computer is 5 million times faster. The answer for how the "slow" brain can solve this task lies in its special architecture and particular information representation and processing. It is thus 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. 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.

Our work has received inspiration from research about biological visual systems, [1, 2]. This is not to say that the model presented in this paper are necessarily identical with biological visual systems. The ultimate criterion of our work is performance from a technical point of view.

A cognitive process is an abstract concept which can be considered as an information processing function. A cognitive system is composed of many cognitive processes each responsible for a different task. By the complex structure of mutual interaction of the cognitive processes the cognitive system becomes very sophisticated with new limits of computation. A cognitive process only describes a functionality, but it does not say anything about the way of implementation, thus it can be implemented in many ways. One existing implementation of cognitive processes is the cerebral cortex of mammalian animals, where a very complex biological computational architecture provides the computational power for cognitive processes.

Such an architecture is built up by numerous, simple computational elements that can perform only primitive functions like addition, subtraction in a rather short time. These computational elements are connected to each other in a very complex network, like the neurons in the brain. The neural architecture can be much more efficient in certain tasks than the complex, classical algorithms, by virtue of the decomposition of the problem into thousands of simple independent operations which can be done simultaneously. The elaboration of such simple operations require simple hardware units that can be implemented in a chip with a clear and simple architecture. The resulting architecture is able to perform the computation in a fully parallel way, thus tremendously reducing the computational time. It seems thus to be promising to base the cognitive models on parallel architectures to achieve an efficient operation.

This paper introduces a model strongly based on the cognitive functions of the visual cortex for extracting image features of contour line segments. The model is based 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 pixels arranged in a matrix. These algorithms thus represent a pixel-to-pixel transformation between two matrices.

The notion of an image feature, or simply a *feature*, is defined as a visual object, which can range from a single pixel or edge element through an oriented line segment until a more complex corner or even a triangle. This suggests the introduction of a hierarchical organization of features along the abstraction dimension. So far, many work has dealt with the hierarchical organization of features according to scale factors [3, 4, 5, 6]. The abstraction hierarchy first introduced by Granlund [7] employs symmetry properties implemented by Gabor functions.

Accordingly, the more complex a feature is, the higher level of abstraction it is classified. A one-pixel-size feature can be considered as a feature of the lowest level abstraction. Similarly to the neural networks in the cerebral cortex, the proposed model implements a pixel-to-feature transformation, which should more precisely be referred to as a low-level-feature to high-level-feature transformation. The result of the transformation is thus a higher level feature abstraction of the input image. The abstract features can also be re-transformed into the lower level features they are composed of. In the case of a feature composed of pixels, this re-transformation will result in a pixel level representation of the features of higher level abstraction. The re-transformation of features into lower level features excludes 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 gives an introduction to the visual pathway, how the brain processes an image. Section 3 describes the proposed architecture of the model for high speed image processing. Section 4 is devoted to the model evaluation and experimental results. The fundamental ideas of the hardware realization of our model is discussed in Section 5. Finally, Section 6 concludes the paper.

2. The Visual Pathway from the Retina to the Primary Visual Cortex

The main goal of this paper is to present a cognitive model based on the visual pathway with a special respect on the primary visual cortex. The purpose of this section is to give an overview of the biological and cognitive aspects of early visual information processing, on which the model is based.

Visual processing begins in the retina. The photoreceptors that include 120 million rods and more than 5 million cones are located in the outer plexiform layer of the retina. The rods are sensitive to light intensity and are responsible for phototransduction [8], while cones are sensitive to the wavelength of the light [9]. These photoreceptors modulate the activity of the bipolar cells, which in turn connect with more than one million ganglion cells in each eye. The axons of the ganglion cells leave the eye at the optic disc and form the optic nerve, which carries information from the retina to the brain.

The bipolar cells and the ganglion cells are organized in such a way that each cell responds to light falling on a small circular patch of the retina, which defines the cell's *receptive field.* Both bipolar cells and ganglion cells have two basic types of receptive fields: on-center/off-surround and off-center/on-surround. The center and its surround are always antagonistic and tend to cancel each other's activity [10, 11]. On the other hand, the on/off or off/on arrangement of the receptive field makes ganglion cells more responsive to differences in the level of illumination between the center and surround of its receptive field. Uniform illumination of the visual field is less effective in activating a ganglion cell than is a well placed spot or line or edge passing through the center of the cell's receptive field.

The main target of the axons of the ganglion cells are the lateral geniculate nucleus (LGN) of the thalamus, and the superior colliculus. The LGN is the main conduit to the primary visual cortex where conscious visual perception occurs. The superior colliculus is involved in guiding eye movements and other automatic visuo-motor responses. The primary visual cortex (which is also referred to as V1, the striate cortex, or area 17) populates approximately 2 billion neurons in a two-dimensional sheet about 2–3 mm thick. Visual information processing totals up to a vast portion of cortical activity and is composed of more than a dozen separate areas. In macaque monkeys, the visual cortex constitutes about 50% of the surface area of the entire cerebral

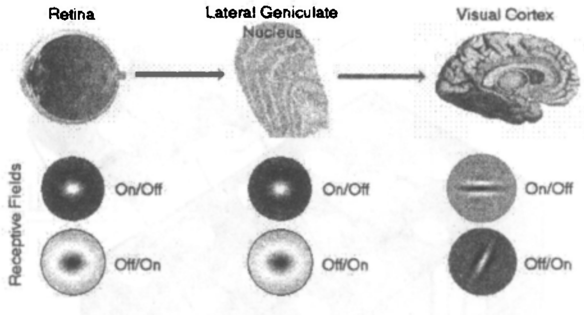


Figure 1. The visual pathway from the retina through the lateral geniculate nucleus to the visual cortex. The shape of the corresponding classical receptive fields varies from circular in the retina and LGN to elongated in the cortex. Orientation selectivity occurs only in cortical neurons.

cortex, while in humans this fraction is about 20%. The primary visual cortex topographically maps the visual field, with neighboring neurons responding to neighboring parts of the visual field.

Neurons in the primary visual cortex can be classified in two major classes according to their response characteristics: simple-cells and complex cells [2]. Simple cells tend to receive afferent projections mostly from the LGN, while complex cells receive projections mostly from other cortical cells [12]. Both of these cells exhibit a property known as orientation selectivity, meaning that they do not respond simply to light or dark in the visual field, but more typically to bars or edges of light with a particular orientation [13].

The visual cortex has a columnar organization on the cellular level. In 1977, Hubel and Wiesel suggested that iso-orientation domains are packed in essentially linear parallel stripes, which Hubel [1] subsequently referred to as the "ice-cube" model. The model of Hubel, and later V1 models [14] 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, as shown in Figure 2.

While simple cells respond to an oriented edge at a particular position of the visual field, complex cells exhibit more robust functionalities. An example of cortical processing in primary visual cortex is *length-tuning* or *end-inhibition*. Hubel and Wiesel first described complex cells in which the response to a stimulus increases with the length of the stimulus up to some optimum value, after which further increases in length decreased the response [15].

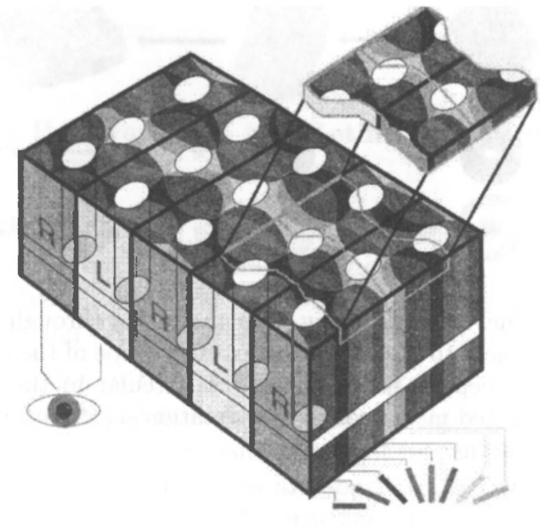


Figure 2. Schematic map of the visual cortex at work. This amazingly orderly "mosaic" of the working brain is formed by three groups of neurons performing different tasks: 1) Black lines mark the borders between columns of neurons that receive signals from the left and right eye and are responsible for the binocular perception of depth. 2) White ovals represent groups of neurons responsible for color perception (blobs). 3) The 'pinwheels' are formed by neurons involved in the perception of shape, with each color marking neurons responsible for a particular orientation of the visual field. (Reprinted with permission from [16])

3. Cognitive model of the visual pathway

A scene projected to the retina becomes a two-dimensional image, which has to be 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 work of Hubel and Wiesel states the existence of simple and complex neurons in the visual cortex [2]. Tao goes further, and introduces the complexity of neurons as a quantitative descriptor [12]. The larger synaptic distance a neuron is from the input, the more complex it is. This suggests that neurons connected in a complex network can be hierarchically classified into different levels corresponding to the synaptical distance from the input. The neurons

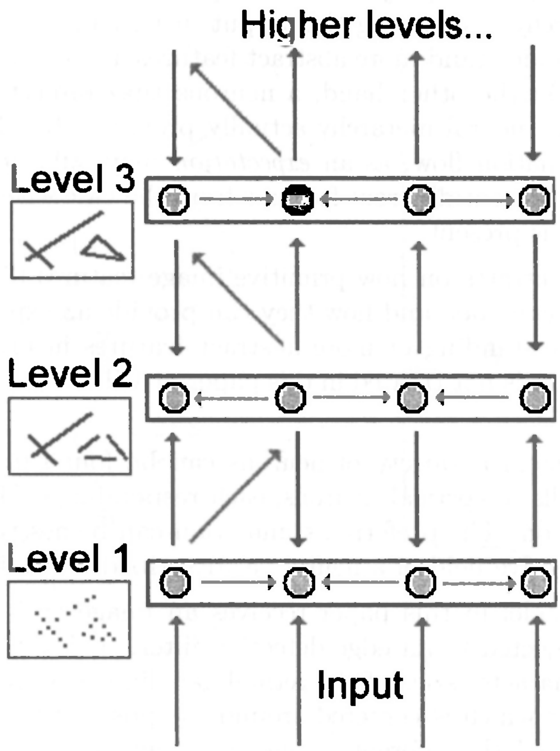


Figure 3. The neural hierarchy

in the n^{th} level may receive input from levels l_i where $l_i \geq n-1$. This means that a neuron in a certain level can receive input from one level bellow or from the same or higher levels. The first level consists of neurons that directly receive an input signal. In this paper we refer to such an organization as a neural hierarchy. Since each level is representing the abstraction of the level bellow, it can be supposed that the higher is the level of a neuron, the more abstract feature it can represent (Figure 3).

The main goal of the authors is to propose a cognitive model, which is able to *understand* (i.e. represent at an abstract level) the basic primitives (features) of an image, analogically to the cerebral cortex. In neurobiology a feature is *understood* when it causes the intensive firing of a set of neurons. In the proposed model a feature is represented by the activation of a single neuron instead of a set. A feature is considered to be *understood* by the model when the neuron corresponding to the actual feature has a high output. The neurons representing 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 neurons that project their outputs to lower levels in the neural hierarchy actually provide a top-down information flow. This information flow, as an *expectation* may influence the neurons in lower layers to represent different features from the case when only bottom-up information flow is present.

This paper concentrates on how primitive image features (line segments) are understood by the model, and how they can provide an expectation for lower levels. The understanding of more abstract features in higher levels of the neural architecture is not treated in this paper, it will be the subject of further research.

In the visual system a variety of neurons can be found from ganglion cells through LGN cells to cortical neurons, each responding to different preferred afferent stimulation. The preferred stimulation can be described by the properties of the receptive field of a neuron, as described in Section 2.

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 receptive field 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. The sum of the values of the filter matrix have to be zero so that no constant component is added to the result. The matrix as a non-directional derivative filter should be symmetric along all the axes. These two constraints explain the choice of the filter matrix:

$$M_1 = \begin{pmatrix} -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} \\ -\frac{1}{8} & 1 & -\frac{1}{8} \\ -\frac{1}{8} & -\frac{1}{8} & -\frac{1}{8} \end{pmatrix} \tag{3.1}$$

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.

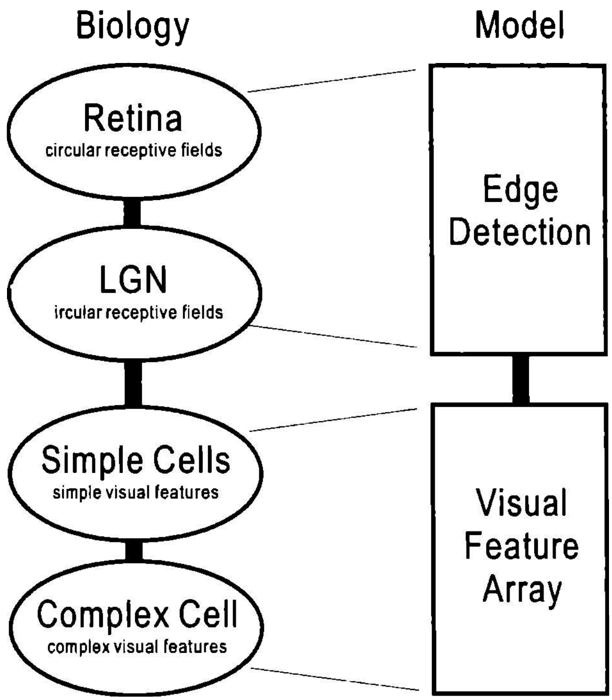


Figure 4. The biological system and the components of the proposed model that cover the biological functionalities.

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. 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 correspondence between the biological functionalities and the model components that cover them are shown in Figure 4.

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, as explained in Section 2. 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.

As a result of the edge detection an edge detected image is available in the matrix I where

$$I \in \mathbb{R}^{n \times m}, \tag{3.2}$$

n and m representing the image dimensions. The elements of the matrix I are bounded, such that

$$I_{i,j} \in [0,1], \tag{3.3}$$

where $I_{i,j}$ represents the pixel in the i^{th} row and j^{th} column of the matrix I.

Similarly to the visual cortex, several different features can be extracted from the edge detected image I. The extraction of the features begins with the longest line segments, those spanning through the largest angle in the visual field, and thus causing activation in the largest number of ganglion cells, or pixels in the context of a CCD imager. 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

$$\forall i,j,k \quad (F_k)_{i,j} \in \{0;1\}, \tag{3.4}$$

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 be extracted are ordered by the number of pixels they contain in order to ensure that

$$\mathcal{F}_k \supseteq \mathcal{F}_l, k < l, \tag{3.5}$$

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:

$$\mathcal{V} \in \mathbb{R}^{n \times m \times r} \tag{3.6}$$

For the tensor $\mathcal{V}$ we introduce the notion of *Visual Feature Array* or *VFA*, where r represents the total number of visual features extracted from the image. 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. The representation shown in Figure 2 shows that the neurons tuned to different visual features in the visual cortex are organized in a rather sophisticated system. In the VFA the same 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.

Let's take a closer look on the third dimension of the VFA.

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. There are other visual features in the brain, such as sensitivity to spatial frequency, eye preference or binocular depth cues, but these features are not included in our model yet. 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

$$\alpha \in [0 \ldots \pi], \alpha = k \cdot \theta, k \in \mathbb{N}, \tag{3.7}$$

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:

$$r = \frac{\pi}{\theta} \quad h. \tag{3.8}$$

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.

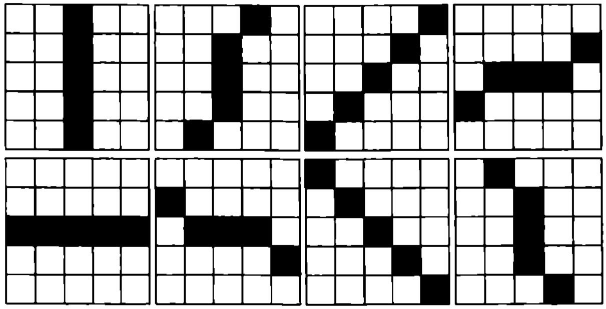


Figure 5. The matrices in the model that represent the receptive fields of cortical orientation tuned end-inhibited cells of 5-pixel-length.

In the visual cortex there are receptive field characteristics that actually define the visual feature the particular neuron is responsive to, as described in section 1 and shown on the bottom of Figure 1. 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 have chosen 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 for all the possible five-pixel-long lines are shown in Figure 5.

Once the VFA is constructed from the edge detected image I, it can be subjected to further transformations in order to extract more abstract features from it. As it was described above, one layer in the VFA contains the pixels that belong to a well-defined feature (i.e. a line of a certain length and orientation). A grouping transformation can be defined on the VFA, which unifies the layers and thus groups the features of the VFA according to different feature properties.

Two basic grouping transformations are defined:

$$\mathcal{G}_o \quad \mathcal{V} \to \mathcal{V}^{(o)}, \tag{3.9}$$

and

$$\mathcal{G}_l \quad \mathcal{V} \to \mathcal{V}^{(l)} \tag{3.10}$$

The result of $\mathcal{G}_o(\mathcal{V})$ is $\mathcal{V}^{(o)}$, which contains iso-orientation layers, where all the features of the same orientation are present in one layer. This step was inspired by the iso-orientation columns found in the primary visual cortex by Hubel and Wiesel, as described in the section 2. On the other hand such a grouping transformation is necessary to find the line crossings and vertices in the VFA.

The result of $\mathcal{G}_l(\mathcal{V})$ is $\mathcal{V}^{(l)}$, which contains iso-length layers, where all the features of the same length are present in one layer. This transformation can be useful in segmenting short and long line segments from each other. Short line segments of arbitrary orientation are usually the components of textures of natural objects (trees or bushes). Longer, parallel and orthogonal line segments usually make part of artificial (man made) objects or scenes, such as an urban scene.

The nodes of $\mathcal{V}$, $\mathcal{V}^{(o)}$ or $\mathcal{V}^{(l)}$ can send their outputs to higher or lower levels of the neural hierarchy. Sending the inputs further up allows further transformations and the recognition of more complex features or objects. Sending the output back in the neural hierarchy allows feedback and reinforcement in lower neural structures.

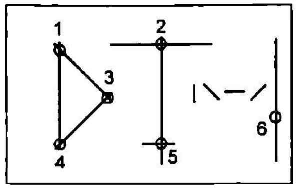
For instance, line crossing and vertex detection can easily be done by sending the output of the $\mathcal{V}^{(o)}$ neurons further *up* in the neural hierarchy. A layer of neurons organized in a two dimensional matrix $C \in \mathbb{R}^{n \times m}$ receives input from $\mathcal{V}^{(o)}$ and provides an output according to the function f as

$$C_{i,j} = f(\mathcal{V}^{(o)}_{i,j,1}, \mathcal{V}^{(o)}_{i,j,2}, \ldots, \mathcal{V}^{(o)}_{i,j,o}), \tag{3.11}$$

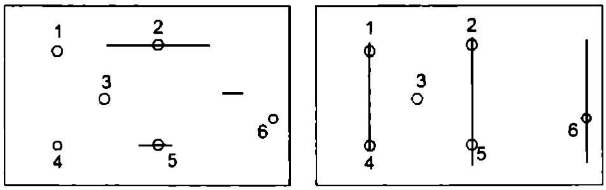
where o is the number of line orientations. A neuron in $C_{i,j}$ will have a high output if it receives more than one active inputs, meaning that there is more than one differently oriented line at the same image location. Figure 6 shows an example, where the red circles are neurons in C indicating line crossings, while the blue circle indicates no line crossing.

It is to note that the use of the original VFA $\mathcal{V}$ is not appropriate in finding the vertices, because two colinear line segments may overlap each other. If so, their overlap will be considered as a vertex, which is not desirable. If the $\mathcal{V}^{(o)}$ is used, only one neuron from one position and one orientation will send input to C, and thus two overlapping collinear line segments will not activate C.

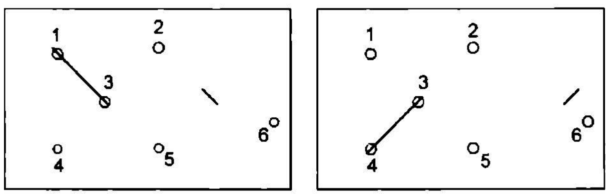
One can consider a third type of grouping transformation on the VFA, which simply groups all the layers into one final layer containing all the extracted features. This transformation equals sending the output of the VFA neurons *down* in the neural hierarchy, and can be used to reconstruct an image by reactivating the pixels that belong to 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. The



(a)



(b) (c)



(d) (e)

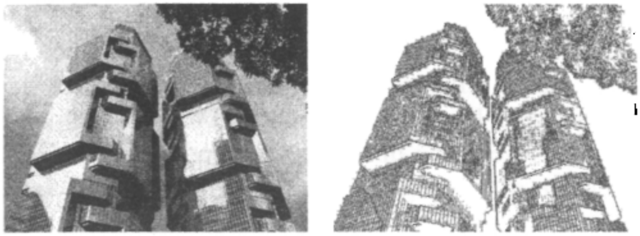
Figure 6. Detection of line segment crossing and vertices. Image (a) shows the edge detected image, images (b)-(e) show the results of four grouping transformations of the VFA $(\mathcal{V}_1^{(o)}, \mathcal{V}_2^{(o)}, \ldots, \mathcal{V}_4^{(o)})$. Red circles indicate line crossings, the blue circle shows an example of no crossing. Red circles contain active neurons in the VFA in more than one group, while the blue circle contains active neurons only on (c).

comparison or merging of the reconstructed and the original edge detected image actually adds information to the original image.

We introduce 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.

4. Model evaluation, results

The proposed model has two important advantages compared to classical solutions. By virtue of the simple but numerous computational units (neurons)



(a) (b)

Figure 7. Original test image (a) and the result of the primary edge detection (b)

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 this paper only a computer based simulation of the model is presented, which allowed to evaluate its functionalities. The evaluation of the performance was however not possible due to the lack of a hardware implementation. In the rest of this section the different sections of the information flow within the model will be presented.

The input test image used to evaluate the model is shown in Figure 7a. This image is subjected to a primary edge detection as discussed in section 3. 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 7b.

The edge-detected image within the 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 transformations with 3, 9 and 33 pixel-long segments are shown in Figure 8. Using the grouping transformation $\mathcal{G}_l$ that yields the VFA $\mathcal{V}^{(l)}$, having 5 layers each of them containing the line segments of all the possible orientations of a certain length. Three out of these five layers are shown on Figure 8.

The union of the five layers of $\mathcal{V}^{(l)}$ 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, which was not recognized as a visual feature (a line segment of certain length and orientation). The final,

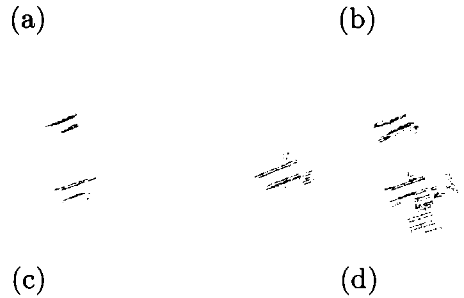


Figure 8. The reconstruction of the edge-detected image from line segments of 3 pixels (a), 9 pixels (b), 33 pixels (c) and the combination of all sub-VFAs (d)

fully reconstructed, negative filtered image composed from the five layers of $\mathcal{V}^{(l)}$ is shown in Figure 8d.

5. Hardware realization

One of the most advantageous properties of the proposed architecture is its native parallelism. A software implementation running on the fastest *von Neumann*-based processor cannot provide fast, $O(1)$ time responses even if they are extended with Single Instruction Multiple Data (SIMD) instruction sets like Intel's Streaming SIMD Extensions (SSE). Parallel processing with multiple simple computational elements, on the other hand, can provide tremendous speedups. The Field-Programmable Gate Array (FPGA) is such a microelectronic device the programming logic of which can be set up according to the users' needs, and some models even allow to be reconfigured during operation time. Thus, the proposed architecture can be implemented in an FPGA, and then can be used as a coprocessor or accelerator card in a PC environment to solve dedicated tasks. Moreover, it can be a stand-alone image processing device that solves the task without the execution of any conventional algorithm.

The proposed architecture requires about 10 000 computational elements to perform the edge detection on a 100×100 pixel image. Then for each direction of each length another 10 000 computational elements are necessary, that is in total 124 × 10 000. The state of the art FPGA has about 6 000 000 logic cells that is sufficient for about 100–200 000 computational elements. This number copes with what is necessary, while the processed image is still relatively small. However, the famous Moore's Law also applies to FPGAs saying that in about every two years the number of transistors on a silicon chip doubles, thus the number of logical cells is expected to double, too. In addition, some of the modern FPGAs also have the capability of being reprogrammed in runtime. Applying this feature allows the use of only one chip for the processing that can be done by reprogramming the architecture for each task, sacrificing extra processing time. In conclusion, a primary visual cortex based image contour detector chip can be realized in near future by some compromises.

A simple (low resolution) version of the model is being implemented in an FPGA. A serious bottleneck in this solution is the small number of parallel input/output data that can transmitted to and from the FPGA. Apart from this problem, the implementation will give ground to test and evaluate the model operating on a dedicated hardware.

6. 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 a hierarchical classification of visual features. 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.

It is important to note that the goal of this model was not to achieve a qualitative advance in image contour detection, but to make the first step towards a biology and cognitive science inspired vision system. The new approach is expected to lead to a cognitive system overpassing the performance of classical computational methods. Meanwhile, the quality of the contour detection achieved by the proposed model is comparable to classical edge detection algorithms.

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] HUBEL, D.: *Eye, Brain and Vision.* W.H. Freeman & Company, 1995, ISBN 0716760096.

[2] HUBEL, D. H. and WIESEL, T. N.: Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiology*, **160**, (1962), 106–154.

[3] GRANLUND, G. H. and KNUTSSON, H.: Signal processing for computer vision. *Kluwe Academic Publishers.*

[4] KOENDERINK, J. J. and VAN DOORN, A. J.: The structure of images. *Biol. Cybernet.*, **50**, (1984), 363–370.

[5] LIFSHITZ, L. M.: *Image segmentation via multiresolution extrema following.* Tech. rep., University of North Carolina, 1987.

[6] WITKIN, A.: Scale-space filtering. In *Proc. 8th Internat. Joint Conf. Artificial Intelligence*, 1983, pp. 1019–1022.

[7] GRANLUND, G. H.: In search of a general picture processing operator. *Comput. Graphics Image Process.*, **8**(2), (1978), 155–178.

[8] POLANS, A., BAEHR, W., and PALCZEWSKI, K.: Turned on by Ca^{2+}! the physiology and pathology of Ca^{2+}-binding proteins in the retina. *Trends in Neurosciences*, **19**(12), (1996), 547–554.

[9] DARTNALL, H. J. A., BOWMAKER, J. K., and MOLLON, J. D.: Human visual pigments: Microspectrophotometric results from the eyes of seven persons. *Proc. of Royal Society of London, B.*, **220**, (1983), 115–130.

[10] BARLOW, H. B.: Summation and inhibition of the frog's retina. *J. Physiology*, **119**, (1953), 69–88.

[11] KUFFLER, S. W Discharge patterns and functional organization of mammalian retina. *J. Neurophysiology*, **16**, (1953), 37–68.

[12] TAO, L. SHELLEY, M. MCLAUGHLIN, D., and SHAPLEY, R.: An egalitarian network model for the emergence of simple and complex cells in visual cortex. *PNAS*, **101**(1), (2004), 366–371.

[13] HUBEL, D. H. and WIESEL, T. N.: Receptive fields and functional architecture of monkey striate cortex. *J. Physiology*, **195**, (1968), 215–243.

[14] GRINVALD, A., MALONEK, D., SHMUEL, A., GLASER, D., VANZETTA, I., SHTOYERMAN, E., SHOHAM, D., and ARIELI, A.: *Imaging of Neuronal Activity.* Cold Spring Harbor Laboratory, 1999.

[15] HUBEL, D. H. and WIESEL, T. N.: Receptive field and functional architecture in two nonstriate visual areas (18–19) of the cat. *J. Neurophysiology*, **28**, (1965), 229–289.

[16] GRINVALD, A., MALONEK, D., SHMUEL, A., GLASER, D., VANZETTA, I. SHTOYERMAN, E., SHOHAM, D., and ARIELI, A.: *Imaging of Neuronal Activity*, chap. Intrinsic signal imaging in the neocortex, pp. 1–17. Cold Spring Harbor Laboratory, 1999.