A Neuromorphic Architecture for Visual Processing

Giacomo Indiveri, Luigi Raffo, Silvio P. Sabatini and Giacomo M. Bisio
DIBE-Department of Biophysical and Electronic Engineering
University of Genova-Via Opera Pia 11/A-16145 Genova-ITALY
ph: +39 10 3532163, fax: +39 10 3532795, e_mail: luigi@dibe.unige.it

Abstract

Early vision tasks and in particular texture analysis are solved introducing a hierarchical neuromorphic architecture organized in three layers with inter- and intra-layer local connections. These connections allow us to integrate the different specific computations implemented in the single layers.

1 Introduction

Researchers in machine vision have typically relied on general-purpose architectures to implement a variety of algorithms for the processing of visual information. In this approach, notwithstanding the continuous advances of microelectronic technology, the simulation of machine vision algorithms often remains well below the expected performance level. These difficulties are overcome in visual systems of humans and animals that accomplish complex computational tasks such as image analysis, segmentation and classification are accomplished with a surprising speed and efficiency. In this framework, here we investigate the computational capabilities of neuromorphic architectures inspired by biological modeling of the visual cortex. Specifically, we present a three layer hierarchical neural network architecture to be used in early vision processing tasks (e.g. texture segregation, edge and contour enhancement). Taking into account both the linear properties of visual cortical cells and the non-linear properties of intracortical processing, we defined the structure and the functionality of a neural model of the visual cortex applied to machine vision.

2 Neurophysiological Basis for Perception

In mammals, ascending the visual pathway, cells perform more and more complex functions going from luminous contrast sensitivity, in the retina, up to symbolic information extraction, in the visual cortex. The pattern of activity that best activates a response of a cortical cell is called its *trigger feature*. The area of the visual field in which this pattern best elicits the neural response is called the *receptive field* of the cell. Cortical cells respond optimally to moving or flashing oriented bars and gratings. According to the increasing complexity of trigger features, one can distinguish three classes of cortical cells: *simple cells*, *complex cells* and *hypercomplex cells* [1].

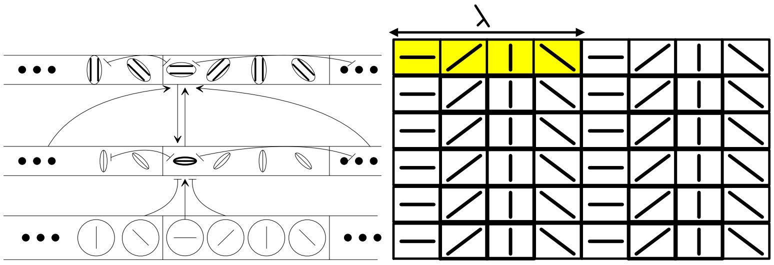
Simple cells responds almost linearly to visual stimulation within their receptive fields [2]. A linear transfer function can be associated to such cells and their excitation can be modeled as a convolution between the input image and a two dimensional function, often referred to as its *receptive field profile*, completely characterizing the behavior of the cell. To describe the behavior of complex and hypercomplex cells the concept of receptive field profile is not sufficient and it is necessary to introduce non-linear mechanisms of interaction [3].

All the three classes of cortical cells, however, have a common feature: many neurophysiological experiment have revealed that all cortical cells' receptive fields are composed of elongated ON (sensitive to light stimuli on darker background) and OFF (dark stimuli on brighter background) subregions that render them orientation-selective.

The functionality of the visual cortex depends not only on the complexity of its component cells but also on their spatial organization. A great deal of neurophysiological measurements [4], have revealed that orientation-selective cells are spatially organized. Assemblies of cells with near the same orientation preference are grouped together forming the so-called *columns*, and, on a larger scale, the orientation preference of columns is distributed in a regular periodic pattern. The set of orientation present in a spatial period is called *hypercolumn* [4]. The particular disposition of these cells on the cortical surface determines the cortical *orientation map*.

3 Network Architecture

Each cell acts as a local feature extractor in its receptive field; simple, complex and hypercomplex cells can be functionally organized in a three-layers hierarchical network (see Fig. 1(a)); the output of each cell depends on both inter-layer and intra-layer connections; the manner in which cortical cells are arranged in the layers, together with intra-layer connections, is instrumental in enhancing feature selectivity. Each layer carries out a specific functional task (e.g. the simple layer performs feature extraction, the complex layer performs feature enhancement and the final layer reduces noise and segregates different features).



(a) (b)

Figure 1: (a) Inter layer connections: in the bottom layer columns of simple neurons are connected with complex neurons. In the top layer hypercomplex neurons both receive and send excitatory signals from the middle layer; (b) Discrete form of a cortical orientation map. In gray background is an hypercolumn of width $\lambda = 4$.

While neurons belonging to the simple layer have a linear transfer function, complex and hypercomplex neurons are characterized by a sigmoid function $g(.)$. The output $z(\mathbf{u})$ of complex and hypercomplex neurons is:

$$z(\mathbf{u}) = g(e(\mathbf{u})) \tag{1}$$

where $\mathbf{u} = (u_1, u_2)$ is the coordinate vector on the cortical plane and $e(\mathbf{u})$ is the input excitation.

In each layer, neurons, which have receptive fields selective to oriented featural elements, are arranged in a cortical map (see Fig. 1(b)).

The Simple Layer In this layer simple neurons are organized in columns composed of four neurons with receptive field profile with different sign and symmetry but with the same orientation, spatial frequency, spatial extension and position. Cells belonging to columns of the same hypercolumn have receptive fields centered on the same region of the image, that overlap almost completely. Cells belonging to adjacent hypercolumns have receptive fields that overlap only partially.

The response $e^p(\mathbf{u})$ of a neuron at point $\mathbf{u} = (u_1, u_2)$ in the layer and position p ($p = 1 \ldots 4$)in its column, can be formulated in integral form as:

$$e^p(\mathbf{u}) = \int_{\mathcal{I}} (I(\mathbf{x}) - \bar{I}(\mathbf{x})) w^p(\mathbf{x}; \mathbf{u}) d\mathbf{x} \tag{2}$$

where $I(\mathbf{x})$ represents the intensity of the pixel at point $\mathbf{x} = (x_1, x_2)$ in the image plane $\mathcal{I}$, $\bar{I}(\mathbf{x})$ is the local average luminosity of the image and the kernel $w^p(\mathbf{x}; \mathbf{u})$ describes the neuron's receptive field profile. The dependence of w^p on the cortical coordinates $\mathbf{u}$ reflects the changes in orientation and position of receptive fields on the image plane. The operation $I(\mathbf{x}) - \bar{I}(\mathbf{x})$ is a rough model of the retinal adaptability to background luminance. The particular choice of the two-dimensional function is not critical for the performance of the network. In the present model, we chose to model the receptive field profile with a *DOOG* function [5] defined as follows:

$$\tilde{w}(x_1, x_2) = A\, G(0, 0, \sigma_1, \sigma_2) + B\, G(0, x_2^c, \sigma_1, \sigma_2) + C\, G(0, -x_2^c, \sigma_1, \sigma_2) \tag{3}$$

where $G(x_1^c, x_2^c, \sigma_1, \sigma_2) = \exp(-[(x_1 - x_1^c)^2/\sigma_1^2 + (x_2 - x_2^c)^2/\sigma_2^2])$ is a Gaussian function and A, B, C, σ_1 and σ_2 are the parameters determining the shape of the DOOG function (see Fig. 2). The coordinates (x_1^c, x_2^c) denote the center of the receptive field.

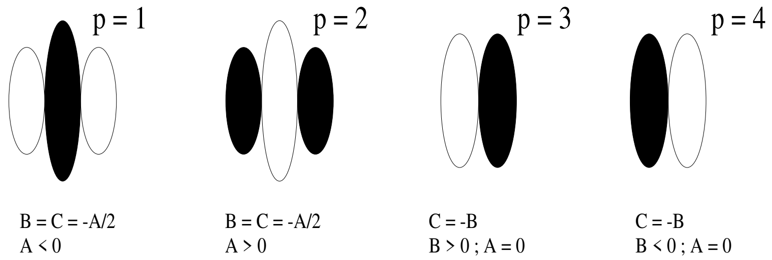


Figure 2: Shapes and parameters of the four types of receptive fields used for simple neurons.

The four even/odd symmetric receptive fields have the same spatial frequency selectivity. Moreover, $e^1(\mathbf{u}) = -e^2(\mathbf{u})$ and $e^3(\mathbf{u}) = -e^4(\mathbf{u})$, and $e^1(\mathbf{u})$ is orthogonal to $e^3(\mathbf{u})$. The interaction among the four neurons of a column can be functionally described as a maximum operation over their excitations: Taking into account the relationships among the four receptive fields, the maximum operation is a sort of energy measurement of the column excitations [6].

The Complex Layer The functionality of neurons belonging to this layer arises from a combination of outputs of the preceding layer and from functional loops within the complex layer itself and with the subsequent hypercomplex layer. In this way, a complex neuron receiving input from the corresponding column becomes selective to oriented bars of any luminous contrast and of any phase:

The excitation $e_c(\mathbf{u})$ in the complex layer is the result of four contributions due to direct excitation z_s from the corresponding column in the simple layer; feed forward inhibition from a set $M(\mathbf{u})$ of columns in the simple layer; recurrent inhibition from a set $N_c(\mathbf{u})$ of complex neurons; positive feedback z_h from the corresponding neuron of the hypercomplex layer:

$$e_c(\mathbf{u}) = z_s(\mathbf{u}) - w_{sc} \sum_{\mathbf{u}' \in M(\mathbf{u})} z_s(\mathbf{u}') - w_{cc} \sum_{\mathbf{u}' \in N_c(\mathbf{u})} z_c(\mathbf{u}') + w_{hc} z_h(\mathbf{u}) \tag{4}$$

where w_{sc}, w_{cc} and w_{hc} represent the connection weight from simple to complex (feed-forward), from complex to complex (intralayer) and from hypercomplex to complex neurons (feedback) respectively. The sets $M(\mathbf{u})$ and $N_c(\mathbf{u})$ are properly chosen to increase orientation contrast, also reducing the noise sensitivity. In our model the feed-forward inhibition does not depend on the position (and thus on the orientation) of the neuron considered in the layer; the columns belonging to $M(\mathbf{u})$ are the ones located in (u_1+1, u_2) and (u_1-1, u_2). Concerning recurrent inhibition, we implemented a lateral *functional cross orientation inhibition* schema [7]; $N_c(\mathbf{u})$ depends on the orientation preference θ of the target neuron ($N_c(\mathbf{u}) = N_c(\theta(\mathbf{u}))$), where θ is the orientation preference of the neuron at the plane coordinate $\mathbf{u}$. Specifically, a neuron with orientation preference θ receives inhibitory input from two complex neurons selective to $\theta + 90$, that belong to the two closest hypercolumns that lay along an axis orthogonal to θ.

The Hypercomplex Layer The excitation of hypercomplex neurons results from two contributions: the feed-forward actions from a set $L(\mathbf{u}) = L^+(\mathbf{u}) \cup L^-(\mathbf{u})$ of neurons in the complex layer; and the cross-orientation inhibition from a set $N_h(\mathbf{u})$ of neurons in the hypercomplex layer:

$$e_h(\mathbf{u}) = w_{ch}\left[\sum_{\mathbf{u}'\in L^+(\mathbf{u})} z_c(\mathbf{u}') - \sum_{\mathbf{u}'\in L^-(\mathbf{u})} z_c(\mathbf{u}')\right] + w_{hh}\sum_{\mathbf{u}'\in N_h(\mathbf{u})} z_h(\mathbf{u}') \tag{5}$$

where w_{hh} and w_{ch} define the connection weight for hypercomplex to hypercomplex connections and for complex to hypercomplex connections respectively. The set $L(\mathbf{u})$ depends on the orientation preference of the target neuron. Specifically, the connection schema is defined as follows: if the target neuron is selective to θ then the complex neurons, that provide the input, belong to the neighboring hypercolumns that lay on an axis parallel to θ and are selective to θ. The number of neighboring hypercolumns involved in the schema goes from three up to seven. The neurons in $L(\mathbf{u})$ that lay within a radius ρ define the subset $L^+(\mathbf{u})$ and provide an excitation for the target neuron, conversely those outside this radius define the subset $L^-(\mathbf{u})$ and provide inhibition.

Moreover the feedforward action of $L(\mathbf{u})$ neurons combines with the positive feedback, of weight w_{hc}, that goes from hypercomplex to complex neurons creating a *cooperative loop.*

4 Application to Preattentive Vision

We will show, that the network here proposed achieves satisfactory performance in a variety of experiments related to preattentive vision tasks (edge detection, texture analysis and texture boundaries). The results reported below, related to texture analysis, are all derived with the following architectural specification: $\sigma_1/\sigma_2 = 3$, an overlap fraction between adjacent receptive fields of 0.75, four different orientation selectivities ($\lambda = 4$) and three different spatial frequencies (i.e., three different sizes for the receptive fields of the columns).

In Fig. 3(a) an image with four different natural textures arranged in quadrants, is shown. The upper two quadrants contain oriented textural elements with high frequency characteristics, while the bottom quadrants contain textural elements with lower frequency components.

The response of the network is illustrated in Figs. 3(c) and (d). Each image represents the activity of neurons selective to a specific orientation, the luminous intensity of the pixel codes the activity of the corresponding neuron. If the pixel is light then the neuron is active, if the pixel is dark the neuron is inhibited and if a pixel is of intermediate value then the corresponding neuron is silent (the neuron is not selective to the stimulus present in its receptive field). Fig. 3(c) shows the output of the system for neurons with small receptive fields (high frequency tuning): in the first image (on the left) the output of the hypercomplex neurons selective to 0° is shown, in the second image the output of hypercomplex neurons selective to 45° is shown, in the third the output of neurons selective to 90° and in the last image (the rightmost one) the output of the neurons selective to 135°. The textures in the upper quadrants are well distinguished, while the textures in the lower quadrants are not perfectly evidenced. Similarly, Fig. 3(d) shows the output of the system for neurons with larger receptive fields (intermediate frequency tuning), for which the two textures in the lower part of the image are well separated.

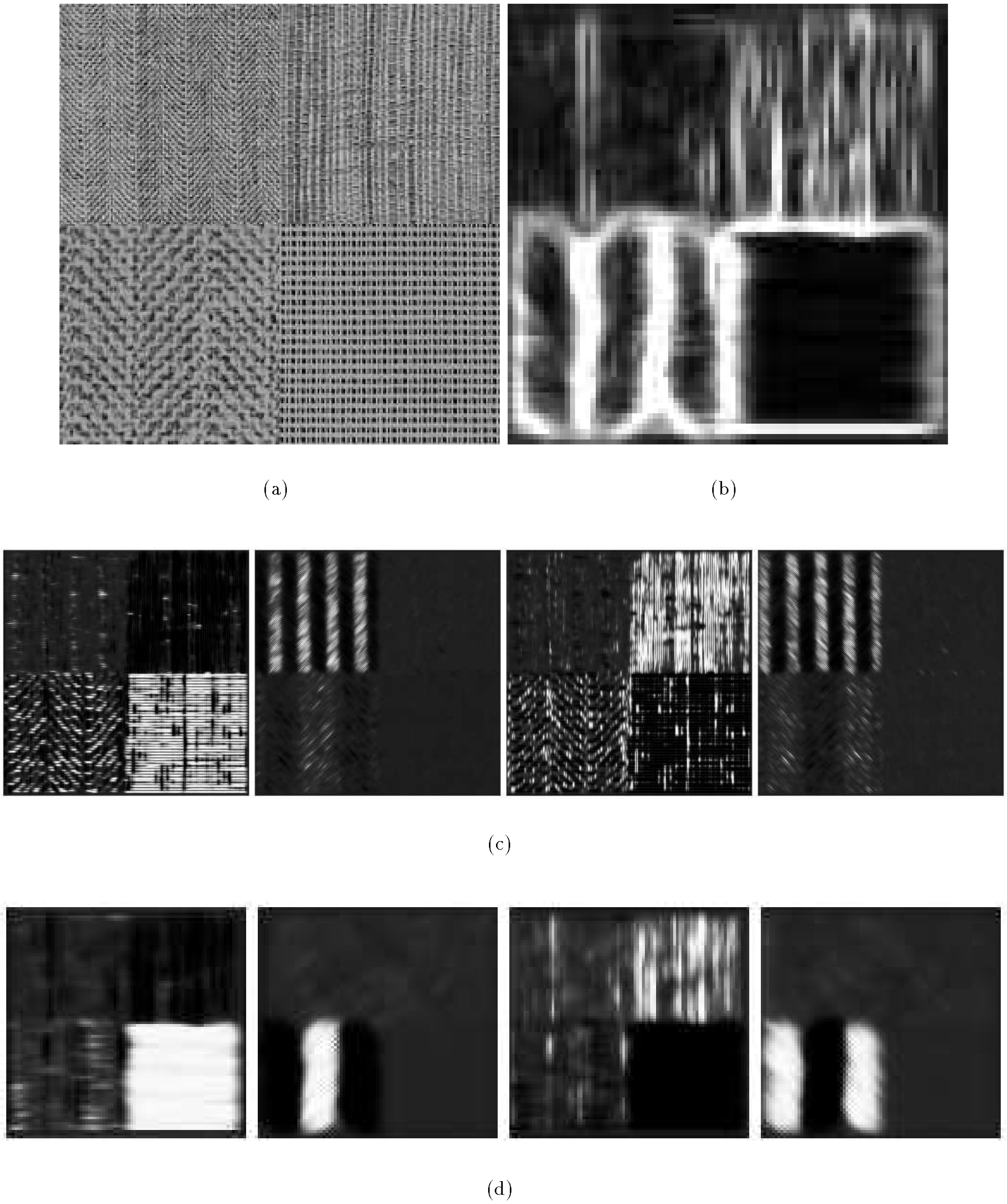


Figure 3: (a) Image of natural textures, (b) texture boundaries obtained by applying, for a second time, the convolution stage (simple layer stage) on the output images merged together of (a), (c) Output of the system for neurons with small receptive fields ($\sigma_x = 0.75$), (d) Output of the system for neurons with medium receptive fields ($\sigma_x = 2.25$).

Exploiting the advantage of having an architecture that integrates different operators we can obtain boundaries of textured images. Once obtained the representations relative to texture segregation (output of the hypercomplex layer), texture boundaries can be computed by processing again with the simple layer of the architecture an input image, obtained by merging together the output of the hypercomplex layer (see Fig. 3(b)).

5 Conclusion

In this paper we proposed a novel approach trying to unify the computational paradigms addressing early vision within a single framework: low and intermediate-level tasks of visual perception can be based on multiple and cooperating simple agent architectures, formulated as networks of formal neurons organized in a three layer neural network locally connected. In each layer there is a specific intermediate representation of the image: the first layer extracts oriented featural elements, while the second and third one are used for controlling and coordinating interactions among features. This approach provides a link between the computational nature of early vision mechanisms, the structure of the algorithms implementing them and the parallel hardware that can be used for efficient visual information processing [8] [9] [10].

References

[1] D. Hubel and T. Wiesel, "Receptive fields and functional architecture in the two nonstriate visual areas (18 and 19) of the cat," *J. Neurophysiol.*, vol. 160, pp. 106–154, 1965.

[2] J. Kulikowski and T. Vidyasagar, "Space and spatial frequency: Analysis and representation in the macaque striate cortex," *Exp. Brain Res.*, vol. 64, pp. 5–18, 1986.

[3] R. Emerson, M. Korenberg, and M. Citron, "Identification of complex-cell intensive nonlinearities in a cascade model of cat visual cortex," *Biol. Cybern.*, vol. 66, pp. 291–300, 1992.

[4] D. Hubel and T. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *J. Physiol.*, vol. 160, pp. 106–154, 1962.

[5] J. Malik and P. Perona, "Preattentive texture discrimination with early vision mechanisms," *J. Opt. Soc. Amer.*, vol. 7, pp. 923–932, 1990.

[6] J. Bergen and M. Landy, "Computational modeling of visual texture segregation," in *Computational Models of Visual Processing* (M. Landy and J. Movshon, eds.), pp. 253–271, MIT Press, 1991.

[7] G. Bisio, D. Caviglia, G. Indiveri, L. Raffo, and S. Sabatini, "A neural network architectural model of visual cortical cells for texture segregation," in *Proc. Int. Conf. on Neural Networks*, (San Francisco, CA), March 1993.

[8] C. Mead, *Analog VLSI and Neural Systems.* Addison-Wesley, Reading, 1989.

[9] H. Kobayashi, J. White, and A. Abidi, "An active resistor network for Gaussian filtering of images," *IEEE J. on Solid State Circuits*, vol. 26, pp. 738–748, May 1991.

[10] L. Raffo, S. Sabatini, D. Caviglia, and G. Bisio, "Anisotropic active resistor meshes for implementing image processing operators," *Electronics Letters*, vol. 29, pp. 960–961, May 1993.