A neural network for detection of orientation, velocity and direction of movement, based on physiological rules

G.E. LA CARA, M. RITROVATO, M. URSINO
Department of Electronics, Computer Science-and Systems
University of Bologna
viale Risorgimento 2, I 40136
BOLOGNA, ITALY

Abstract: - Aim of this work is to present two neural network models for detection of velocity, orientation and direction of movement in visual images. Both models mimic a single hypercolumn in the primary visual cortex. They differ as to the arrangement of inhibitory circuitry: in the first ("*anti-phase* inhibition model") inhibition is in phase opposition with excitation, but with a similar orientation tuning; in the second ("*in-phase* inhibition model"), inhibition is in phase with excitation, but with larger orientation tuning. Simulation results, performed by using bars with different length and motion direction, show that the models can explain velocity tuning, orientation tuning and direction selectivity of simple cells quite well with a suitable choice of intracortical synapses. The models can be used to test the hypothesis on the disposition of cortical synapses, and could provide practical tools in order to carry out a primary analysis of the movement detection of individual points in a visual scene.

Key-Words: movement detection, direction selectivity, velocity tuning, neural network, simple cells, model

1 Introduction

Detection of the velocity of movement of individual points is one of the fundamental tasks that any vision system must be able to cope with. In the primary visual cortex (V1), this task is first performed by the so-called "simple cells", which are sensitive to orientation, velocity and direction of motion of input stimuli within a particular region of the visual space (named, the receptive field, RF). Analysis of the neural circuits by which the cerebral cortex may detect movement is of the primary importance, not only to reach a deeper understanding of the neurophysiology of V1, but also to design artificial neural networks devoted to vision problems.

In the physiological literature, the response of simple cells has been studied, by using different alternative stimuli: they include drifting gratings of various orientation and spatial and temporal frequencies [1], moving bars [2,3], or flashing bars [4,5,6]. These stimuli are aimed at analysing the main spatio-temporal properties of cortical visual cells (such as orientation and direction preference, velocity tuning, or the organisation of the receptive field) and at discovering the possible mechanisms (thalamic or cortical) involved in their aetiology. Despite the great number of experimental studies performed, and mathematical models proposed, the exact intracortical circuit responsible for velocity selectivity is still largely hypothetical.

The aim of this work is to investigate the properties of two alternative neural network models for velocity sensitivity and direction selectivity, based on physiological considerations. These models share several fundamental aspects, and differ only as to the particular arrangement of intracortical inhibition. This choice is justified by the observation that the exact arrangement of inhibition in the primary visual cortex is still a matter of debate, and alternative intracortical circuits have been proposed, without definite conclusions.

This study has been conceived with two fundamental purposes: i) to reach a deeper understanding on the possible mechanisms by which the primary visual cortex extracts some essential features of the input image, such as orientation, velocity, direction. This may be important to improve neurophysiological knowledge; ii) to represent a possible initial step usable within artificial vision systems. For instance, the proposed neural networks may be exploited as a detector of movement, able to signal the existence of moving objects with a particular orientation, a particular direction, and within a specific velocity range.

The paper is structured as follows. In the subsequent section, the main aspects of the two models are quantitatively presented, laying emphasis on their neurophysiologic justifications. The third section presents the results obtained by stimulating the models with moving bars having different velocities, orientation and length. Finally, results are

discussed in the last section, laying emphasis on the differences between the two models and on their possible use within artificial systems for movement detection.

2 Problem Formulation

In the present paper two different models have been implemented (*in-phase* and *anti-phase* model) which differ as to the disposition of the intracortical inhibitory circuit. The models consider the architecture of a single hypercolumn, composed of 360 excitatory neurons and 90 inhibitory interneurons, parameterized by their preferred orientation. Two orientation angles differing by 180 degrees identify opposite direction of movements. In the models the output of neurons is represented by the firing rate (spikes/s).

Each model consists of two consecutive layers: a first layer simulates the response of the thalamic cells in the lateral geniculate nucleus (LGN), which represents the first step in the visual processing stream. A second layer mimics the elaboration in V1.

The LGN includes both ON-center and OFF-center cells. Their response is computed as a function of normalized luminance. We assume that these cells have a separable spatio-temporal RF. Hence, this is described as the product of a spatial term and a temporal term, i.e.

$$RF_{x,y}(i,j,t) = \varphi_{x,y}(i,j) \cdot \gamma_x(t - t_x) \qquad (1)$$

where t is time, $RF_{xy}(i,j,t)$ represents the spatio-temporal RF of a thalamic cell centered at position x,y of the retina, $\varphi_{x,y}(i,j)$ is its spatial component and $\gamma(t)$ the temporal component.

The spatial component of the RF at the coordinates (i,j) is described as the difference between two concentric Gaussian functions, having the same spatial constant in both directions. Hence for a LGN cell at position x, y:

$$\varphi_{x,y}(i,j) = A_1 e^{-\left[\frac{(i-x)^2+(j-y)^2}{r_1^2}\right]} - A_2 e^{-\left[\frac{(i-x)^2+(j-y)^2}{r_2^2}\right]} \qquad (2)$$

$\gamma(t)$ describes the response to a light impulse at time $t=0$ by means of a biphasic curve. T_{xy} is a time delay, which depends on the position of the LGN cell. Hence:

$$\gamma_x(t - t_x) = K_1 \times \frac{[c_1(t - t_x - t_{01})]^{n_1} e^{-c_{01}(t-t_{01})}}{n_1^{n_1} e^{-n_1}} +$$

$$- K_2 \times \frac{[c_2(t - t_x - t_{02})]^{n_2} e^{-c_{02}(t-t_{02})}}{n_2^{n_2} e^{-n_2}} \qquad (3)$$

The parameters in (2), (3) have the following meaning: r_1 and r_2 are the radii of the center and surround, respectively; A_1 and A_2 set the strength of the response in the center and surround. The values for the parameters K_1, c_1, t_{01}, n_1, K_2, c_2, t_{02}, n_2, c_{01}, c_{02} are given to fit experimental data about the LGN cells impulsive response [7]. Finally, the response of each LGN cell (Eq.1) is passed through a non-linear function, which exhibits lower threshold and upper saturation, according to the literature [8].

The RF of each simple cell in the hypercolumn is obtained from the convergence of excitatory input from 15 LGN cells, disposed to constitute a regular lattice oriented along the preferred orientation of the cell (Fig.1).

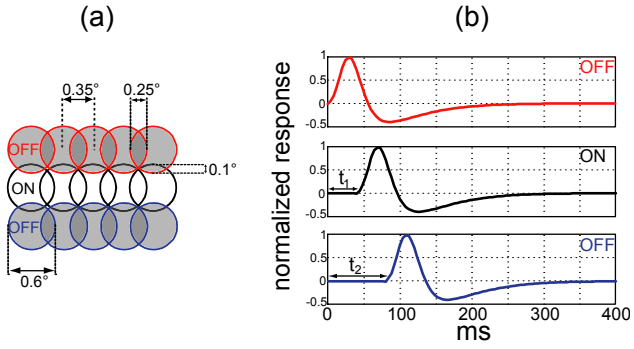

Fig.1. a) Thalamic input to a simple cell, arising from 15 thalamic cells arranged in a regular lattice. b) Thalamic impulsive response (normalized) for neurons belonging to the three different subfields.

The strength of synapses from these LGN cells to the target cell are described by means of a Gabor function, to have cells with a symmetric RF, composed of a central ON region plus two lateral OFF regions (or viceversa). All the parameters are assigned in acceptable agreement with physiological data [8], [9], [10].

A basic assumption of both models is that the time lag in the impulse response of LGN cells (i.e., parameter T_{xy} in Eq.3)) increases regularly from one-subregion to the next . The temporal patterns shown in Fig.1b represent the impulse responses of the non lagged neurons (T_{xy}=0 ms in (3)) for the upper subregion, and of two different lagged neurons (T_{xy}=40 ms and T_{xy}=80 ms) for the middle and lower subregions, respectively.

The convergence of the previous LGN responses determines an inseparable spatio-temporal RF . This causes a moderate direction preference for the

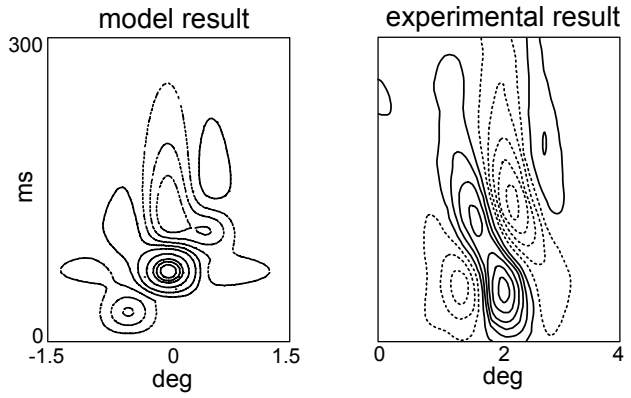

Figure 2. Cortical ON cell receptive fields: model result and experimental result [1]. Solid line: ON subfield; dashed line: OFF subfield.

thalamic input to cortical cells. Fig. 2 shows the RF obtained from the model compared with an RF obtained from an experimental registration [1].

The previous portion of the model considers just feedforward excitation from the LGN. The cortical neurons, however, do not only receive the LGN input but also excitatory synapses from the other cortical cells in the hypercolumn, and the inhibitory synapses from inhibitory interneurons. In order to describe the cortical circuitry, the following equations have been used:

$$\Delta V_c^{ON}(\theta,t) = \Delta V_{ct}^{ON}(\theta,t) + \Delta V_{ce}^{ON}(\theta,t) - \Delta V_{ci}^{ON}(\theta,t) \quad (4)$$

$$\tau \frac{dc^{ON}(\theta,t)}{dt} = -c^{ON}(\theta,t) + k_c\left[\Delta V_c^{ON}(\theta,t) - \upsilon\right]^+ \quad (5)$$

where $V(\theta,t)$ is the membrane potential of the cortical cell with orientation preference θ at time t, $c(\theta,t)$ is the output activity of the cortical cell at time t, $\Delta V_{ct}(\theta)$ is the change in membrane potential caused by the LGN input, $\Delta V_{ce}(\theta,t)$ and $\Delta V_{ci}(\theta,t)$ are the changes in membrane potential caused by excitatory and inhibitory intracortical connections, respectively. At equilibrium, the value of cortical cell activity is obtained by comparing the variation in membrane potential with a threshold, υ, using a single wave rectifier []+ which cuts negative values, and multiplying the value so obtained by a gain factor k_c.

In the model, intracortical excitation occurs via synaptic connections between neurons in the same hypercolumn, according to a feedback mechanism. Inhibition is realized through inhibitory interneurons, which modulate information from LGN cells to cortical cells via a feedforward mechanism (see Fig. 3). Two different circuits have been implemented, which differ only as to the inhibitory circuit. In the first, named *anti-phase* model, thalamo-cortical excitation and feedforward inhibition are arranged according to a "push-pull" schema, that is the interneurons send their inhibition to a simple cell with similar orientation preference, but opposite spatial phase of the RF (i.e., ON vs. OFF and OFF vs. ON). Conversely, in the *in-phase* model the RF of the inhibitory interneuron has the same spatial phase and orientation as those of its target simple cells (Fig.3).

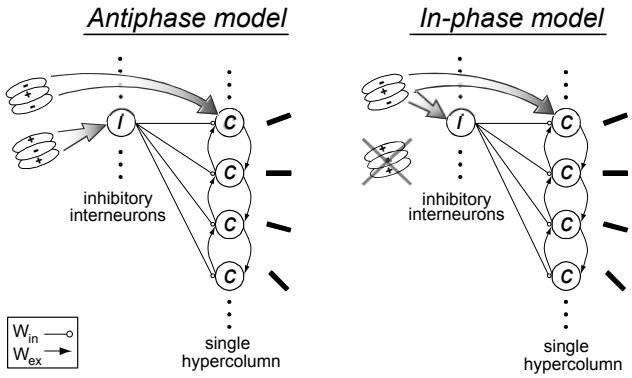

Fig.3. The arrangement of thalamic excitatory, cortical excitatory and cortical inhibitory inputs to simple cells in the hypercolumn.

The output activity of the inhibitory interneurons depends only of the LGN input, that is they do not receive intracortical synapses. Hence, the activity of the inhibitory interneuron with orientation preference φ can be written, for anti-phase model, as:

$$i^{OFF}(\varphi,t) = k_c\left[\Delta V_{ct}^{OFF}(\varphi,t) - \upsilon\right]^+ \quad (6)$$

where ΔV_{ct}^{OFF} is the LGN input in case of a RF with a central OFF region and two external ON regions.

By contrast, the activity of the inhibitory interneuron for the in-phase model is computed as:

$$i^{ON}(\varphi,t) = k_c\left[\Delta V_{ct}^{ON}(\varphi,t) - \upsilon\right]^+ \quad (7)$$

where ΔV_{ct}^{ON} is the same as in Eq.4.

According to physiological data [11], the synaptic strength between cortical cells, and between cortical cells and inhibitory interneurons decrease exponentially with the distance in orientation preference: two neurons with similar orientation preference have a strong synaptic connection, whereas neurons with a large difference in orientation preference have a weak synaptic connection.

Accordingly, for the anti-phase model the following equations have been used to compute the excitatory and inhibitory cortical inputs to a cortical cell with preferred orientation θ:

$$\Delta V_{ce}^{ON}(\theta,t) = \sum_{\phi} w_{ex}(\theta-\phi)c^{ON}(\phi) \quad (8)$$

$$\Delta V_{ci}^{ON}(\theta,t) = \sum_{\phi} w_{in}(\theta-\phi)i^{OFF}(\phi) \quad (9)$$

For the in-phase model Eq.9 is replaced by a similar equation, in which the term $i^{OFF}(\varphi)$ is substituted by $i^{ON}(\varphi)$. In Eqs.8 and 9 $w_{ex}(\theta-\varphi)$ represents the synapse between two excitatory neurons, and $(\theta-\varphi)$ is the distance between the preferred orientations of the post-synaptic and pre-synaptic cells; similarly, $w_{in}(\theta-\varphi)$ represents the synapse between an inhibitory interneuron and a target excitatory cortical cell (anti-phase: OFF vs ON; in-phase: ON vs ON).

3 Problem Solution

The mathematical model has been numerically solved on Pentium-based personal computer, using the software package Matlab®. We used the Euler method to integrate the differential equations. The input stimuli for the models are moving bars. In all cases, the response is evaluated as the maximal activity (spikes/s) evocated by the bar.

The first simulations were performed to analyze the velocity tuning curve of the neurons in both models, i.e., their activity vs. the velocity of the bar. To this end, we used bars with an optimal length, optimal orientation, which move in the orthogonal direction (i.e., in the direction perpendicular to that of orientation). Moreover, the width of the bar is generally adjusted to obtain the best response. The bar starts its movement well outside the RF of the cell, and terminates only when the RF has been completely crossed. In our simulations, the previous characteristics have been mimicked by using a vertical bar with length 4 deg and width 0.5 deg. The results are presented in Fig.4, and compared with experimental data in the literature [2].

Results show that the neuron response in both models is "velocity-tuned" in type, i.e., the response is maximal at a given optimal velocity, but falls to zero at low-velocities and at high velocities. The lower cut-off velocity is similar in both models (about 1deg/s); the upper cut-off velocity is greater in the anti-phase model (about 30 deg/s) than in the in-phase model (about 10 deg/s); i.e., the second model exhibits greater velocity tuning. The effect of orientation of the bar, and of movement direction is presented in the upper panels of Fig.5 for the two direction of movement of the bar (from 0 to 360 deg), which is always orthogonal to orientation. Hence, two angles which differ by 180 deg represent a bar with identical orientation, but moving in opposite directions. The Y-axis of this figure represents the response of a cell with optimal orientation 0 (or 180 deg), stimulated by bars with all possible orientations and directions of movements.

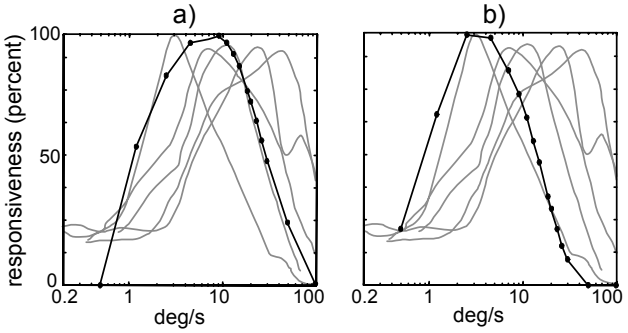

Figure 4. Velocity response curves: a) Anti-phase cortical output; b) In-phase cortical output. The gray lines represent experimental data [2], black lines represent model results.

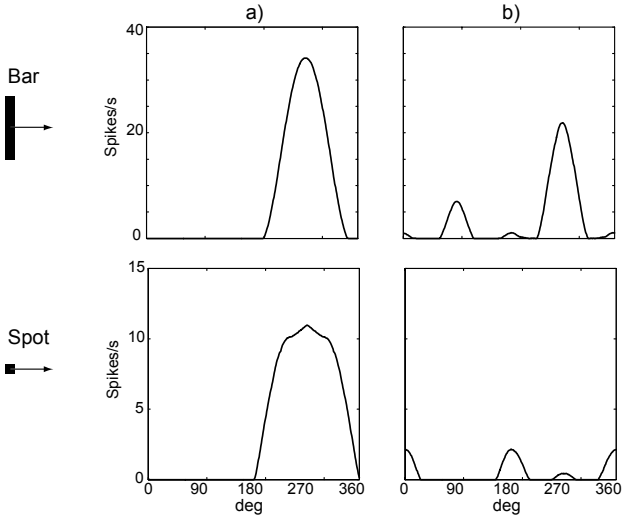

Figure 5. Activity of simple cells vs. orientation (i.e., the orientation tuning curve) in response to a moving bars. In the upper panels the bar has optimal length (4 deg) and moves through the orthogonal direction (i.e., in the direction perpendicular to that of its orientation) at velocity of 11 deg/s. In the lower panels the bar has length equal to 0.5 deg (i.e., spot light) and moves at velocity of 11 deg/s again. a) Anti-phase cortical output; b) In-phase cortical output.

Two fundamental aspects in Fig.5 deserve a comment. First, both models exhibit orientation selectivity, i.e., the response is maximal when the bar has an optimal orientation, equal to the orientation of the RF (0 deg), but the response rapidly declines with orientation difference. Second, both models exhibit a significant direction preference, i.e., the response is much greater in one direction than in the opposite one. Direction selectivity is much stronger

in the anti-phase model, leading to almost complete suppression of the response in the non-preferred direction, than in the in-phase model.

In order to quantify direction selectivity (i.e., the capacity to discriminate one direction of motion compared with the other one) an index commonly used in the literature is the direction index (DI), defined as follows:

$$DI = \frac{\text{response in Preferred Direction - response in Non Preferred Direction}}{\text{response in Preferred Direction}}$$

According to this definition, a *DI*=1 means a perfect direction selectivity (i.e., a strong response for an optimally oriented bar moving in one direction, but complete suppression of the response in the opposite direction). *DI*=0 means no direction selectivity, i.e., equal response in both directions. Fig.6 describes *DI* as a function of the velocity of the bar. The left panel represents Direction Index for the thalamic input (i.e., the input from the first layer of Fig.1), while the middle and right panels show the *DI* for the two input exhibits just a moderate direction selectivity, caused by the presence of time lags in the response of the thalamic cells. This moderate direction selectivity is then sharpened by the intracortical circuitry. Direction selectivity increases with the velocity of the bar, reaching an optimum for bars having a velocity in the range 5-10 deg/s, which is the velocity of maximal response for the cells. Further simulations have been performed to analyze the effect of a reduction in the length of the bar. Results are shown in the lower panels of Fig.5. These panels show that the two models behave in a very different way in response to a short bar or a spot of light (length less than 0.5 deg, i.e., a bar which is completely contained within the central region of the RF). In the anti-phase model, a reduction in the length of the bar simply causes a widening of the orientation tuning curve, i.e., the cell looses a part of its orientation selectivity, and becomes less selective to an alteration in the orientation of the bar. By contrast, in the in-phase model, the response to a spot light is completely suppressed at the optimal orientation and, moreover, we have the appearance of a response to different directions of movement (0 deg and 180 deg). In other words, in the in-phase model a spot light, provides its maximal response when it is moving along the major axis of the RF, along a direction orthogonal to the optimal direction for a long bar. The latter result agrees with experimental data by Worgotter and Eysel [4,5].

4 Conclusion

The present study aspires at developing a neural network model, based on physiological considerations, for detection of the main local features in an input image, with particular reference to movement. At present, the network implements a single hypercolumn in the primary cerebral cortex, i.e., it analyzes only a specific small portion of the image, with a dimension of the same order as the major axis of the receptive field (about 3 deg in the present work). Of course, the same elementary structure may be replicated at different points in the space, to have a complete description of the overall visual image.

The results obtained with the use of long bars (i.e., bars with a length greater than the long axis of the RF) demonstrate that the network is able to detect several important properties of the local pixel, i.e.: i) the presence of local discontinuities, such as bars or edges. In fact, i no local discontinuity occurs, i.e., if luminance is locally constant, the contributions of the positive and negative subregions in the RF annul each others, and so the cells provide zero response; ii) the orientation of the bar or edge, and its direction of movement. As shown in Fig. 5, the neuron with maximal activity within the hypercolumn signals orientation of the bar; iii) the velocity of the bar or edge. The system responds only if the bar or edge moves with a specific velocity, within a tuned range (in this work within the range 2-10 deg/s). Lower or higher velocities are cut off. iv) The direction of movement. The system is able to discriminate whether the bar, with a given orientation, is moving in one direction or in the opposite one. For instance, the cell in Fig. 5 detects a direction 270 deg but cut the direction 90 deg. Of course, in order to have a cell that reveals the opposite direction, it is sufficient to invert the spatial disposition of the time lags in Fig. 1.

Further simulations, not shown here, reveal that the network is able to detect both light bars on a dark background, or dark bars on a white background. Moreover, the response is largely independent of the contrast of the input image.

Another important point, to be stressed, is that the velocity band of the cell (which, in the present work, is centered at about 5-8 deg/s) may be shifted to higher or lower velocities by changing the width of the receptive field. In fact, as observed in physiological experiments [2,3], a larger receptive field is generally associated with a velocity high-pass cell, i.e. a cell which responds better to high velocities, high temporal frequencies and low spatial frequencies of the input stimulus. This means that, by replicating the hypercolumn using different

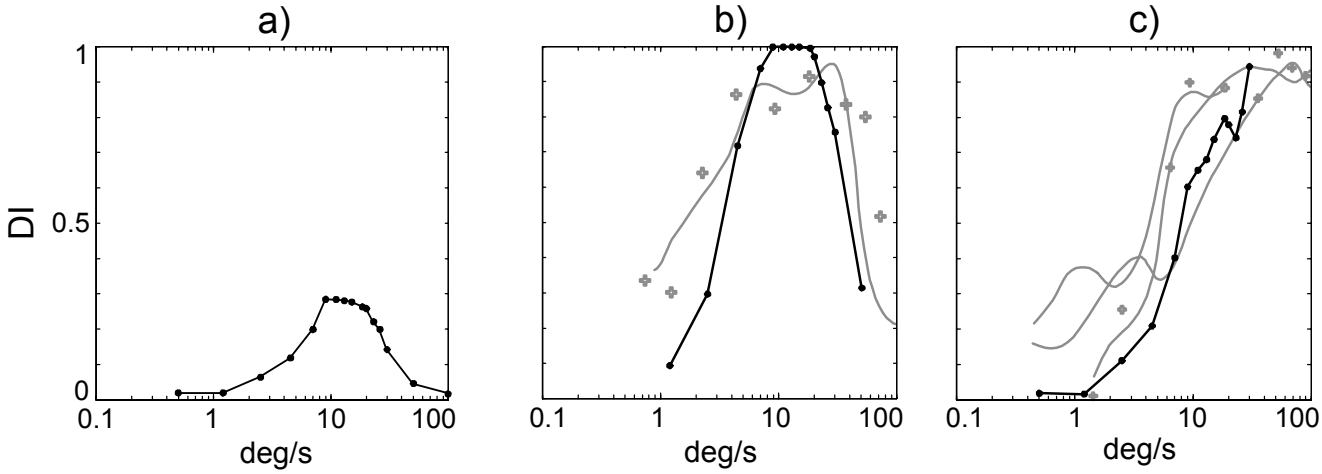

Figure 6. Velocity-DI curves computed for the two models: a) Thalamic input; b) Anti-phase model; c) In-phase model. The gray lines represent experimental data [3], black lines represent model results

dimensions of the RFs, it may be possible to realize a multi-scale analysis of the input image.

At present, the proposed network is able to extract velocity, orientation and direction of movement at a given location, and at an assigned level of temporal and spatial frequency resolution. A subsequent, important problem is how to link the information from different hypercolumns (i.e., from different pixels in the image) to extract features at a higher-level. In our opinion, this may be achieved using lateral excitatory and inhibitory connections among hypercolumns, which realize contextual influences (similar to those hypothesized by the Gestalt psychology). For instance, excitatory connections may be implemented among neurons in different hypercolumns, which concur to the formation of a smooth contour (good continuity criterion) or between neurons which respond to the same direction and similar velocity of movement (common fate criterion). Hence, the present model may represent the basic block for the construction of more complex visual processing networks, able to extract higher-level features from the image.

Finally, it is interesting to comment briefly on the differences between the two proposed models. Both models behave in a similar way in response to a long bar (where "long" means a dimension greater than the major axis of the RF, in this work: 3 deg). The only significant differences are that the anti-phase model exhibits greater direction selectivity, and a less-tuned velocity response (mainly imputable to a greater response at high-velocities, above 10 deg/s). By contrast, a striking difference between the two models is evident when using short bars or spots (where "short" means a dimension comparable to the width of the central sub-region in the RF; in this work: 0.6 deg). In the anti-phase model the preferred direction for a moving spot is equal to the preferred direction for long bar (both moving in the direction of the minor axis of the RF), but with a loss in orientation selectivity. By contrast, in the in-phase model the preferred direction for a spot is orthogonal to the preferred direction for a long bar (Fig. 5). The model responds better when a long bar is moving in the direction of the minor axis of the RF, but when a spot is moving in the direction of the major axis. However, the response to a spot is much smaller compared to the response to a long bar. Briefly, we can summarize these results saying that the anti-phase model responds to moving long bars and spots in a similar way, whereas the in-phase model is able to distinguish long bars from spots (and partly suppress the last ones). It is interesting to observe that both pieces of behaviour have been observed in the physiological literature [2,3,6]. At present it is difficult to understand the implications, virtues and short-comings of these differences in real vision problems.

In conclusion, we presented neural network models for the detection of the main local properties of an image. The models replicate, although in an idealized manner, some important properties of the early visual processing pathway (LGN and V1). Simulations show that the networks are able to detect orientation, velocity and direction of a moving bar or edge. Integration of several similar networks, to accomplish a multi-scale resolution, and to realize contextual influences, may be the natural evolution of this study

References:

[1] DeAngelis, G.C., Ohzawa, I., Freeman, D., Receptive-field dynamics in the central visual pathways', *Trends.Neurosci.,* Vol.18, 1995, pp. 451-458.

[2] Orban, G.A., Kennedy, H. and Maes, H., Response to movement of Neurons in areas 17 and 18 of the cat: velocity sensitivity, *J. Neurophysiol.,* Vol.45, N.6, 1981, pp. 1043–1058.

[3] Orban, G.A., Kennedy, H. and Maes, H., Response to movement of Neurons in areas 17 and 18 of the cat: Direction selectivity, *J. Neurophysiol,.* Vol.45, N.6, 1981, pp. 1059–1073.

[4] Wörgötter, F., Eysel, U.T., Axial response in visual cortical cells: spatio-temporal mechanisms quantified by Fourier components of cortical tuning curves, *Exp. Brain Res.,* Vol 83, 1991, pp. 656–664.

[5] Wörgötter, F., Eysel, U.T., Axis of preferred motion is a function of bar length in visual cortical receptive fields, *Exp. Brain Res.,* Vol. 76, 1989, pp. 307–314.

[6] Henry G.H., Dreher B., Bishop P.O., Orientation Specificity of cells in cat striate cortex, *J. Neurophysiol.*, Vol. 37, 1974, pp. 1394-1409.

[7] Cai, D., DeAngelis, G.C., D. Freeman, 'Spatiotemporal Receptive Field Organization in the Lateral Geniculate Nucleus of Cats and Kittens, *J. Neurophysiol*., Vol.78, 1997, pp. 1045-61.

[8] Smith, E.L., Hamamoto, J., Yoshida, K., Cheng, H., Chino, Y.M., Transfer characteristics of lateral geniculate nucleus X neurons in the cat: effects of spatial frequency and contrast, *J. Neurophysiol.,* Vol.74, 1995, pp. 2548–2557.

[9] Usrey, W.M., Reid, R.C., Alonso, J.M., Rules of connectivity between geniculate cells and simple cells in cat primary visual cortex, *J.Neurosci.*, Vol.21, 2001, pp. 4002–4015.

[10] Jones J.P., Palmer L.A., An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex, *J.Neurophysiol.*, Vol.58, 1987a, pp.1233–1258.

[11] Pei, X., Volgushev, M., Vidyasagar, T.R., Multiple mechanisms underlying the orientation selectivity of visual cortical neurones, *Trends.Neurosci.*, Vol.19, 1996, pp. 272–277.