CEAI, Vol. 4, No. 1, pp. 76-86, 2002 Printed in Romania

MODELLING SOME OF THE EARLY VISUAL PROCESSES INVOLVED IN SPACE PERCEPTION

Diana LUNGEANU

Medical Informatics Department, University of Medicine and Pharmacy Timisoara,

P-ta E. Murgu 2, 1900-Timisoara, e-mail: dlungeanu@medinfo.umft.ro

***Abstract**: This paper presents a computational approach in studying vision and models proposed for two of the early vision mechanisms: one for obtaining the disparity map in stereo vision, and the other for edges enhancement in contour integration. A computational perspective in modelling early visual processes is important for understanding the biological mechanisms, as well as for integrating the known cortical architecture of primary visual area (V1) and the physiological functions (some already proved in the experiments, some only presumed). Therefore, our approach is based on the neuro-physiological findings about the functional columns in V1. We also developed computer applications for implementing and testing the proposed models and the results of the simulation were consistent with the biological data.*

***Keywords:** computational vision, modelling, early vision, stereopsis, contour integration.*

1. INTRODUCTION

Every waking moment we are bombarded with sights from the outside world. The brain is vital for processing the information that floods in through the senses. However, it is still unclear why the world looks as it does and, especially, how the perception is achieved, what are the precise mechanisms involved in visual information processing. Therefore, theoretical models and computer simulations trying to imitate the behavior of biological systems have an important role in understanding some of these mechanisms and phenomena [1-3].

We chose to approach vision modelling from a computational standpoint, focusing on the information processing mechanisms needed to extract the visual information embodied in the energy states or changes flooding in through the eyes. This implies to understand what it is in the external world, where the objects are located, their changes in time, etc. [1-5]. Combining computational perspective with experimental studies in psychology, psychophysics, or neurosciences, a functional understanding of vision can be achieved.

The scientists know now that the brain states represent states of some other systems (the outside world or the body itself) and that the transitions between the states are equivalent to computational operations on these representations [1]. However, there are many unknowns about these representations and processing chains; therefore, the theoretical models are expected to play an important role in understanding them.

In spite of these, until quite recently, most of the computer scientists' approaches were top-down (as those in AI), while the neurobiological scientists preferred the bottom-up research strategies (reflecting the biological constraints and the connectionist approaches). As a result, most of the important results were obtained mainly during the last decades, by inter-disciplinary research, involving scientists from different fields. A turning-point in vision research was the defining of the three levels approaching in nervous systems, by David Marr [4]: *the computational level* of the abstract problem analysis, *the level of the algorithm*, which specifies the procedure to be followed in order to perform the task, and *the physical implementation level*. Marr claimed the independence of the levels, but the models of stereo vision he developed did not ignore the implementation aspects. Further research in vision proved that the levels are highly inter-dependent and, consequently, the implementation and computation cannot be addressed separately [1, 4, 5].

For that reason, we based our models on the neuro-physiological findings about the cortical organization in functional columns and also on some other existing models [6-20]. Moreover, we emphasized the importance of the parallelism in processing and also the important role of the interactions between these parallel-processing chains.

2. SPACE PERCEPTION IN EARLY VISION

The space perception implies a hierarchy of parallel processes [21]. It also implies combining information from different channels in order to obtain coherent and meaningful figures – Figure 1. The computational studies suggested three primary types of representation: *early representations* (location, contrast, sharpness of edges, direction and speed of motion, etc.), corresponding to physical features; *intermediate representations* (about 3-D shape and orientation of small surface regions), which are viewer-centered; and *higher-level representations*, which are object-centered [2, 4].

The recovering of the 3-D structure of the world from 2-D images needs some constraints to be imposed, concerning the properties of the world we live in [4, 6]. In addition to these, among the most powerful cues for 3-D space perception are: binocular disparities, motion, illumination differences, texture, continuous contours, non-visual information.

We have tried to model the stereo-vision and the contour integration mechanisms as parallel processes in early vision. We chose to begin with these aspects in modelling the space perception, as they are strong cues in 3-D perception and play an important role in object discriminating and locating.

The architecture of the striate cortex is function-dependent, having columns and hypercolumns of cells performing the same function for the same location in the visual field (for different aspects, as resolution, angles, orientations, disparities, angular velocities, etc.). These columns are subdivisions at the sub-millimetre scale, going deeply into the cortex structure (through all the six layers) and they can be seen as an outcome of the self-organising tendency during development [1]. There are also cortical modules, seen at an intermediate scale, between the maps and the columns, with similar internal wiring. We respected this functional organisation in the models we developed – Figure 2.

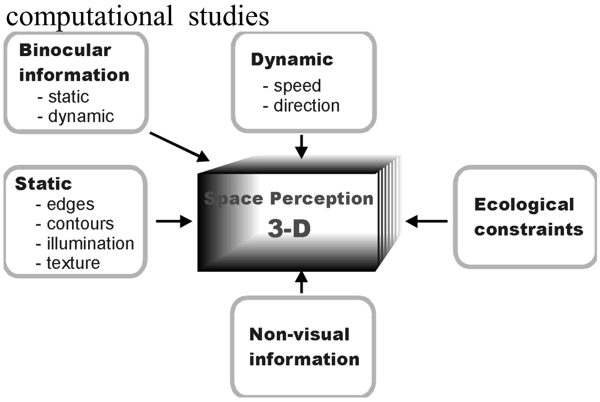

Fig.1. Space perception implies combining information from different sources.

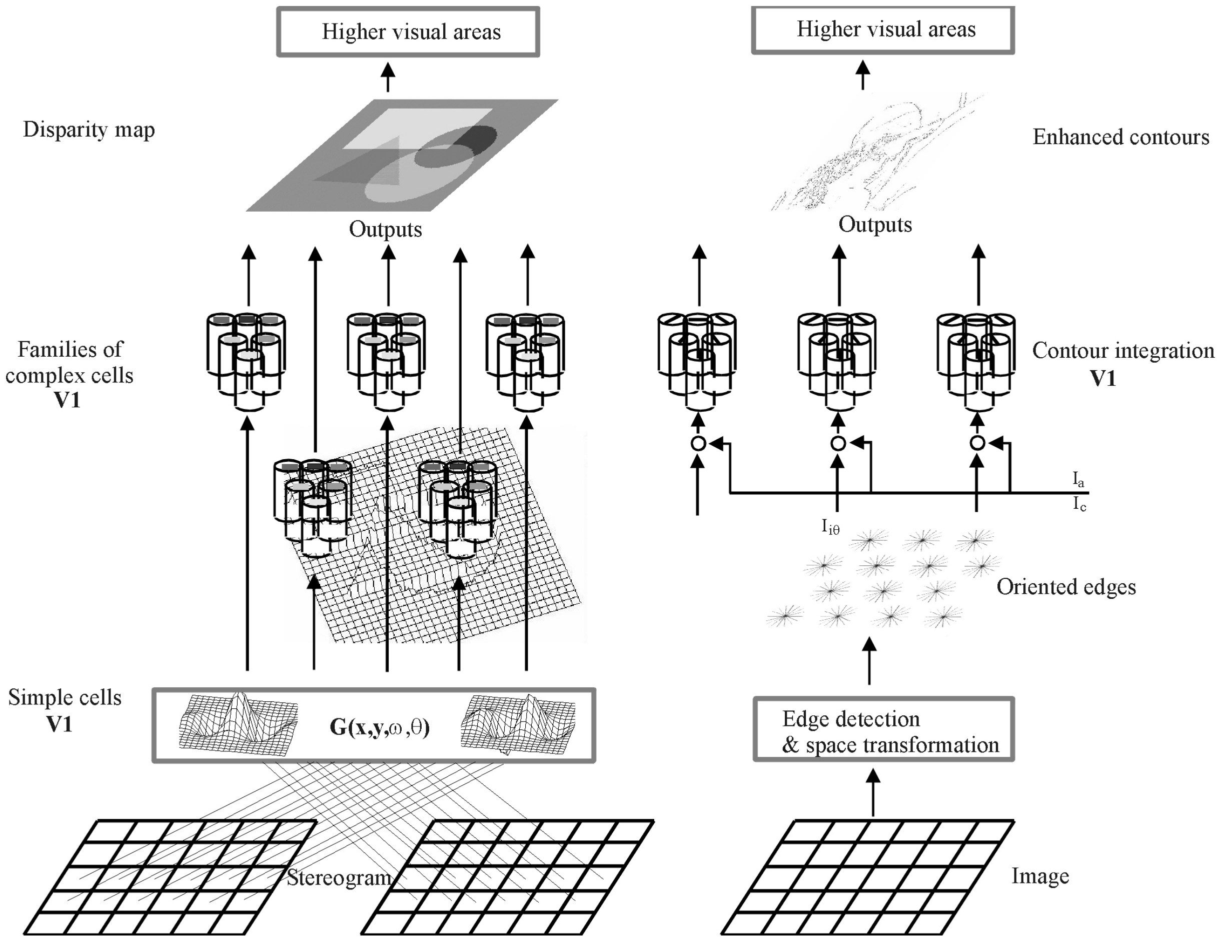

Fig.2. Stylized image of the hierarchical series of processing stages in stereo-vision and contour integration, seen as an architecture- function inter-relation.

3. DISPARITY MAP IN BINOCULAR VISION

The Proposed Model

The stereo-vision model uses families of complex cells sensitive to different disparities for every spatial location in the visual field. Each complex cell computes its response as a combination of inputs from pairs of simple cells in quadrature phase. We used Gabor wavelets to model the receptive fields of such pairs of simple cells (describing the "left" and the "right" receptive fields) – Equation 1.

$$G(x,y) = \exp\left(-\pi\left[(x-x_0)^2/\alpha^2 + (y-y_0)^2/\beta^2\right]\right) \exp\left(-2\pi i\left[u_0(x-x_0)+v_0(y-y_0)\right]\right) \quad (1)$$

Where: (x_0,y_0) specify the spatial location, (α,β) specify the filter's effective width and length, and (u_0,v_0) specify the filter's modulation wave-vector (spatial frequency $\omega_0 = \sqrt{u_0^2 + v_0^2}$; orientation $\theta_0 = \arctan(v_0/u_0)$).

Within each family of complex cells, the cell tuned on the actual disparity will have the highest response and, therefore, they constitute a distributed representation of the binocular disparity. As the Figure 2 (left) shows, these complex cells families are organized in functional columns.

During all the simulations, we assumed that the epipolar constraint was satisfied and we covered a disparity range between –4 and +4.

Results and Discussion

We developed a computer application (STEREO-VISION 2) as a tool for implementing, developing, and testing models of stereopsis. We have implemented two models for obtaining disparity map: the proposed "biological model" and the classical model of Marr and Poggio [4, 7] in order to have a comparison term for our simulations.

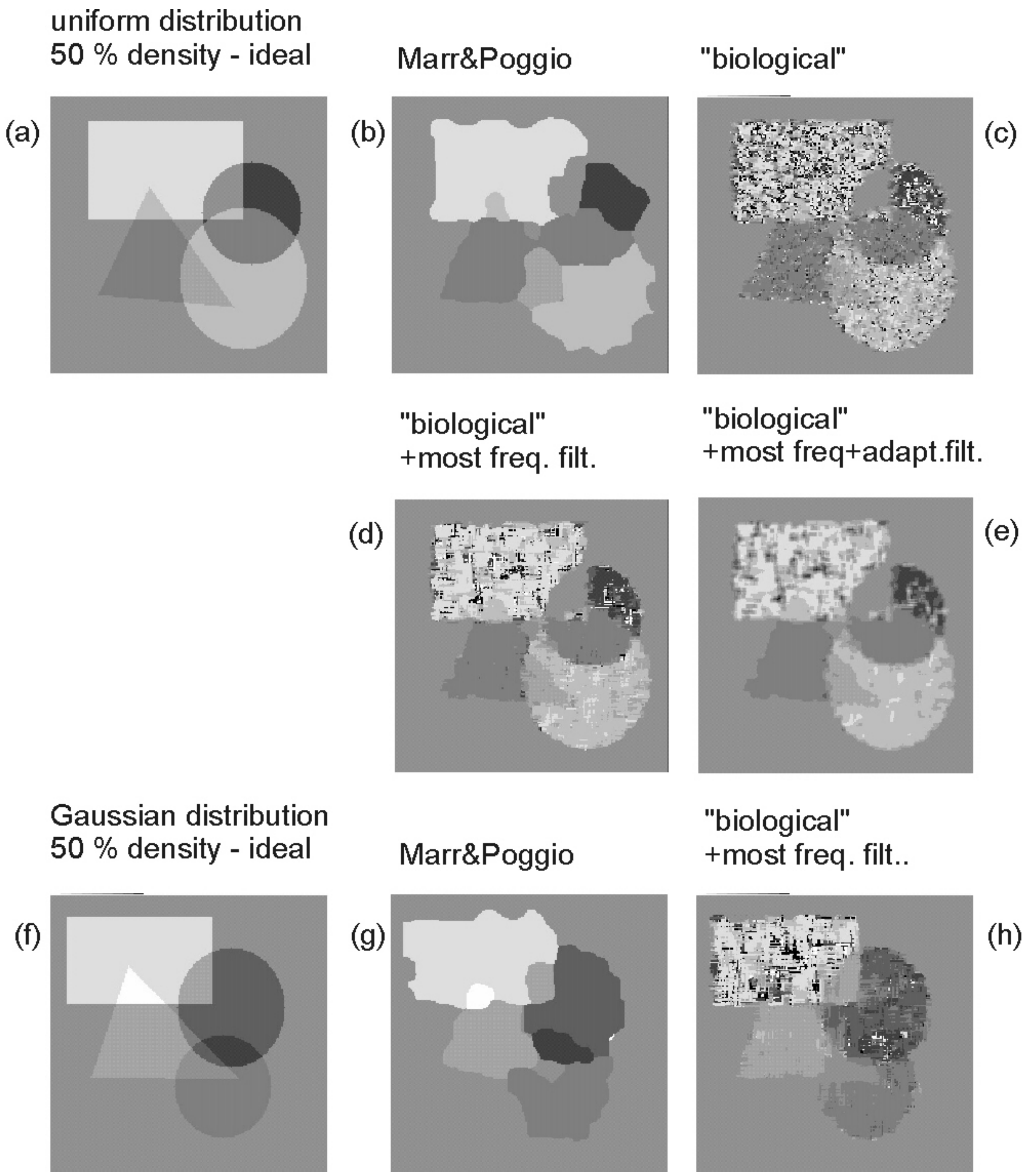

Fig.3. Disparity maps obtained on RDSs of black and white 200x200 pixels uniformly and normally distributed. (a) The "ideal" disparity map for a RDS with uniform distribution. (b) Result obtained with the Marr and Poggio algorithm. (c) Result obtained with the "biological" algorithm (u=1/8, v=0) and no subsequent filtering applied. (d) and (e) The same as (c), but with filtering applied. (f) The "ideal" disparity map for a RDS with normal distribution. (g) and (h) Results obtained on the RDS with normal distribution.

Computational models of stereopsis are based, to a large extent, on research with random-dot stereograms (RDS); the use of RDSs as benchmarks for stereo algorithms is a standard practice. Therefore, STEREO-VISION 2 offers tools for creating RDSs with various dimensions, densities, and distributions (the uniform, Gaussian, and inverse-Gaussian distribution); moreover, it offers the possibility of adding noise to one or to both images of a stereogram (noise with different distributions and densities). The application works with *.BMP files, which can be RDSs or other pairs of images, too.

The results of the simulation (the disparity maps) are obtained in an internal format (values of disparity), but they are also presented as wireframe images and as two-dimensional images using grey levels (the nearer the surface, the lighter the grey of the corresponding zone in the disparity map). In addition, the coarse disparity map can be smoothed using a filtering method: Gaussian weighting filtering, most frequent value filtering, median value filtering, and an adaptive smoothing (as a general tool for early vision [22]).

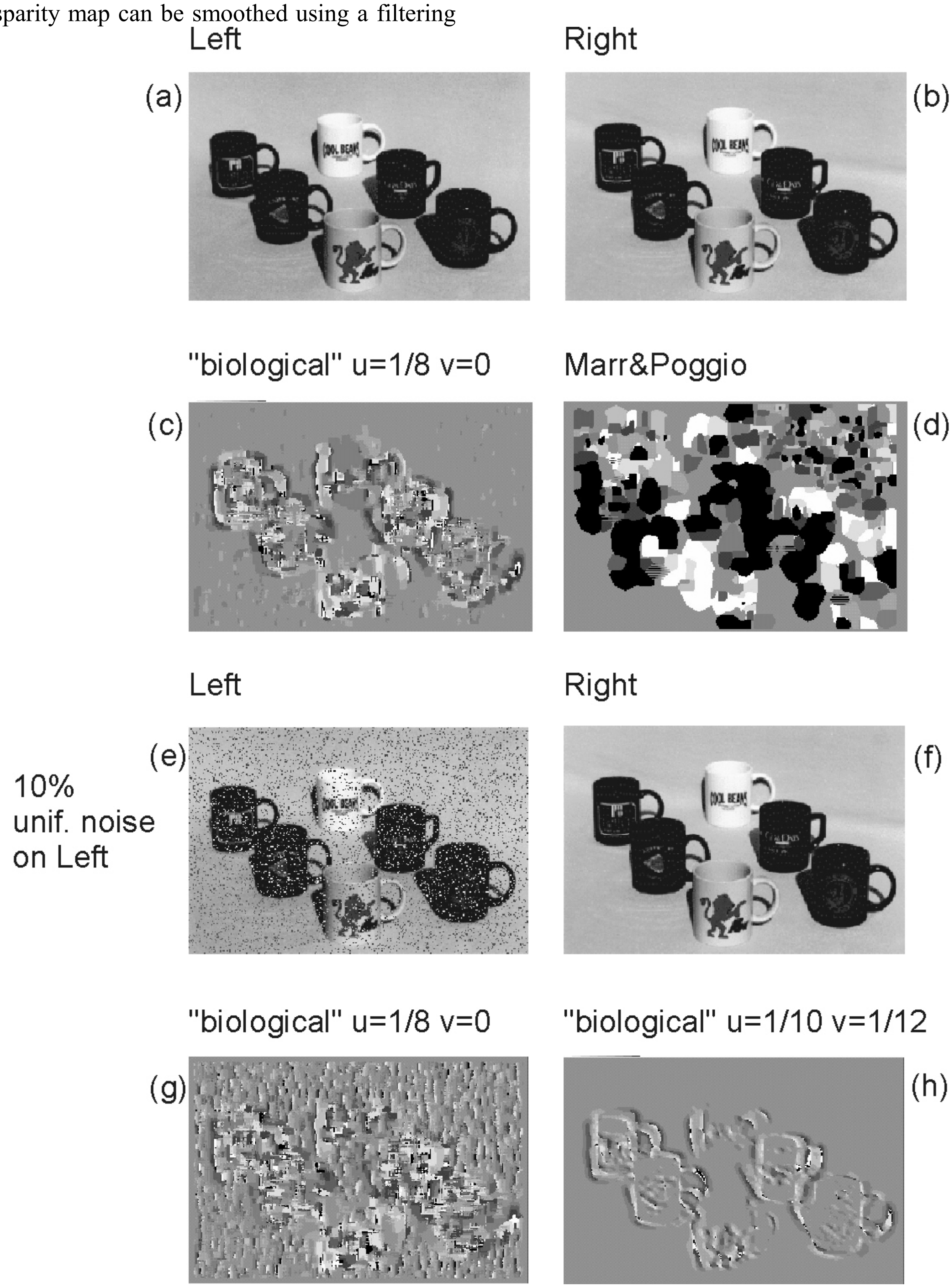

Fig.4. Results obtained on "real-world" coloured stereograms. (a) and (b) The left and right images. (c) Disparity map obtained with the biological algorithm (one-dimensional Gabor wavelets). (d) Result obtained with the Marr and Poggio algorithm. (e) and (f) The noisy stereogram. (g) and (h) Results obtained on the noisy stereogram for one- and two-dimensional Gabor wavelets.

We tested the two models on RDSs with uniform and Gaussian distributions; we also used accurate stereograms and noisy ones, as well as one-dimensional and two-dimensional Gabor functions; moreover, we tested the effects of the different filtering methods. We presented the detailed results in previous papers [22]. Figure 3 presents some of the results obtained on RDSs without any additional noise added. We can see that the "biological" algorithm is better in preserving the shapes and edges, while the algorithm of Marr and Poggio tends to round the corners (the cohesiveness constraint is too strong for this model). Unfortunately, the disparity map generated by the "biological" model was noisy; even though, successive filtering applied can lead to a significant improvement.

The tests on noisy RDSs and on real-world pairs of images led to the conclusion that the performance of the "biological" model was significantly better than that of the Marr and Poggio model. Moreover, while for the RDSs two-dimensional Gabor wavelets (used to model the behaviour of the complex cells) worsened the results, for the real-images they were the solution to obtain reliable disparity maps – Figure 4.

The explanation must rely on the fact that the RDSs are artificial images created by the horizontal displacement of sets of points in the images and therefore they could be successfully solved taking into account only the horizontal disparity. On the other hand, when noise was added or real images were tested, additional information had to be used in order to solve the matching problem. We have to emphasize that we used only filtering techniques on the original images (Gabor wavelets) and a distributed representation of the disparity information, both of them plausible from the biological point of view and consistent with the data from neuro-physiological experiences.

4. EDGE ENHANCEMENT IN CONTOUR INTEGRATION

4.1. *The Proposed Model*

The model for contours integration consists of a network based on a hexagonal grid of orientation hypercolumns – Figure 5. We chose such an organization for a better covering of the visual space and a proper description of the primitive features in the scene (preserving both spatial distribution and orientation of the edges). Each hypercolumn consists of a family of oriented edge elements, in fact neural oscillators that are pairs of excitatory-inhibitory neurons (*x*–*y*), covering the whole range of possible orientation.

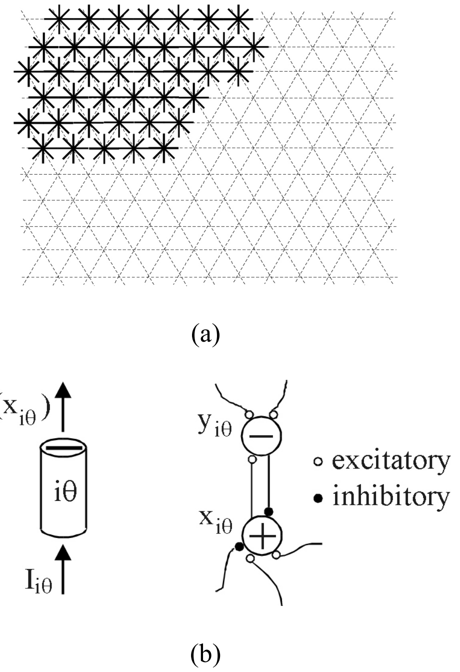

Fig.5. The model uses hypercolumns of orientation. (a) The hexagonal grid. (b) The elementary edges are neural oscillators characterized by their spatial location *i* and their orientation θ.

The behaviour is described by Equations 2:

$$
\begin{aligned}
\dot{x}_{i\theta} &= -\alpha_x x_{i\theta} + I_{i\theta} + G_0 h_x(x_{i\theta}) \\
&\quad - \sum_{\Delta\theta} V_{i\theta}(\Delta\theta) h_y(y_{i,\theta+\Delta\theta}) \\
&\quad + \sum_{j\neq i,\phi} G_{i\theta,j\phi} h_x(x_{j\phi}) + I_a \\
\dot{y}_{i\theta} &= -\alpha_y y_{i\theta} + h_x(x_{i\theta}) + \sum_{j\neq i,\phi} W_{i\theta,j\phi} h_x(x_{j\phi}) + I_c
\end{aligned}
\tag{2}
$$

Where: i is the spatial location and θ is the orientation of the edge element (composed of $x_{i\theta}$ and $y_{i\theta}$); $I_{i\theta}$ is the input for the edge element; $V_{i\theta}(\Delta\theta)$ models the inhibition within the same hypercolumn, similar to short-distance connections; G_0 models the self-excitatory connection; $G_{i\theta,\ j\phi}$ models the excitatory influence of the neighbours with similar orientation, a lateral connectivity aimed to enhance the edges belonging to plausible common contours; $W_{i\theta,\ j\phi}$ models the inhibitory influence of the neighbours for different orientation, similar to long-range connections within the same cortical layer; α_x, α_y are the membrane constants for the two edge-neurons and they are positive values; h_x, h_y are the transfer functions for the two edge-neurons; I_c allows the control from higher visual areas and I_a models the influence of the general background activity in the network.

4.2. *Results and Discussion*

We implemented the model and the resulted application was called VISION. It was imagined as a friendly tool: it accepts *.BMP* files as inputs and initializes the hypercolumns matrix based on the gradient filter, allowing the user to choose the matrix dimension, the visual space dimension, as well as the graphics options. Moreover, the user can set and test the model's parameters. The simulation can be performed continuously, or step-by-step, at the user's choice, who can also specify the intermediate results (s)he wants to visualize. The intermediate results can also be saved on disk, both as *.BMP* files, and as text files, for allowing a numerical checking of the performance.

We tested the model on different kinds of images, from very simple ones (lines, circles), to more complex images, and even on real world scenes (detailed presentation of the application in [23]). As the experimental studies have proved, in order to have biologically plausible models of the visual mechanisms, we cannot ignore the parallel architecture of the brain itself, which is highly function-dependent. On the contrary, we have to base our reasoning process on the experimental data and to assume a similar organisation in the models we build. Therefore, in our work, we used a similar distributed representation of information for both models.

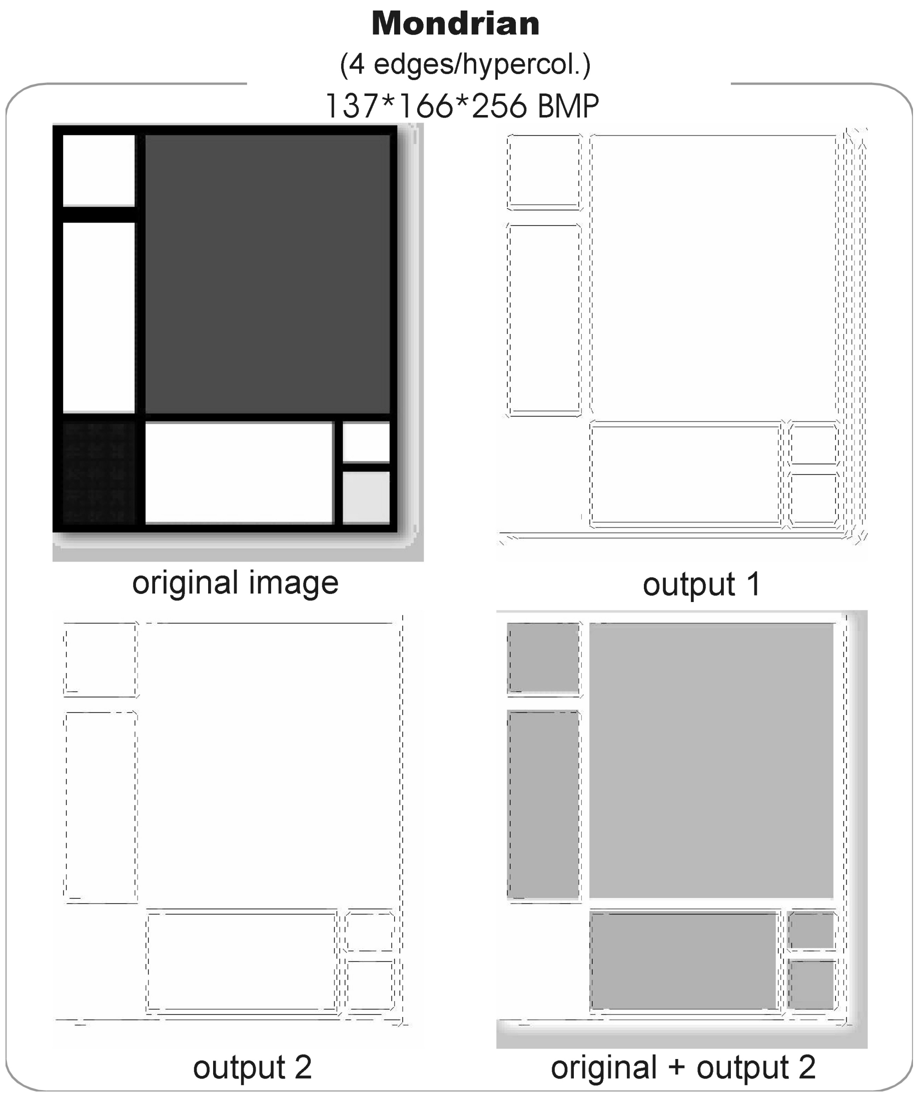

Fig.6. Results obtained on a Mondrian image, where the second set obtained as output was superimposed on the original image (the colour levels were shifted to higher values to allow the superposition).

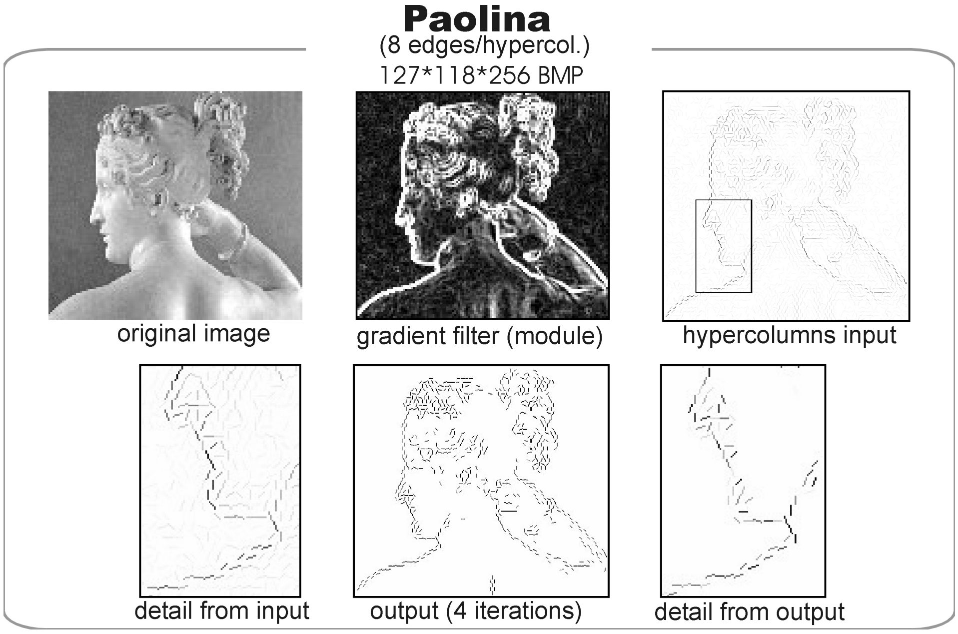

Fig.7. Image where details were enlarged in order to emphasize the elimination of the spurious edges.

Results in Figure 6 were obtained on a noise-free image with 8 bits grey level resolution. We have to mention that the contours are lightly shifted to the right, due to the gradient filter applied for the initial edge detection and to the procedure used for visualising the contours as *.BMP* files. However, the intrinsic information in the image was not altered.

Figure 7 presents results on a different image, where some details from the image were emphasized.

As we expected, the simulations on different kinds of images emphasised that the performances depend not only on the model parameters, but also on the processed image. This is biologically plausible: at the cortical level there is a high degree of redundancy and parallelism in processing [1, 4, 17], which suggests a finer adjustment of the biological parameters. Miller and Zucker [17] emphasised that it is unlikely to have (or to rely on) individual responses for neurons at the cortical level. On the contrary, the biological systems should rely on a certain level of activity in large groups of cortical neurons (the "cliques") in order to be able to process the details in images. Apart from this, we have to mention that some of the apparent discontinuities in the contours are due to the hexagonal grid we used for the visual space representation, and therefore, most often, no filling in is necessary.

The images presented show how highly dependent is the model's behaviour on the parameters used for the simulation. Nevertheless, the model proved to be quite robust in general and also to the noise.

5. CONCLUSIONS

As the experimental studies have proved, in order to have biologically plausible models of the visual mechanisms, we cannot ignore the parallel architecture of the brain itself, which is highly function-dependent. On the contrary, we have to base our reasoning process on the experimental data and to assume a similar organisation in the models we build. Therefore, in our work, we used a similar distributed representation of information for both models.

Implementing the two models [22-24] and testing them, the results obtained in the computer simulations confirmed the biological plausibility and the necessity of combining different types of information for a sound 3-D space perception. As Figure 8 shows, the “quality” of the disparity map is highly dependent on the Gabor filtering parameters: for the same spatial frequency, using one-dimensional Gabor function or two-dimensional Gabor functions with different orientations may lead to different results. Moreover, the two-dimensional filtering seems to compensate for the inherent noise and some slight vertical displacements.

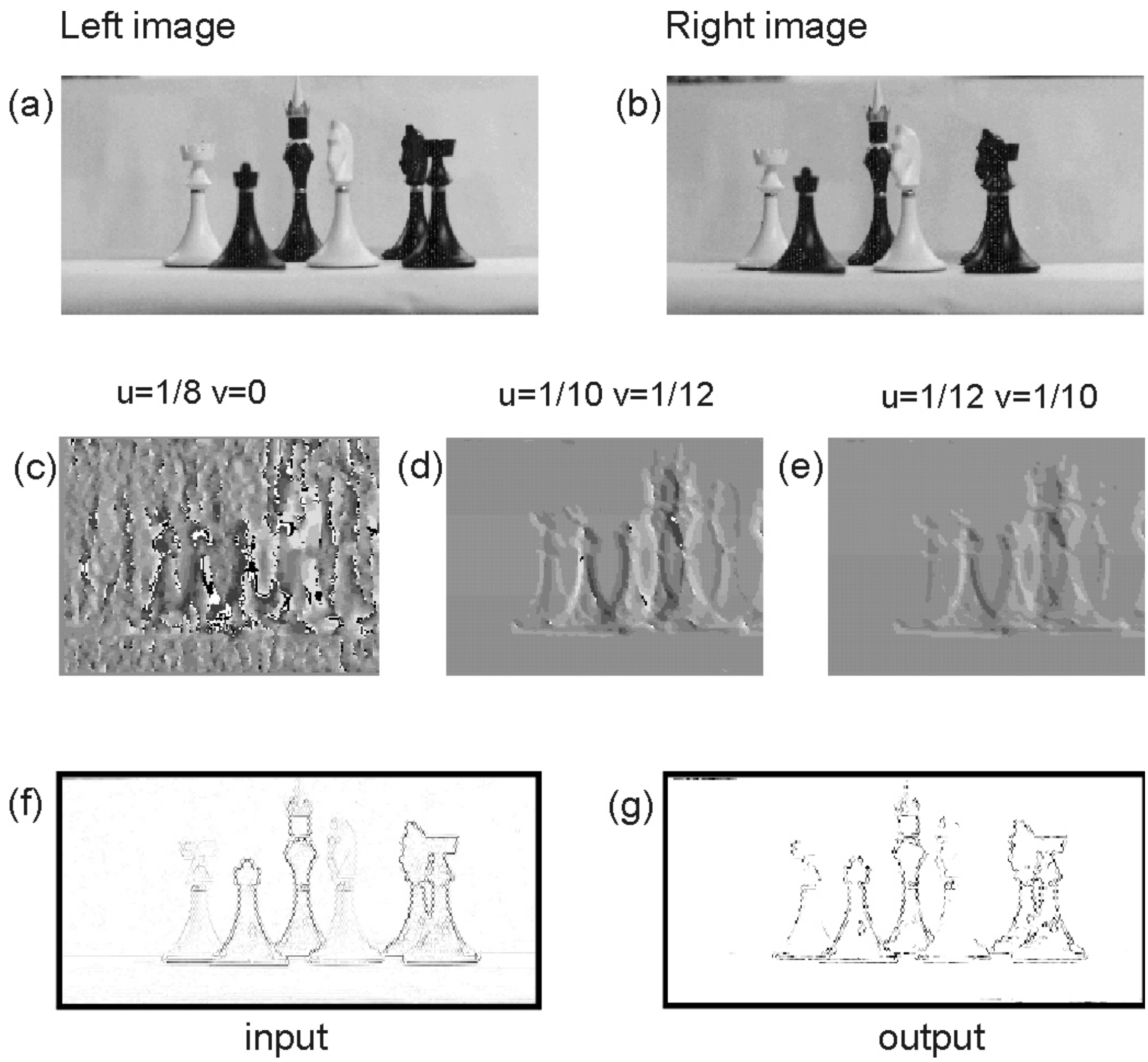

Fig.8. Results obtained on real-world images. (a) and (b) Initial images. (c), (d), and (e) Disparity maps obtained for different parameters, represented as gray-level images. (f) and (g) Input and output edges in the contour integration model.

On the other hand, the recurrent network used for edge enhancement is highly dependent on the parameters, as well. Therefore, when we used the same parameters ($V_{i\theta}$, $G_{i\theta,\,j\phi}$, $W_{i\theta,\,j\phi}$) and neighbourhood profile all over the image, this spoiled the model's performance – Figure 8 (g): false edges were eliminated, but other useful contours were lost, too.

If we had used a non-uniform sampling and sets of overlapping orientation-selective cells groups (finely tuned, as it is the biological situation), the results from different parallel processes could have been combined. At this processing level, also inputs from higher visual areas might intervene. We have not implemented this type of connection, yet – we have tried to see whether only the orientation hypercolumns architecture could lead to a contour enhancement at the striate cortex level. Not surprisingly, the apparent "shortcomings" of the models (noisy disparity maps, noisy contours) confirmed the need of inter-connections between parallel processing chains for eliminating spurious features and reciprocally complementing missing information. The biological and physiological data are consistent with these conclusions, i.e. the necessity of combining information from different sources (both visual and non-visual) to obtain the 3-D perception [1, 20, 21]. Experiments showed that when the visual system is provided with information from only one source, it needs some "accommodating" delay to reach the space sensation.

The simulation results also accounted for the necessity to investigate the synchrony in the activity of large networks, as a must for increasing the robustness of the visual system.

We think our results are encouraging, as long as we used only known properties of simple and complex cells in V1 and the known columnar organisation at the striate cortex level. Further research is focused on effectively combining the results from the two models.

6. REFERENCES

[1] Churchland, P. S. and Sejnowski, T. J. - "The computational brain", The MIT Press, Cambridge, 1992.

[2] Hildreth, E. – "Computational vision". In: Wilson, R. A. and Keil F. C., eds. *The MIT Encyclopedia of the Cognitive Sciences*, The MIT Press, Cambridge, **172-173**, 1999.

[3] Jordan, M.I. and Russell, S. – "Computational intelligence". In: Wilson, R. A. and Keil, F. C., eds. *The MIT Encyclopedia of the Cognitive Science,*. The MIT Press, Cambridge, **lxxiv-xc**, 1999.

[4] Marr, D. – "Vision",.W.H. Freeman, New York, 1982 (13^{th} printing 1996).

[5] Li, Z. and Atick, J.J. – "Toward a Theory of the Striate Cortex", *Neural Computation* 6, **127-146**, 1994.

[6] Frisby, J.P. – "Stereo correspondence and neural networks". In: Arbib, M. A., ed. *Handbook of Brain Theory & NN*. The MIT Press, Cambridge, **937-941**, 1995.

[7] Marr, D. and Poggio, T. – "Cooperative computation of stereo disparity", *Science* 194,. **283-287**, 1976.

[8] Ludwig, K. O., Neumann, H. and Neumann, B. – "Local stereoscopic depth estimation", *Image and Vision Computing* 12, **16-35**, 1994.

[9] Qian, N. – "Computing of stereo disparity and motion with known binocular cell properties", *Neural Computation* 6, **390-404**, 1994.

[10] Nasrabadi, N. M. and Choo, C. Y. – "Hopfield network for stereo vision correspondence", *IEEE Trans. on Neural Networks* 3, **5-13**, 1992.

[11] Hongo, S., Sonehara, N. and Yoroizawa, I. – "Edge-based binocular stereopsis algorithm – a matching mechanism with probabilistic feedback", *Neural Networks* 9, **379-395**, 1996.

[12] Siper, B. J. and Klarquist, W. N. – "Patch-based stereo in a general binocular viewing geometry", *IEEE Trans. on Pattern analysis and Machine Intelligence* 19, **247-253**, 1997.

[13] Qian, N. and Mikaelian, S. – "Relationship between phase and energy methods for disparity computation", *Neural Computation* 12, **279-292**, 2000.

[14] Hirakura, Y., Yamaguchi, Y., Shimizu, H. and Nagai, S. – "Dynamic linking among neural oscillators leads to flexible pattern recognition with figure-ground separation", *Neural Networks* 9, **189-209**, 1996.

[15] Li, Z. – "A neural model of contour integration in the primary visual cortex", *Neural computation* 10, **903-940**, 1998.

[16] Riesenhuber, M., Bauer, H. U., Brockmann, D. and Geisel, T. – "Breaking rotational symmetry in a self-organising map model for orientation map development", *Neural Computation* 10, **717-730**, 1998.

[17] Miller, D.A. and Zucker, S.W. – "Computing with self-excitatory cliques: a model and an application to hyperacuity-scale computation in visual cortex", *Neural Computation* 11, **21-66**, 1999.

[18] Law, T., Itoh, H. and Seki, H. – "Image filtering, edge detection, and edge tracing using fuzzy reasoning", *IEEE Trans. Pattern Analysis and Machine Intelligence* 18, **481-491**, 1996.

[19] Okajima, K. – "A model visual cortex incorporating intrinsic horizontal neural connections", *Neural Networks* 9, **211-222**, 1996.

[20] Verschure, P.F.M.J. and Konig, P. – "On the role of the biophysical properties of cortical neurons in binding and segmentation of visual scenes", *Neural Computation* 11, **1113-1138**, 1999.

[21] Tyler, C.W. – "Cyclopean riches: cooperativity, neuroentropy, hysteresis, stereoattention, hyperglobality, and hypercyclopean processes in random-dot stereopsis". In: Papathomas, T. V., Chubb, C., Gorea, A. and Kowler, E., eds. *Early vision and beyond*, The MIT Press, Cambridge, **5-15**, 1995.

[22] Lungeanu, D., Popa, C., Hotca, S. and Macovievici, G. – "Modelling biological

depth perception in binocular vision: the local disparity estimation", *Medical Informatics (Taylor&Francis)* 23, **131-143**, 1998.

[23] Lungeanu, D., Cotirlea, V. and Hanga, L. – "From primary edges to contours: modelling contour integration in human early vision". In: Hasman, A., Blobel, B., Dudeck, J., Engelbrecht, R., Gell, G., and Prokosch, H. U., eds. *Medical Infobahn for Europe. Proceedings of MIE2000 and GMD2000,* IOS Press, Amsterdam, **68-76**, 2000.

[24] Lungeanu, D. – "Modelling space perception in human early vision. A computational approcah". In: Patel, V., Rogers, R. and Haux, R. eds. *MEDINFO2001. Proceedings of the 10th World Congress on Medical Informatics*. IOS Press, Amsterdam, **919-923**, 2001.

7. ACKNOWLEDGEMENTS

I have to thank Professor Stefan Holban (Polytechnic University Timisoara) and Professor George I. Mihalas (University of Medicine and Pharmacy Timisoara) for their permanent support during my Ph.D. work, part of it being included in this article. I am also deeply indebted to my colleagues from the Polytechnic University Timisoara, Carmen Popa-Lile, Viviana Cotirlea, and Sergiu Hotca, for our fruitful collaboration.