Advanced biologically plausible algorithms for low-level image processing

Valentina I.Gusakova[1], Lubov N. Podladchikova[1], Dmitry G. Shaposhnikov[1], Sergey N.Markin[1], Alexander V. Golovan[1,2], Seong Whan Lee[2]

[1]A.B.Kogan Research Institute for Neurocybernetics, Rostov State University, Russia
[2]Center for Artificial Vision Research, Korea University, Korea

ABSTRACT

At present, in computer vision, the approach based on modeling the biological vision mechanisms is extensively developed. However, up to now, real world image processing has no effective solution in frameworks of both biologically inspired and conventional approaches. Evidently, new algorithms and system architectures based on advanced biological motivation should be developed for solution of computational problems related to this visual task. Basic problems that should be solved for creation of effective artificial visual system to process real world images are a search for new algorithms of low-level image processing that, in a great extent, determine system performance. In the present paper, the results of psychophysical experiments and several advanced biologically motivated algorithms for low-level processing are presented. These algorithms are based on local space-variant filter, context encoding visual information presented in the center of input window, and automatic detection of perceptually important image fragments. The core of latter algorithm (the cascade method) are using local feature conjunctions such as noncolinear oriented segment and composite feature map formation. Developed algorithms were integrated into foveal active vision model, the MARR [18, 22, 23]. It is supposed that proposed algorithms may significantly improve model performance while real world image processing during memorizing, search, and recognition.

Keywords: Active vision, low-level processing, facial images, perceptually important regions

1. INTRODUCTION

At present, in computer vision, the approach based on modeling the biological vision mechanisms is extensively developed [1-4. 6, 10, 19]. Earlier [18, 22, 23], the biologically plausible active vision model for Multiresolutional Attentional Representation and Recognition (MARR) was developed. The MARR model is Foveal Vision System imitating space-variant resolution from the fovea to the retinal periphery and foveation at the most informative visual objects. Such systems have evident computational advantages as compared with artificial visual systems based on uniform image representation [25]. The MARR model was examined while processing test facial images and demonstrated high rate of recognition invariantly to scale, rotation, and shift. However, model algorithms should be modified for processing natural facial video images.

Face recognition in natural scene is one of important areas in computer vision. It demands a fast and reliable solution of many particular visual tasks, such as face region identification, tracking, recognition, etc. Some of these tasks, in particular, invariant recognition of facial images changing over time have no effective solution in frameworks of both biologically inspired and conventional approaches up to now [9, 12, 14, 15, 17, 21, 24, 27, 30]. Evidently, computational problems of face recognition in natural scene can be solved by means of advanced biologically plausible algorithms and system architectures that may extend advantages of foveal vision.

It is widely accepted that the basic problem that should be solved for creation of an effective artificial visual system to process real world facial images is a search for high-performance algorithms for following visual tasks: (1) automatic selection of the most informative image (scene) regions for detailed processing because only these regions are important for robust face recognition [11, 12, 14, 17, 29, 30]; (2) low-level image processing that, in a great extent, determines system performance as a

whole [19, 28]. Several biologically plausible algorithms for solution of (1) and (2) problems based on some known neurobiological data [13, 16, 20], and results of psychophysical experiments are considered in the present paper.

2. THE RESULTS OF PSYHOPHYSICAL EXPERIMENTS

The goal of present experiments is verification of the model suggestion [18] about local nonuniformity of the human visual perception in the peripheral vision field that may influence next fixation point choosing before estimation of object perceptual importance. Special experimental methodology and setup were developed [26]. According to preliminary instruction, a subject should fix his gaze on a foveal stimulus during whole experiment. Sensitivity to simple stimuli (0.3° spots) and orientation selectivity (0.1°x1° segments of different orientations changing by 45° step) were tested in consequent peripheral vision field points separated by 2°. Each test point was characterized by sensory tuning sharpness, subject response time, and number of eye movements. A particular experimental trial was repeated no less than 6 times for each subject and the obtained data were averaged over trials. Sensory tuning sharpness for each vision field point was estimated according to probability (P) of correct subject response. Ten subjects were tested in 30-60 points of the peripheral vision field.

Three types of points were revealed in the peripheral vision field, namely: high-selective, middle-selective (P=0.9-0.8), and low-selective (P≤0.7) points. In primary map of each subject, high- and low-selective points are spatially arranged and alliterated while testing both orientation selectivity and sensitivity to simple stimulus. To examine possible local nonuniformity of sensory tuning, several averaged estimations for different 2° microareas in primary maps were received. (Fig.1, 2). In addition to macrodynamics of sensory tuning and response time**,** their local variations are remarkable. (Fig.1). The number of low-selective points is significantly different (p>0.95) inside high- and low-selective microareas presented in Fig.2.

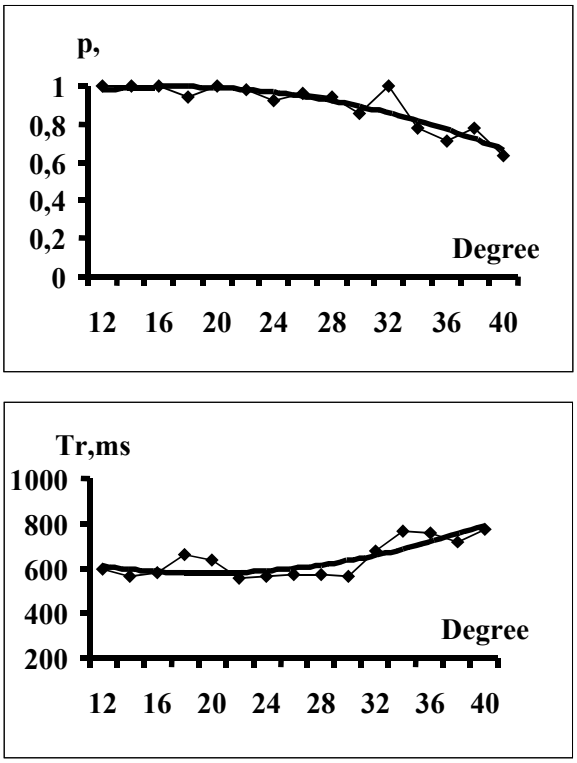


Fig. 1.Example of distributions of averaged sensory tuning: (P) and reaction time, (Tr) in consequent 2° columns of primary map**.** Subject G.

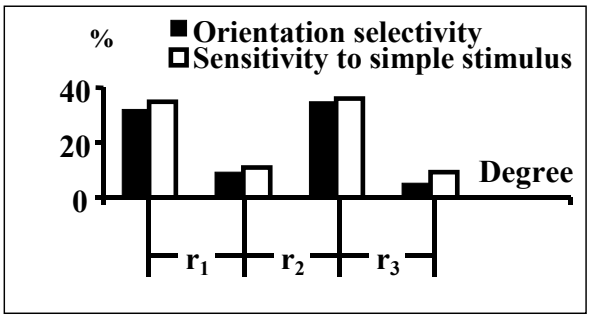


Fig. 2. Averaged (ten subjects) estimation of alteration of low- and high-selective microareas in similar fragments (r_1=2°÷8°, r_2=r_3=2°) of subject primary maps.

The obtained results agree with model suggestion about the local spatial nonuniformity of sensory tuning in the human peripheral vision field. Such local spatial nonuniformity of sensory tuning may determine the priority in fixation points choosing during visual perception before estimation of object perceptual importance. Evidently, low-selective microareas of the vision field may be ignored while next fixation point choosing. To the contrary, high-selective microareas may attract the gaze.

3. CASCADE METHOD FOR DETECTION OF THE MOST INFORMATIVE FACIAL REGIONS

Basing on the results of psychophysical experiments about most informative local features [8,18] and some neurobiological data [13,16,20], cascade method for selective detection of perceptually important facial regions has been developed (Fig.3). .

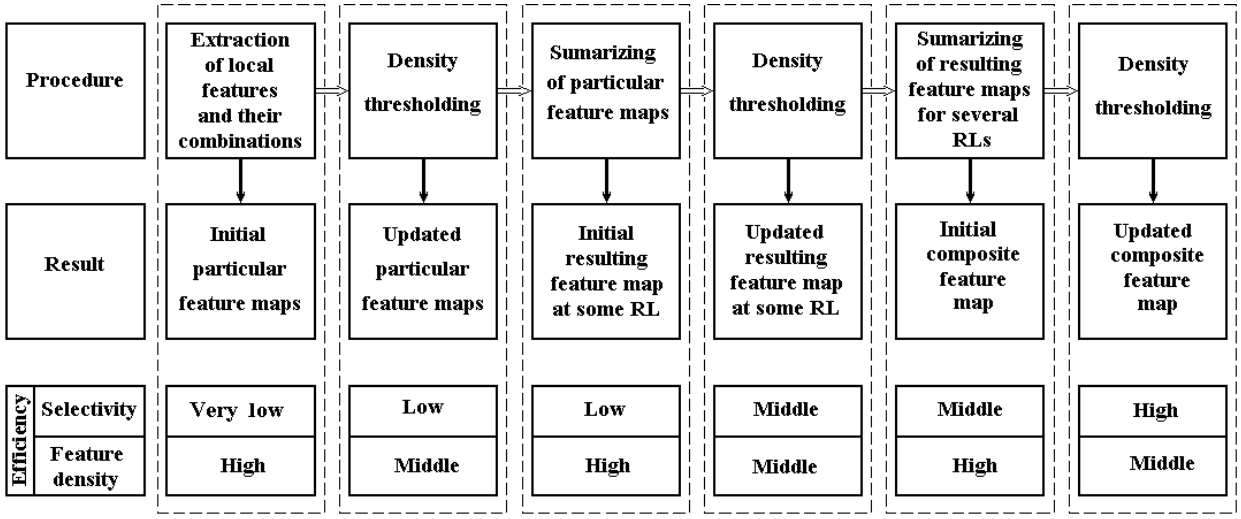


Fig.3. Cascade method schematics: procedures, their results, and efficiency (selectivity and feature density) for detection of perceptually important facial regions.

The method includes three procedures: (1) formation of 5 particular feature maps based on extracted local features and their combinations (such as noncollinear or orthogonal segments) in every 5x5 image fragment for each resolution level; (2) density thresholding of feature maps; (3) composite map formation.

As the first stage in elaboration of algorithms for pre-attentive feature map formation, we performed the quantitative comparison of primary features inside perceptually important face image regions (eyes, mouth, and nose) and other facial

compartments [7]. On the base of the obtained results about concentration of oriented segments and their specific combinations inside the most informative facial regions, 5 parameters are proposed for feature extraction in each 5x5 image fragment and for particular feature map formation.

For each (*i,j*) image point, we examined the following combinations of primary features extracted from a local image fragment of 5x5 size centered at this point (*i,j*) which may imitate some biological mechanisms [13, 16, 20] for encoding local feature conjunctions:

(i) Number of oriented edges extracted from the fragment:

$$O_{ij}^{(l)} = \sum_{p,q} \mathrm{sgn}\left(\overline{Z}_{i-p,j-q}^{(l)}\right),$$

where $\overline{Z}$ is solution of equation (2) for the corresponding image point

Here, the function $\mathrm{sgn}(z)$ is defined as follows:

$$\mathrm{sgn}(z) = \begin{cases} 1, & if\ z > 0, \\ 0, & otherwise. \end{cases}$$

(ii) Estimation of difference between edge orientations in central and peripheral points of the current local image fragment that is determined as follows:

$$D_{ij}^{(l)} = \sum_{p,q} \eta\left(\left| \varphi_{ij}^{(l)} - \varphi_{i-p,j-q}^{(l)} \right| \right)$$

$\varphi_{ij}^{(l)}$ orientation of the edge detected in the fragment center;

$\varphi_{i-p,j-q}^{(l)}$ orientations detected in fragment peripheral points.

Function $\eta(\varphi)$ is determined as follows:

$$\eta(\varphi) = \begin{cases} \varphi, & if\ \varphi \le 4 \\ 8 - \varphi, & otherwise \end{cases}$$

Thus, for collinear orientations, η = 0; for orthogonal ones, η = 4; η =3, 2, 1. for other relations.

(iii) Number of edges perpendicular to the central edge, $P_{ij}^{(l)}$:

(iv) Difference in numbers of horizontal and vertical edges in a local fragment, $V_{ij}^{(l)}$.

(v) Number of horizontal edges detected in a local fragment, $H_{ij}^{(l)}$:

Each of the parameters $O_{ij}^{(l)}, D_{ij}^{(l)}, P_{ij}^{(l)}, V_{ij}^{(l)}, H_{ij}^{(l)}$ allows constructing a pre-attentive map characterizing distribution of particular feature over the image. The particular feature map includes image points in vicinity of which a number of the corresponding particular features exceed a certain threshold. These points may be considered as points of interest for the original image for detailed processing. Thus, the parameters described above allowed to construct the corresponding feature maps **O, D, P, V, H**. It was shown, that each of selected particular maps may preferably identify the arrangement of image informative regions but some points of interest are identified outside these regions.

Density thresholding procedure was applied to each initial particular feature map to increase the selectivity for detection of most informative facial regions. Coefficient of selectivity (CS) was estimated as ratio of number of points of interest extracted from eye, nose, and mouth regions (N_{emn}) to number of such points over whole image (N_{image}). The parameters for density thresholding procedure for each feature map were determined in accordance with averaged results of computer experiments for 10 test images. The feature density providing maximal CS is considered as a basic value during calculation of corresponding threshold. In particular, it is equal to about 0.6 for all oriented segments extracted from 5x5image fragment at resolution level 2. A number of extracted points of interest sharply decreased after such procedure both over whole image and most informative regions. As a rule, CS estimated over modified feature map increased in 1.2 -1.5 times depending on image and particular

feature map. For example, the mean CS for ten test images for the initial feature map based on estimation of all oriented elements at the resolution level 1 was equal to 0.24; it increased up to 0.31 after density thresholding. For each resolution level, we obtained updated feature maps $\mathbf{O}^{*(l)}$, $\mathbf{D}^{*(l)}$, $\mathbf{P}^{*(l)}$, $\mathbf{V}^{*(l)}$, $\mathbf{H}^{*(l)}$ after density thresholding procedure.

To increase points of interest extracted in most informative regions, we constructed a composite feature map from a set of particular feature maps at multiple resolution levels. During construction of the final composite map, we used density thresholding and summarizing procedures applied to intermediate maps. For each resolution level, modified maps $\mathbf{O}^{*(l)}$, $\mathbf{D}^{*(l)}$, $\mathbf{P}^{*(l)}$, $\mathbf{V}^{*(l)}$, $\mathbf{H}^{*(l)}$ described above were summarized to obtain the resulting pre-attentive map:

$$\mathbf{R}^{(l)} = \mathbf{O}^{*(l)} + \mathbf{D}^{*(l)} + \mathbf{P}^{*(l)} + \mathbf{V}^{*(l)} + \mathbf{H}^{*(l)}.$$

In the next stage, the resulting map $\mathbf{R}^{(l)}$ of points of interest for each resolution level was modified by density thresholding procedure

It was shown during computer experiments, that the composite maps based on summarizing particular feature maps for different resolution levels provide high-selective automatic detection of the eyes, nose, and mouth regions while processing home-made (n=10) and ORL (n=40) face image database[1], i..e., more than 90% of extracted points of interest are located inside these regions (Fig.4).

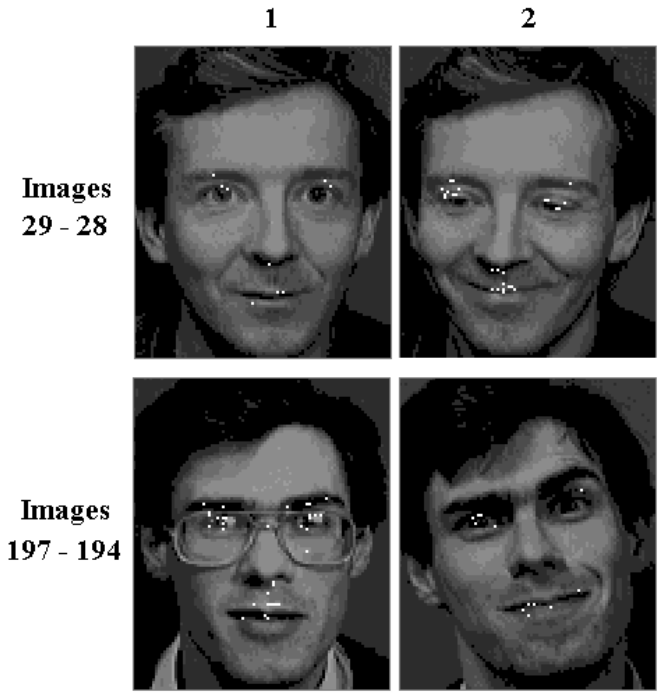


Fig. 4.. Two examples of composite maps for face images from ORL database in a frontal (1) and transformed (2) views.

[1] The ORL database is available at http://www.cam-orl.co.uk/facedatabase.html

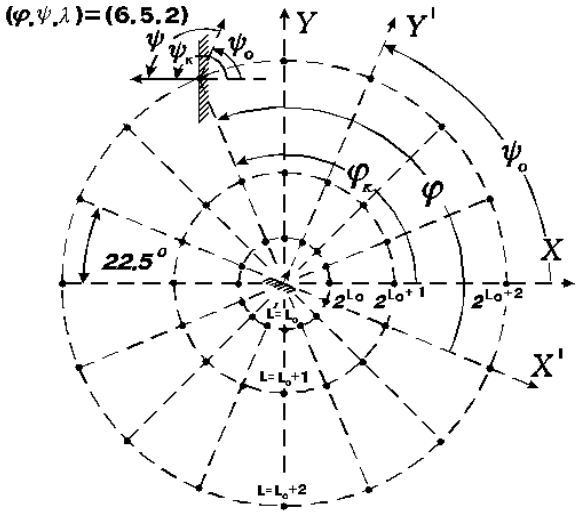
4. ALGORITHM OF IMAGE VIEWING BASED ON PRELIMINARY DETECTION OF THE MOST INFORMATIVE IMAGE REGIONS

Some fixation points of "attention window" selected by automatic algorithm developed earlier [22] were located outside the most informative image regions. It is evident that detailed processing must be performed mainly in such regions. Elaborated cascade method providing effective and fast detection of the perceptually important facial regions may be useful for reliable face recognition in natural scene by the MARR model. In this case, image viewing is performed after preliminary identification of perceptually important regions (Fig.5). For invariant representation of features located inside of different facial regions the present version of "attention window" is modified:
(1) co-ordinates of the most informative image regions are considered as one of main factor deterring gaze shift; they are included into algorithm of image viewing;
(2) resolution levels in space-variant structure of attention window became adjustable according to distance between most informative image regions.

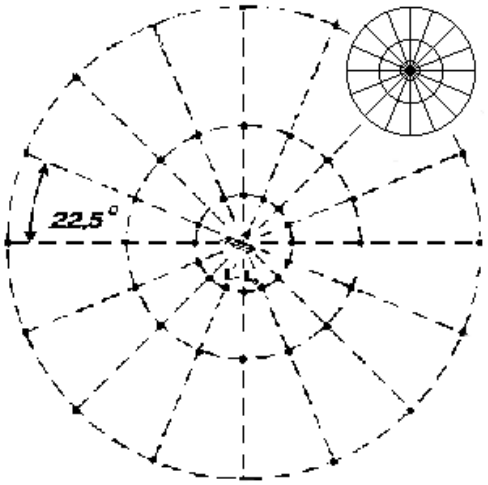
Fig.5 Image viewing trajectory based on prelimanary identification of perceptually important facial regions. The groups of white points detected by means of cascade method indicate on location of eye, nose, and mouth regions.

5. MODIFIED ALGORITHMS OF LOW-LEVEL PROCESSING IN THE MARR MODEL

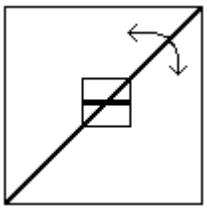
The results of psychophysical experiments described in Section 2 are used for development of new biologically-inspired algorithms for space-variant image representation in “attention window” in the MARR model. In particular, the obtained experimental data allows to choose 3 resolution levels for image representation, which are sharply different from each other. In this case the number of fixation points providing specific image description was sharply decreased (acceding to preliminary estimation , from 20 points in old model up to 3-5 ones). Besides, on the base of the data about spatial nonuniformity of sensory tuning (see Section 2) and neurophysiological data about composite structure of neuron receptive fields {23}. Local space-variant filter has been created for extraction of oriented segments on several resolution levels and their conjunctions in vicinity of each of 49 points of “attention window” (Fig.6). In previous version of the MARR model (Fig.6a), an oriented segment was identified in vicinity of each i,j-th point of attention window on one resolution level only. It is supposed that image fragment processing by proposed space-variant filters included in macrostructure of attention window allow to increase specificity of image representation and recognition accuracy in the MARR model.



a



b



c

Fig.6 Attention window structure in the MARR model.
a - previous version of the model, b - modified local space-variant filter, c - detection of two segments of different size by local space-variant filter.

6. CONCLUSIONS

The results of psychophysical experiments and several biological plausible algorithms for low-level processing facial images have been presented. Some of these algorithms were tested while processing natural facial images by the MARR model.
The obtained results can be summarized as follows:

(**1**) Model suggestion [18] about local spatial nonuniformity of sensory tuning and subject response time in the human peripheral vision field are confirmed in psychophysical experiments;

(**2**) cascade method for detection of most informative facial regions is developed. It provides high selective detection of perceptually important facial regions and high number of points of interest inside these regions. Thus, elaborated cascade method results in elimination of points of interest outside informative regions detected at preliminary stages and accumulation of points of interest inside of facial informative regions. As a whole, proposed cascade method may be compared with well-known averaging method used in signal processing area for increasing signal/noise ratio.

The core of the elaborated cascade method is formation of multiple particular feature maps (n=10 in the current implementation) mainly based on feature combinations extracted by means of local 5x5 filter. These combinations may correspond to local feature conjunctions in image regions. Thus, by imitating neurobiological mechanisms for detection of feature combinations such as hypercolumn composite sensory tuning and multiple representations of the same sensory surface in many fields of the visual cortex [13, 16, 20], perceptually important parts of face images [11, 29] may be identified by means of relatively simple bottom-up pre-attentive mechanisms. Elaborated algorithms and method providing effective detection of the perceptually important facial regions have several advantages important for various applications and hardware implementation such as high performance,. simplicity of basic computations, and possibility of parallel computations for obtaining multiple particular feature maps.

(**3**) Basing on above results, the algorithms for image viewing and local space-variant filter are proposed.

At present, these modifications of the MARR model are implemented. It is supposed that such modification can significantly improve the performance of the MARR model while processing real world images during memorizing, search, and recognition.

ACKNOWLEDGEMENTS

This work is supported by Russian Foundation of Basic Research, grant N 98-01-00948.

REFERENCES

1 J. Aloimonus, I. Weiss, and A. Bandyopadhyay, "Active vision," *Proc. of the Int. Joint Conf. on Computer Vision*, London, pp. 35-54, 1987.

2 R. Bajcsy, "Active perception," *Proc. of the IEEE*, vol. 76, pp. 996-1005, 1988.

3 C. Bandera and P.D. Scott, "Foveal machine vision systems," *Proc. IEEE Int. Conf. On Systems, Man and Cybernetics*, vol. 2, pp. 596-599, 1989.

4 Burt P.J. Attention mechanisms for vision in a dynamic world. In: Proc. 9th Int. Conf.on Pattern Recognition, Roma. 1988. pp.977-987.

5 M.C. Burl, T.K. Leung, and P. Perona, "Face Localization via Shape Statistics," *In Int'l Workshop on Automatic Face and Gesture Recognition*, Zurich, Switz, June 1995

6 P.J. Burt, "Smart sensing within a pyramid vision machine," *Proc. of the IEEE*, vol. 76, pp. 1006-1015, 1988.

7 Golovan A., Lee S-W, Podladchikova L., Shaposhnikov D. Formation of feature map for detection of most informative regions of facial images. In Proc. ICNC'99 , Rostov-on-Don , 27-29 September (accepted)

8 Gusakova V.I., Faure,A., Podladchikova L.N., Biologically plausible algorithms for image viewing. In Proc. ICNC'99 , Rostov-on-Don , 27-29 September 1999 (accepted)

9 P.J.B. Hancock, V. Bruce, and M.A. Burton, "A comparison of two computer-based face identification systems with human perceptions of faces," *Vision Research*, vol. 38, no. 15/16, pp. 2277-2288, 1998.

10 R. Hecht-Nielsen and Y.T. Zhow, "VARTAC: A Foveal Active Vision ART System," *Neural Networks*, vol. 8, no. 7/8, pp. 1309-1321, 1995.

11 J.M. Henderson and A. Hollingworth, "Eye Movements during Scene Viewing: An Overview," Michigan State University Eye Movement Laboratory, Technical Report, 5, pp. 1-18, 1997.

12 E. Hjelmas and J. Wroldsen. "Recognizing Faces from the Eyes Only," *Proc of the 11th Scandinavian Conference on Image Analysis*, 1999.

13 D.H. Hubel and T.N. Wiesel, "Sequence, regularity and geometry of orientation columns in the monkey striate cortex," *Journal of Comparative Neurology*, vol. 158, pp. 267-293, 1974.

14 M. Jagersandy, "Saliency maps and attention selection in scale and spatial coordinates. An information theoretical approach," *Proc. of 5th Int'l Conf. on Computer Vision*, pp. 195-202, 1995.

15 P. Kruizinga and N. Petkov, "Person identification based on multiscale matching of cortical images". In: B. Hertzberger and G. Serazzi, editors, *Proc. of the Int'l Conf. and Exhibition on High-Performance Computing and Networking*, HPCN Europe'95, Milan, Italy, May 3-5 1995, volume 919 of *Lecture Notes in Computer Science*, (Berlin: Springer-Verlag, 1995), pp 420-427, 1995.

16 Orban G.A. Neuronal operations in the visual cortex. In: Studies of brain function. Berlin-Heidelberg, N-Y, Tokyo. 1984.

17 W. Osberger and A.J. Maeder, "Automatic identification of perceptually important regions in an image," *Proc. 14th Int'l Conf. on Pattern Recognition*, Brisbane, Australia, August 1998, pp. 701-704, 1998.

18 Podladchikova L.N., Gusakova V.I., Shaposhnikov D.G., Faure A., Golovan A.V. and Shevtsova N.A. "MARR: Active Vision Model". In: Proc. of conference on Intelligent Robots and Computer Vision XIV: Algorithms, Techniques, Active Vision,and Materials Handling, Proc. SPIE, D.P. Casasent, Editor, 1997, 3208, 418-425.

19 R.P.N. Rao, G.J. Zelinsky, M.M. Hayhoe, and D.N. Ballard, "Eye movements in visual cognition: A computational study," Technical Report 97.1, University of Rochester, 1997.

20 A.W. Roe and D.Y. Ts'o, "Visual Topography in Primate V2: Multiple Representation across Functional Stripes," *The Journal of Neuroscience*, vol. 15(5), pp. 3689-3715, 1995.

21 H.A. Rowley, S. Baluja, and T. Kanade, "Human Face Detection in Visual Scenes," *Advances in Neural Information Processing System 9, Proc. of the 1996 Conference*, Michael C. Mozer, Michael I. Jordan and Thomas Petsche, eds., pp. 875-881, 1997.

22 I.A. Rybak, V.I. Gusakova, A.V. Golovan, L.N. Podladchikova, and N.A. Shevtsova, "A Model of Attention-Guided Visual Perception and Recognition," *Vision Research*, vol. 38, no. 15/16, pp. 2387-2400, 1998.

23 Rybak I.A., Podladchikova L.N., Shevtsova N.A., Golovan A.V., Gusakova V.I. Sandler V.M. Neural network modeling in studying neuron mechanisms of visual perception RNNS/IEEE Symposium on Neuroinformatics and Neurocomputers, Rostov-on-Don, pp.323-333,1992.

24 F. Smeraldi, A. Makarov, and J. Bigun, "Saccadic Search with Gabor features applied to Eye Detection," Technical report 98/256, Swiss Federal Institute of Technology, Computer Science Department, CH-1015 Lausanne, January 1998.

25 E.L. Schwartz, D.N. Greve, and G. Bonmassar "Space-variant Active Vision: Definition, Overview and Examples," *Neural Networks*, vol. 8, no. 7/8, pp. 1297-1308, 1995.

26 Shaposhnikov D.G., Podladchikova L.N. Local nonuniformity of the visual perception in the peripheral vision field, In Proc. IJCNN'99, Washington, 10-16 July 1999, (accepted)

27 B. Takacs and H. Wechsler, "A Dynamic and Multiresolution Model of Visual Attention and its Application to Facial Landmark Detection," *Computer Vision and Image Understanding*, vol. 70, no. 1, pp. 63-73, 1998

28 A.M. Treisman and G. Gelade, "A feature integration theory of attention," *Cognitive Psychology*, vol. 12, pp. 97-136, 1980.
29 Yarbus A.L. Eye movements and vision. Plenum Press. New York, 1967
30 K.C. Yow and R. Cipolla, "Towards an Automatic Human Face Localization System," *In Proceedings 6th British Machine Vision Conference*, vol. 2, pp 701-710, Birmingham, 1995.