ISSN 1060-992X, Optical Memory and Neural Networks (Information Optics), 2009, Vol. 18, No. 2, pp. 114–121. © Allerton Press, Inc., 2009.

Model-Based Approach to Study of Mechanisms of Complex Image Viewing[1]

L. N. Podladchikova[a], D. G. Shaposhnikov[a], A. V. Tikidgji–Hamburyan[b], T. I. Koltunova[b], R. A. Tikidgji–Hamburyan[a], V. I. Gusakova[a], and A. V. Golovan[a]

[a] *A.B. Kogan Research Institute for Neurocybernetics, Southern Federal University, Stacka Ave. 194/1, Rostov-on-Don 344090*

[b] *Psychology Department, Southern Federal University*

e-mail: lnp@nisms.krinc.ru

Received March 18, 2009; in final form, April 6, 2009

Abstract—A model-based approach to study complex image viewing mechanisms and the first results of its implementation are presented. The choice of the most informative regions (MIRs) is performed according to results of psychophysical tests with high-accuracy tracking of eye movements. For three test images, the MIRs were determined as image regions with maximal density of gaze fixations for the all subjects (n = 9). Individual image viewing scanpaths (n = 49) were classified into three basic types (i.e. "viewing", "object-consequent", and "object-returned" scanpaths). Task-related and temporal dynamics of eye movement parameters for the same subjects have been found. Artificial image scanpaths similar to experimental have been obtained by means of gaze attraction function.

Keywords: visual perception, attention, eye movements, most informative regions, viewing scanpath, psychophysical tests, computer simulation.

DOI: 10.3103/S1060992X09020088

INTRODUCTION

Since the pioneer A.L. Yarbus' works [1] it is well-known that the choice of the most informative image regions (MIRs) for gaze fixation during image viewing and their analysis by the foveal vision depends on a number of factors determined by perceptual and cognitive mechanisms [2–11]. As compared with cognitive mechanisms, contribution of the perceptual mechanisms into this process can be more directly estimated by image features and eye movement parameters. The currently well-accepted consistent facts about perceptual mechanisms of image viewing are:

(1) At preattentive level, image fragments with maximal brightness and contrast as well as big objects are preferable for gaze shift for a current fixation point [1];

(2) To activate visual attention, image regions must be located on some distance (no less than 2–10 degrees) from the current fixation point [1, 3];

(3) "Centre of mass" of the MIRs and the enveloping contour may be considered as important image properties for gaze positioning [2];

(4) Viewing scanpath topology is a specific attribute for the given image during memorizing, recognition, and mental imagery [7, 12, 13];

(5) Viewing scanpath details for the same image depend on the subject motivation (instruction) before psychophysical test [1].

Taking into account poor formalization and contradictions in the known experimental findings, mathematic models are considered as important tools to study image viewing mechanisms [3, 6, 12–21]. However, in most of modelling research contribution of some gaze shift factors is analyzed while other factors are ignored. In particular, some authors consider mainly primary image features [17–19], attention mechanisms [15, 16, 21] or relationship between sensory and motor memories [12, 13]. In the review paper [20] it is noted that development of realistic models to analyze contribution of various factors and mechanisms of different levels to MIRs choice remains an actual problem. Results of computer simulations with such models

[1] The article is published in the original.

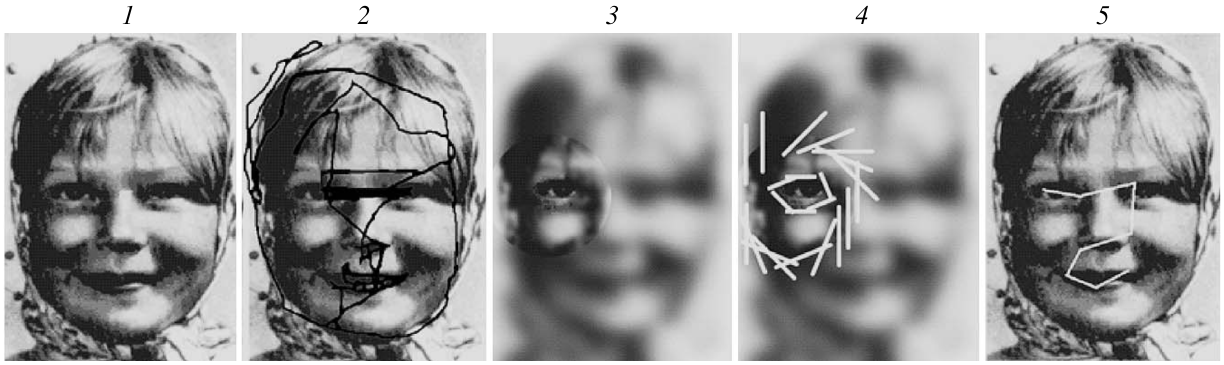

Fig. 1. An example of implementation of the model-based approach to formalize the psychophysical finding. (*1*) Test image Im1; (*2*) Viewing scanpath recorded by A.L. Yarbus [1, modified Fig. 115]; (*3*) Image fragment modification; (*4*) Space-variant analysis of context features around gaze fixation point; (*5*) Simulation of image viewing scanpath by feature-based gaze attraction function [17, 18].

may be useful for generation of suggestions about possible mechanisms of image viewing which can be verified experimentally.

In this paper, a model-based approach based on tight mutual connections between psychophysical experiments and modelling and first results of its implementation has been presented. Prospective goals of our study are: quantitative estimation of various factors received in standard experimental conditions and development of a realistic model of image viewing mechanisms. The preliminary results were presented in [22].

APPROACH AND METHODS

Common approach. Generally, the developed approach (Fig. 1) includes the following steps: (i) choice of published data with detailed protocols of psychophysical experiments; (ii) development of a computer model that can imitate the results of a particular experiment; (iii) analysis of local visual features inside image fragments different from each others by eye movement parameters and gaze fixations; (iv) image modification and a second series of psychophysical tests with the same images; (v) implementation of (ii) and (iii) for the second series of experiments; (vi) determination of gaze attraction function by detailed comparison of experimental and modelling results.

As compared with others studies that use both modelling methods and psychophysical tests [5, 6, 13, 15, 16, 20, 21], the basic peculiarities of our research [14, 17, 18, 22] are as follows: analysis of primary image features inside the MIRs determined as image regions with high density of gaze fixations points; detailed analysis of spatial structure of viewing scanpaths taking into account various objects location and temporal parameters of gaze shift; estimation of context features distribution in space-variant structure of input window.

Modelling methods. Similar to our previous studies [17, 18], the model presented here imitates key properties of the real visual system such as space-variant representation of visual information with reduced resolution from the vision field centre to the periphery, search for the perceptually important image regions for detailed analysis, context description of local features in this regions, and orientation selectivity of the visual cortex neurons.

The model consists of the following modules: input module, feature analysis, and scanpath formation. Each image point (x, y) is described by a set of features identified around its vicinity. 48 feature maps may be received such as: brightness; contrast; colour parameters in RGB and CIECAM colour spaces; number and distribution of particular oriented brightness edges and their angular combinations; etc. The size of central (foveal), para- and perifoveal input window parts and their resolution levels are determined according to the known neurobiological data [1]. Convolution matrixes inside these input window parts have differed dimensions. Initial data are used for image representation at the first resolution level, to present the image at the second and third resolution levels initial image is transformed by recursive Gaussian-like convolution [23]. Edge contrasts and orientations at each resolution level are determined as a difference between two Gaussian convolutions with spatially shifted centres [24]. The discrete angle step of 22.5° is considered as a unit in all angular measurements.

Several algorithms [17, 18] for gaze attraction function and consequent formation of the viewing scanpath are included into the model: (i) supervision algorithm; (ii) MIRs/object-based frame; (iii) feature-based frame [17, 18]; (iv) detailed gaze attraction function based on the multiple parameters that is described as (1):

$$F(x, y, h) = \sum_{i=0}^{n} k_i f_i(x, y) + V(Q(x, y), r_F(x, y), r_P(x, y)) + CFix(x, y, h), \quad (1)$$

where:

$f_i(x, y)$, $i = 1, \ldots, n$ are feature maps and k_i are normalization coefficients for each feature map;

$V(Q, r_F, r_P)$ is a function characterizing a structure of the visual field;

$Q(x, y)$ is a map of the visual field peculiarities;

$r_F(x, y)$ are a set of coordinates of the central (foveal) part of the visual field;

$r_P(x, y)$ are a set of coordinates of the peripheral part of the visual field;

C is a coefficient determined for a current visual task;

$Fix(x, y, h)$ is a distribution of the previous $(h - 1)$ fixation points.

Because each algorithm for formation of feature maps in $f_i(x, y)$ has its own range of values, the normalization coefficients k_i are included into (1) for each map, restricting their weighted sum to interval from 0 to 1. The function $V(Q, r_F, r_P)$ gives an integral description of the visual field structure, defined by results of psychophysical tests and includes several components, namely $Q(x, y)$ which is determined by the peculiarities of the subject's visual field structure including the local spatial non-uniformity of sensory tuning in the peripheral vision field [25], $r_F(x, y)$ and $r_P(x, y)$ which describe parameters of different parts of the subject's visual field. Function $Fix(x, y, h)$ is formed dynamically according to the results analysis of image fragments at the previous gaze fixations and may be described as a set of coordinates of consequent gaze fixations or as their weighted sum.

For gaze fixation points grouping and scanpath classification, the nearest neighbour method [26] and a set of determinative rules have been used.

Experimental methods. Experimental setup allowed us the following: to present simple visual stimuli, integral objects and complex images in different parts of the vision field; to modify the initial images (blurring of image and its fragments, noising, scaling etc.); to record the audio and video data of experiment process with detailed report of the subjects; to take into account the individual peculiarities of the vision field of a subject; to receive visualization and the quantitative analysis of the experimental results (coordinates of gaze fixations, viewing scanpaths, analysis of eye movement parameters). The iView X Hi-Speed 1250 eye tracker is used to record the eye movements. Some illustrations from [1] which may be quantitatively analyzed were chosen as basic test stimuli, namely the photo image "Volghanochka", S. Fridland (Im1), the paintings: "Ne zhdali", I.E. Repin (Im2), "V lesu grafini Mordvinovoy", I.I. Shishkin (Im3), and "Berezovaja rosha", I.I. Levitan (Im4). 9 volunteers participated in the experiments. The images Im2–Im4 were presented for viewing with different instructions, i.e.: "free viewing", "objects viewing", "estimation of objects relationship", and "search for modified image fragments". Possible artefacts such as trend, blinks, short saccade (less than 0.2°), and others have been rejected from the analysis of experimental data.

EXPERIMENTAL RESULTS AND DISCUSSION

Identification of most informative regions. As a first step to identify the image MIRs, the A.L. Yarbus' experimental data [1] during viewing image Im2 were analyzed. Quantitative estimation of the data presented in Fig. 109, 1 [1] by the nearest neighbour method [26] and their projection onto the initial Im2 allowed to detect the MIRs as areas with maximal density of gaze fixation points (Fig. 2a). To receive the quantitative estimations that cannot be extracted from Yarbus experimental data, a set of psychophysical tests have been carried out with Im2–Im4 as test images. The MIRs similar to the A.L. Yarbus' results were identified during viewing of Im2 (compare Figs. 2a and 2b). Perceptually important objects on each test image were determined on the base of the identified MIRs (Fig. 2c).

Image viewing scanpath classification. During visual analysis of the illustrations in [1], three basic types of image viewing scanpaths may be identified (Figs. 3a and 3b) by qualitative analysis. In particular, wide distribution of gaze fixation points and its relatively low density in the MIRs are characteristic for the "viewing" type of scanpaths but two other type have high gaze fixation density in the MIRs and are distinguished by transition frequency from one MIRs to another. The transition frequency (returns) is less for "object-consequent" scanpath type as compared with "object-returned" one. Similar viewing scanpaths were revealed in our psychophysical tests (compare Figs. 3a and 3b and Fig. 3c).

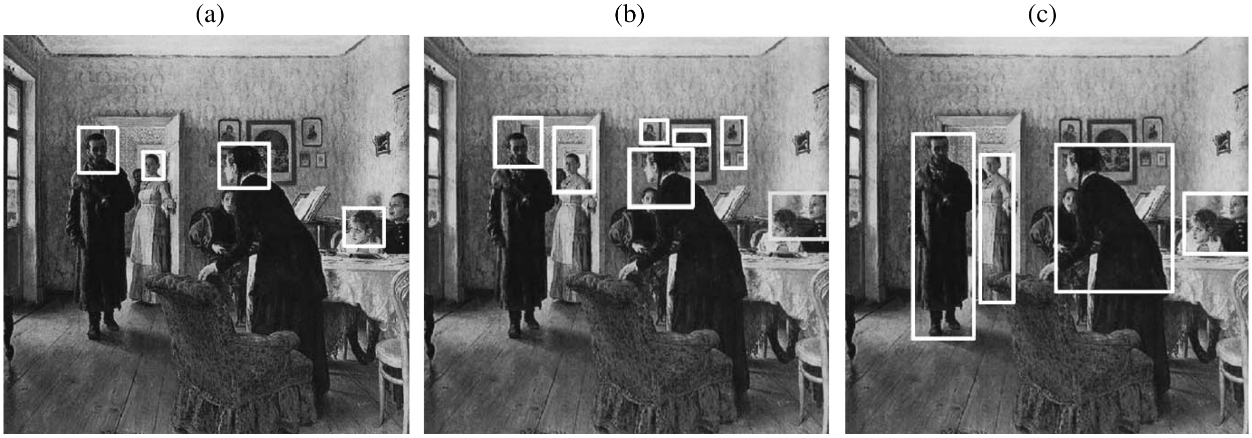

Fig. 2. MIRs detected by distribution of fixation points during viewing of Im2. (a) MIRs identified by analysis of A.L. Yarbus' results; the data for one subject have been used [1, Fig. 109, 1]. (b) MIRs identified in our psychophysical tests; cumulative data for 9 subjects have been used. (c) The detected objects. In each case, the MIRs or objects with MIRs are marked by white rectangles.

The quantitative criteria for trajectories classification have been obtained by a set of artificial scanpaths ($n = 150$) divided into three types by the experimentalist. During the artificial scanpath generation each next gaze fixation point was defined as follows: gaze jump amplitude (from the size of the human foveal part of the vision field to the perifoveal one) and shift angle between two gaze fixations were chosen randomly. The attraction function was estimated for chosen point based on the image MIRs and perceptually important objects locations, shift angle with maximal weighted value was chosen for 10 iterations. Any artificial scanpath varying from chaotic to the object-oriented can be received by changing attraction coefficients in this algorithm. Some examples of artificial scanpaths similar to experimental are presented in Fig. 3d.

Each artificial scanpath was described by number of fixations inside the objects, number of gaze jumps inside and between the objects. Scanpaths division onto two groups, i.e. "viewing" and "object-oriented" may be successfully achieved by estimation of the relation of a number of fixation points inside objects to all recorded fixation points. In particular, for viewing scanpath less than a half of fixation points was located inside objects, for object-oriented scanpaths this relation was no less than 0.85. To divide the object-oriented scanpaths onto object-consequent and object-returned types, a number of gaze jumps inside the object and between the objects were estimated, their relation was equal to 10 in the first case and 0.2 in the second one.

For classification of particular experimental viewing scanpaths ($n = 49$) we have used the results of psychophysical experiments that included more than 90 typical gaze fixation points. It was found that the procedure described above provides scanpath classification with the probability ($p = 0.86$) similar to evaluation of a skilled experimentalist.

Similar to A.L. Yarbus' study [1] the scanpath structure dependence on initial subject instruction was observed in our experiments. Namely, viewing scanpaths dominated in the experiments with the "free viewing" instruction but object-consequent scanpaths were more presented in the tests with the "objects viewing" instruction. In these experimental conditions, a difference has been revealed in gaze fixation duration and saccades amplitude (382 ± 82 ms against 550 ± 162 ms; $2.11° \pm 0.36°$ against $1.44° \pm 0.36°$, for "free viewing" and "objects viewing", correspondingly).

Temporal dynamics of image viewing. Visual analysis of psychophysical data showed that gaze fixations, eyes movements and viewing trajectory had a temporal dynamics during a test. In particular, the first 20 fixations had shorter duration and amplitude of preceding saccades than subsequent ones. Approximately, a half of the first 20 fixations were located in the MIRs, while the subsequent ones were widely spread throughout the image. This data indicate that preattentive mechanisms may dominate at first stages of image viewing.

Temporal dynamics of eye movement parameters during image viewing

Eye movement parameters	1st–20th fixations	21st–40th fixations	41st–60th fixations	61st–80th fixations
Fixation duration, ms	293 ± 11	351 ± 16	329 ± 15	320 ± 14
Saccade amplitude, degrees	2.99 ± 0.10	3.60 ± 0.17	3.06 ± 0.11	3.54 ± 0.14

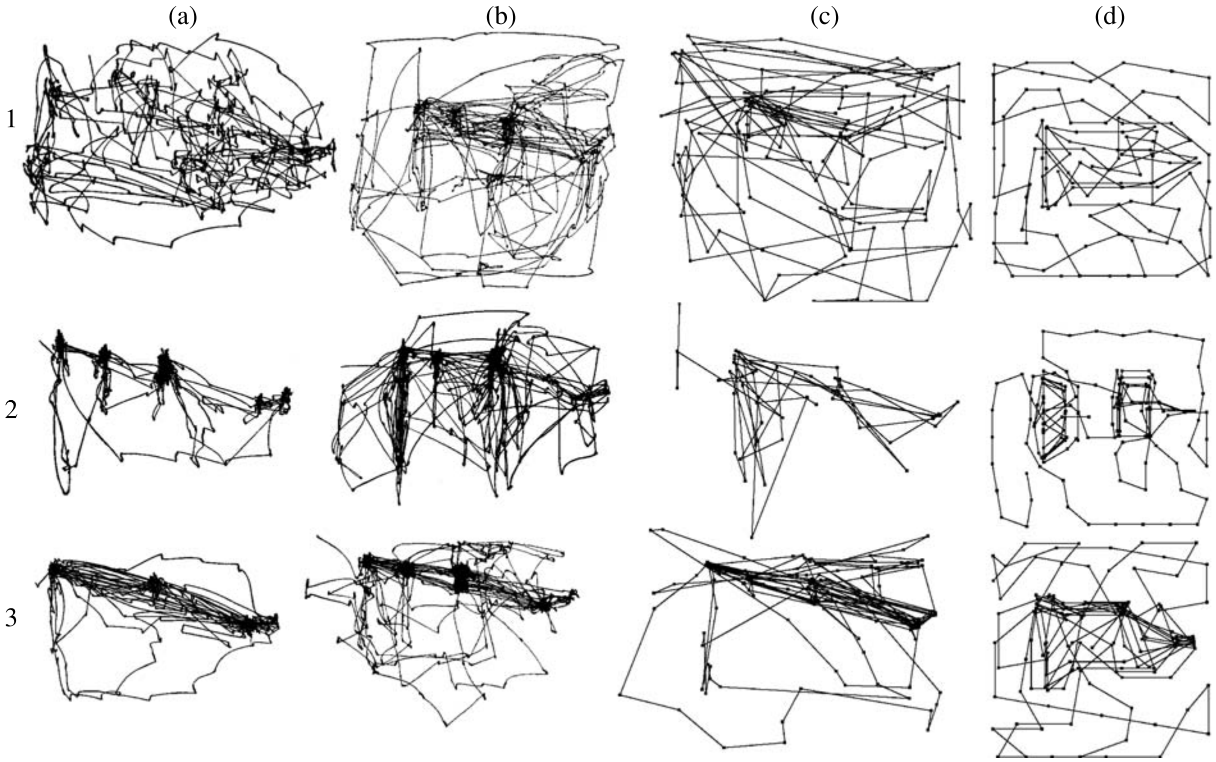

Fig. 3. Examples of viewing scanpaths for Im 2. (a) and (b) the A.L. Yarbus' results [1] (the excerptions from Figs. 109 and 107). (c) Our psychophysical results. (d) Artificial scanpaths. (1), (2), (3) The scanpaths of "viewing", "object-consequent", and "object-returned" types correspondingly.

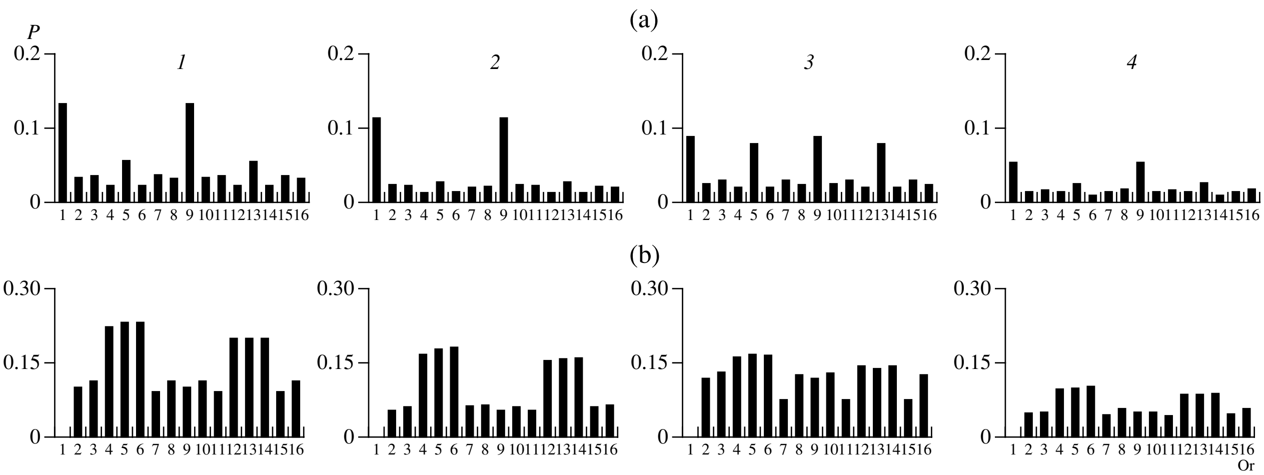

Fig. 4. Histograms distribution of oriented segments (a) and their angular combinations for Im2 fragments that differed from each others by density of gaze fixation points.

Analysis of primary local features. It was found that image fragments with differed gaze fixation density had no difference from each other for most primary features and their combinations. However, for some features the areas with maximal and minimal density of gaze fixations show a dramatic difference (compare columns (1) and (4) in Fig. 4). The data shown in Fig. 4 are consistent with our previous assumption that a set of differently oriented edges with high contrast may be one of gaze attraction factors at the preattentive level, and the results of its verification in psychophysical tests with simple stimuli [14]. Seemingly, this factor contributes into choice of fragments for next gaze fixation during complex image viewing, too. The regions with high density of gaze fixation points often have a big diversity of features not only in the central (foveal) part of input window but also in the nearby one.

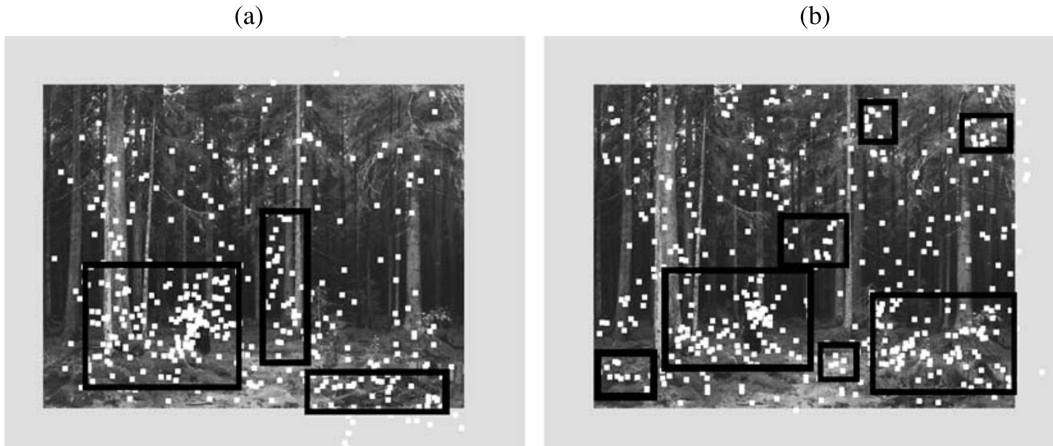

Fig. 5. The MIRs (marked by black rectangles) detected by density of gaze fixation points during free viewing of Im3 (a) and search for modified fragments (b). Cumulative data for 4 subjects have been used.

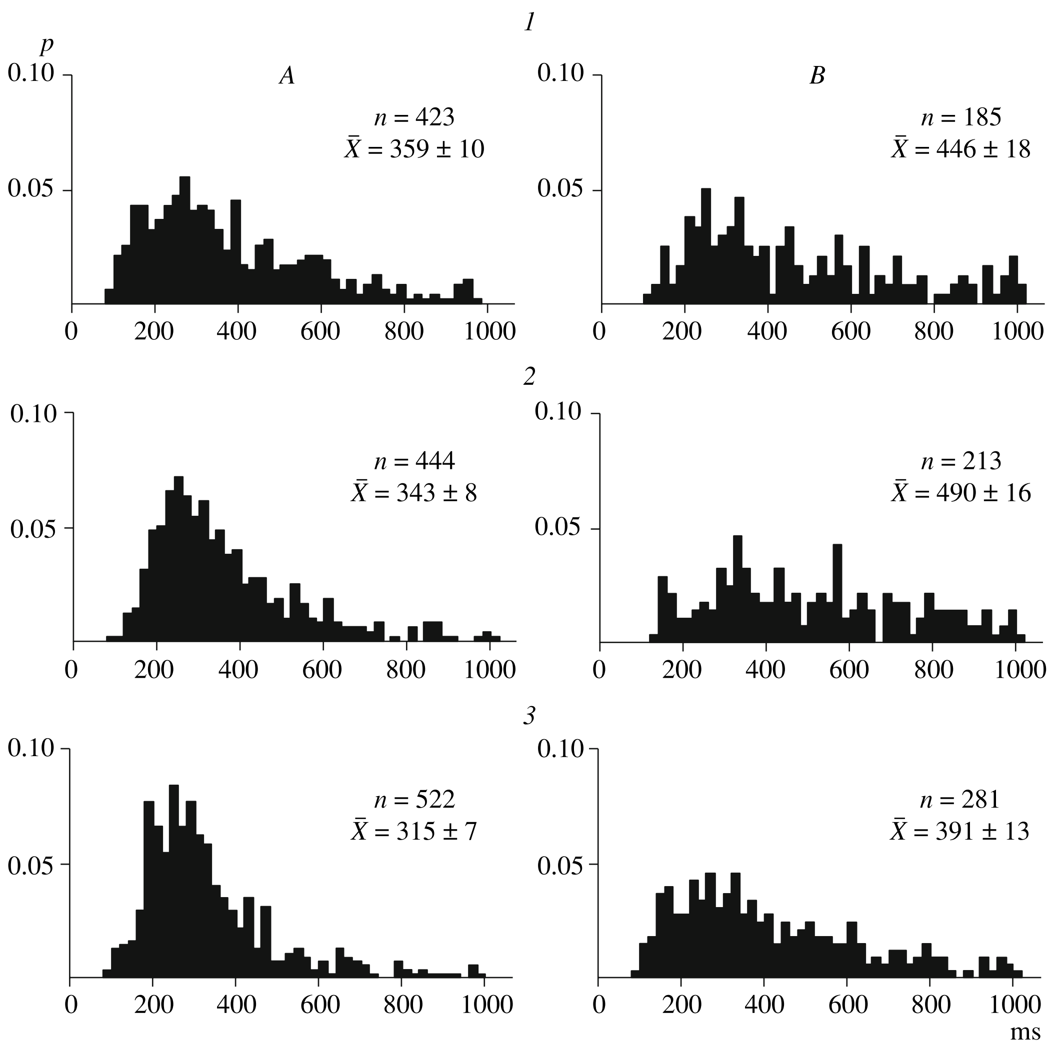

Fig. 6. Averaged histograms of distribution of gaze fixation durations during free viewing (*A*) and search for modified fragments for Im2 (*1*), Im3 (*2*), and Im4 (*3*). For each histogram the number of fixation points (n) and the mean duration of gaze fixations ($\bar{X}$) are shown.

The search for modified image fragments. In the other set of psychophysical tests the changed (blurred) local fragments of the images Im2, Im3, and Im4 were offered to search for. Such fragments were located in the image regions distinguished by the density of gaze fixations in the experiments with free image viewing. The size of modified fragments imitated central (foveal) part of the vision field (1.2–1.3 degrees). The results of two psychophysical test series (free-viewing and search for modified fragments) were compared by MIRs locations, type of eye movement scanpath, and duration of gaze fixations. The basic MIRs in the same images, identified in two experimental sets, seem to be topologically similar, however have several peculiarities (Fig. 5). Besides, in a half of cases the same subjects demonstrated similar type of eye movement scanpath during image free viewing and search for modified fragments. The differences in distribution of gaze fixation durations and their mean values were found between these tests (Fig. 6). Revealed differences between "free viewing" and "search" experiments indicate on specific strategies of the visual perception during solution of these visual tasks that may be a reason for contradiction in the published data about contribution of local visual features to gaze shift (compare [11] and [14]). It is important that contribution of these strategies may be quantitatively estimated by eye movements parameters.

CONCLUSIONS

The first results of model-based approach to study image viewing mechanisms are as follows: the results correlated with A.L. Yarbus results [1] have been obtained in psychophysical experiments with high-precision technique for eye movement registration; three basic types of viewing scanpath were identified and quantitative criteria for their classification were obtained by mean of gaze fixation points density and sequence; an algorithm for gaze attraction function were developed and used to build viewing scanpath similar to experimental; sets of some primary features are significantly distinguished in fragments differed by gaze fixation point density; the fragments with high density of gaze fixation points have more variously features distribution in foveal and parafoveal parts of the vision field.

The main result is identification of three types of image viewing scanpaths, which may reflect domination of different mechanisms of visual perception, such as spatial visual attention, focal attention and estimation of semantic relations between objects. The quantitative estimation of contribution of these mechanisms at different stages of visual perception for various visual tasks may be done by means of viewing scanpath structure assessment. The domination of one or another mechanism may correlate with monitored parameters of eye movements. Evidently, the image properties have significant influence on these parameters. This fact suggest a development of criteria for image description and analysis. In particular, an criteria for image classification by semantic importance and eye movement parameters may be found during further development of our approach.

It is assumed that described approach and the developed model will allow to overcome the limitations in generalization and formalization of psychophysical data which are caused by multifactor dependence and incompleteness or inconsistency of data. Also our approach makes possible to receive quantitative estimation of contribution of several "determinants" [2] to the gaze shift.

ACKNOWLEDGMENTS

The authors thank B.M. Vladimirski for consultation and help in the statistical analysis of the experimental data. The work is supported in part by the Russian Foundation for Humanities, grant 09-06-95218 a/F.

REFERENCES

1. Yarbus, A.L., *Role of Eye Movements in Vision*, Moscow: Nauka, 1965 [in Russian].
2. Barabanshikov, V.A., *Oculomotor Structures of the Perception*, Moscow: Press "Institute of Psyhology RAS," 1997 [in Russian].
3. Didday, R.L. and Arbib, M.A., Eye Movements and Visual Perception: A Two Visual Systems Model, *Int. J. Man-Machine Studies,* 1975, no. 7, pp. 547–569.
4. Gippenreiter, Y.B., Eye Movements in Man Activities, 2004, http://virtualcoglab.cs.msu.su/Gippenreiter [in Russian].
5. Hayhoe, M. and Ballard, D., Eye Movements in Natural Behavior, *Trends in Cognitive Science,* 2005, vol. 9, no. 4, pp. 188–193.
6. Henderson, J.M., Falk, R., Minut, S., Dyer, F.C., and Mahadevan, S., Gaze Control for Face Learning and Recognition by Humans and Machines, *Technical Report, Michigan State University Eye Movement Laboratory,* 2000, vol. 4, pp. 1–14.

7. Mast, F.W. and Kosslyn, S.M., Eye Movements During Visual Mental Imagery, *Trends in Cognitive Science,* 2002, vol. 6, p. 7.
8. Rajashekar, U., Van der Linde, I., Bolik, A.C., and Cormack, L.K., Foveated Analysis of Image Features at Fixations, *Vision Research*, 2007, no. 47, pp. 3160–3172.
9. Rehder, B. and Hoffman, A., Eye Tracking and Selective Attention in Category Learning, *Cognitive Psychology,* 2005, no. 51, pp. 1–41.
10. Velichkovsky, B.M., The Successes of Cognitive Science, *Scientific American*, 2003, no. 12, pp. 87–93 [in Russian].
11. Wolfe, J.M., Birnkrant, R.S., Kunar, M.A., and Horowitz, T.S., Visual Search for Transparency and Opacity: Attentional Guidance by Cue Combination? *J. Vision*, 2005, no. 5, pp. 257–274.
12. Noton, D. and Stark, L., Scanpaths in Eye Movements during Pattern Recognition, *Science*, 1971, vol. 171, pp. 72–75.
13. Priviterra, C.M. and Stark, L., Scanpath Theory, Attention and Image Processing Algorithms for Prediction of Human Eye Fixations, *Neurobiology of Attention*, Elsevier, Academic, 2005, pp. 663–670.
14. Gusakova, V.I., Shaposhnikov, D.G., and Podladchikova, L.N., Possible Mechanisms of Image Viewing: Model and Experiment, *Proc. First All-Russian Science and Technology Conf. "Neuroinformatics-99,"* Moscow, 1999, vol. 2, pp. 148–154 [in Russian].
15. Navalpakkam, V. and Itti, L., Modeling the Influence of Task on Attention, *Vis. Res.*, 2005, vol. 45, pp. 205–231.
16. Oliva, A., Torralba, A., Castelhano, M.S., and Henderson, J., Top-Down Control of Visual Attention in Object Detection, *Proc. Int. Conference on Image Processing (ICIP)*, Barcelona, Spain, 2003, vol. 1, pp. 253–256.
17. Rybak, I., Gusakova, V., Golovan, A., Podladchikova, L., and Shevtsova, N., Attention-Guided Recognition Based on "What"; and "Where" Representations: A Behavioral Model, *Neurobiology of Attention,* Elsevier, Academic, 2005, pp. 663–670.
18. Rybak, I.A., Gusakova, V.I., Golovan, A.V., Podladchikova, L.N., and Shevtsova, N.A., A Model of Attention-Guided Visual Perception and Recognition, *Vision Research, Special Issue on Models of Recognition*, 1998, vol. 35, no. 15/16, pp. 2387–2400.
19. Treisman, A.M. and Gelade, G., A Feature Integration Theory of Attention, *Cog. Psychol.*, 1980, no. 12, pp. 97–136.
20. Walther, D.B. and Koch, C., Attention in Hierarchical Models of Object Recognition, *Prog. in Brain Res.*, Elsevier, 2007, vol. 165, pp. 57–78.
21. Zelinsky, G.J., Specifying the Components of Attention in a Visual Search Task, *Neurobiology of Attention*, Elsevier, 2005, pp. 395–400.
22. Podladchikova, L.N., Shaposhnikov, D.G., Tikidji–Hamburyan, A.V., et al., Integrated Approach to Study the Mechanisms of Images Viewing, *Proc. XI All-Russian Science and Technology Conf. "Neuroinformatics-2009,"* Moscow, 2009, vol. 1, pp. 249–257.
23. Burt, P.J., Smart Sensing Within a Pyramid Vision Machine, *Proc. IEEE*, 1988, no. 76, pp. 1006–1015.
24. Marr, D., *Vision*, New York: W.H. Freeman, 1982.
25. Shaposhnikov, D.G., Gizatdinova, Yu.F., and Podladchikova, L.N., The Peculiarities of Visual Perception in the Peripheral Vision Field, *Pattern Recognition and Image Analysis,* 2001, vol. 11, no. 2, pp. 376–378.
26. Pavlidis, T., *Algorithms for Graphics and Image Processing,* Rockville, Maryland: Computer Science Press, 1982.