CONTEXT-DEPENDENT PERCEPTION FOR FOVEAL MACHINE VISION SYSTEMS

D.G. Shaposhnikov, A.I. Samarin, S.I. Anishenko

A.B.Kogan Research Institute for Neurocybernetics, South Federal University,
194/1 Stachka Ave., Rostov-on-Don, 344090, Russia, e-mail: dima@nisms.krinc.ru

Several applications of artificial foveal visual systems in image processing are considered. The basic attention is focused on description of the Behavioral Model of Vision (BMV) developed in A.B. Kogan Research Institute for Neurocybernetics. The procedures for most informative region detection and space-variant context image representation are presented. The results of the BMV testing while recognition of facial (n=400) and traffic sign (n=202) images show a high recognition rate (95% for facial images and 97% for traffic signs) and demonstrate invariance to the point of view, noise and brightness for both image types. In addition, while processing PET situation images, the BMV detects facial landmarks (eye corners and middle point of nose basement) with very high accuracy (1±0.64 pixels).

Introduction

Many problems of computer and robot vision are far away from their effective solution in the traditional frameworks. Among the prospective approaches to solve these problems, one of the most widely acknowledged approach is simulation of biological vision mechanisms. In particular, artificial Foveal Visual Systems are currently most widely studied and developed [1-6 and many others].

Foveal systems imitate space-variant visual acuity, changing from the centre of the retina (the fovea) to its periphery, and attention mechanisms for human gaze control while image viewing. Computational advantages of Foveal Vision consist in a sharp decrease of information volume for detailed processing and possibility of context encoding of local visual information [6]. Object recognition in a natural scene is one of important areas in real world image processing. It demands a fast and reliable solution of many particular visual tasks, such as face region segmentation, person identification, traffic sign recognition, etc. Some of these tasks, changing over time, have no effective solution in frameworks of both biologically inspired and conventional approaches up to now. One of possible ways to solve computational problems of real world image processing is to develop system architectures combining biologically plausible approach with classical methods. The basic problem that should be solved to create an effective artificial visual foveal system is a development of high-performance algorithms for following tasks:

(1) Low-level image representation;

(2) Detection of informative fragments for detailed processing while image viewing (Most Informative Regions, MIR [7]).

In the present paper, a biologically plausible model and algorithms for solution of problems (1) and (2) are considered.

Behavioral Model of Vision

According to the Foveal Vision approach, a Behavioural Model of Vision (BMV) for invariant representation and recognition of images was developed [8,9]. The BMV [9] includes two main parts: 1) cascade method for detection of MIRs; 2) the algorithm for image description by a number of multidimensional vectors $\vec{F}$ formed by fixations of space-variant input window (Fig. 1) in detected informative fragments. Let us consider the main algorithms of the BMV and its application to solve some practical tasks.

Cascade method for the MIRs detection

The cascade method includes several procedures performed in consecutive order to concentrate potential "points of interest" in an image into the MIRs:

1. Detection of potential "points of interest" determined by specific combinations of primary features characteristic for the MIRs.
2. Density thresholding and grouping of potential "points of interest" into particular feature maps.
3. Classification of potential "regions of interest" by average histograms of oriented elements characteristic for each MIR.

The procedures developed provide an accurate detection of MIRs (95% of points of interest are located in MIRs) and determine sizes and centre coordinates of the regions of interest.

Invariant encoding of local features

The input window (IW) represents a structure (Fig. 1) containing one basic node (the central one) and 48 context (peripheral) nodes that are located at intersections of 3 concentric circles imitating spatially non-uniform retinal structure (different levels of resolution) and 16 rays radiating from the centre.

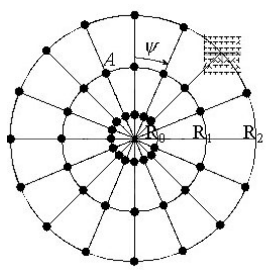


Fig. 1. Structure of space-variant input window with detailed description of context areas around each node (one of them is shown as a rectangular grid).

Feature description of each informative region is provided by multidimensional vector $\vec{F}$ formed by the space-variant IW. Feature vector $\vec{F}$ is determined by dominating edge orientation α in the vicinity of each of 49 IW nodes A_i, i=0, 1…48 and its density.

Let $x_0 = X_0$, $y_0 = Y_0$ be co-ordinates of the central IW node, then co-ordinates (x_i, y_i) of peripheral node A_i, i=1, 2…48 can be determined as follows:

$$\begin{aligned} x_i &= X_0 + R_i \cos \psi_k, \\ y_i &= Y_0 + R_i \sin \psi_k \end{aligned} \quad (1)$$

where R_l , l=0,1,2 is the radius of the l-th concentric circle of the IW (R_0 = 3 pixels, R_1 = 9, R_2 = 15) and $\psi_k = k \cdot 22.5°$, k= 0,1…15 is the angle of the radiating line corresponding to the i-th IW node. Space-variant resolution here is emulated by Gaussian convolution of image fragments with different size kernels depending on distance from the IW centre (5x5 pixels for the central part of the IW, 7x7 - for the intermediate, and 9x9 - for the peripheral part). Each IW node is characterized by edge orientation α that dominates in the node context area (7x7 pixels) and its density ρ as follows:

$$\begin{aligned} &\rho(A_i) = \max_{\varphi} \rho_{\varphi}(A_i) \\ &\alpha(A_i) = \varphi \ \ if \ \ \rho_{\varphi}(A_i) = \rho(A_i) \end{aligned} \quad (2)$$

where

$$\begin{aligned} &\rho_{\varphi}(A_i) = \rho_{\varphi}(x_i, y_i) = \\ &= \frac{1}{S(x_i, y_i)} \sum_{m,n} Sg_{\varphi}(Or(m + x_i, n + y_i)) \\ &Sg_p(x) = \begin{cases} 1; \text{ if } \ x = p \\ 0; \text{ otherwise} \end{cases} \end{aligned} \quad (3)$$

and $Or(x,y)$ is a detected edge orientation (not only dominating one) in the vicinity of the image element with co-ordinates (x,y); $S(x_i, y_i)$ is the context area (see Fig. 1) for the i-th IW node (equal to 49 pixels); m,n = -3,…0,…+3; φ = 0,1…15. The resulting feature vector $\vec{F}(\vec{\alpha}, \vec{\rho})$ is therefore formed as in Eq.4

$$\vec{F}(\vec{\alpha}, \vec{\rho}) = (\alpha(A_0) \ldots \alpha(A_{48}), \rho(A_0) \ldots \rho(A_{48})) \quad (4)$$

During recognition the incoming vector for each fixation of IW in input image MIRs is compared to prototype feature vectors stored in a database by Eq. 5:

$$K^b = \sum_{i=0}^{i<49} \left[\mathrm{sgn}\left(Or_i^b - Or_i^{rw}\right) \cdot \left(1 - abs\left(\rho_i^b - \rho_i^{rw}\right)\right) \right] \qquad (5)$$

where: $\mathrm{sgn}(x) = \begin{cases} 1, & \text{if } x = 0; \\ 0, & \text{in other case;} \end{cases}$

where Or_i is the dominating orientation of contrast segments in the context area of the given IW node (orientation is determined by step 22.5° and denoted as 0, 1, 2, 3…., 15); indices b denote prototype vectors, indices rw – a current vector; ρ - density of dominating contrast oriented segment in the context area of each IW node.

Applications of the BMV

Algorithms and methods of the BMV were used as a basis for development of several applications, namely, face recognition system (FOSFI) [10], traffic signs recognition system (FOSTS) [11] and head movement detection system [12].

Application of the BMV model to face recognition problem is described in details elsewhere [8,10]. Here we present the results of face recognition system in some range of image transformations. **FOSFI** was tested with variations of point of view about $\pm 30^o$ in rotation and $\pm 15^o$ in tilt; size change ±50%, and brightness fluctuations ±30%. Recognition rate for images without point of view transformations was 98%. In FOSFI facial images were represented by fixation of the IW in the centre of each detected MIR.

FOSTS was tested while recognition of traffic signs taken under various environmental conditions (cloudy or sunny weather, etc.) from different lines of movement using a digital camera [4]. Extraction of traffic sign images from a background was performed by means of colour segmentation by the model CIECAM97 [11]. A standard database of traffic sign (n=105) was used to obtain the prototype image descriptions.

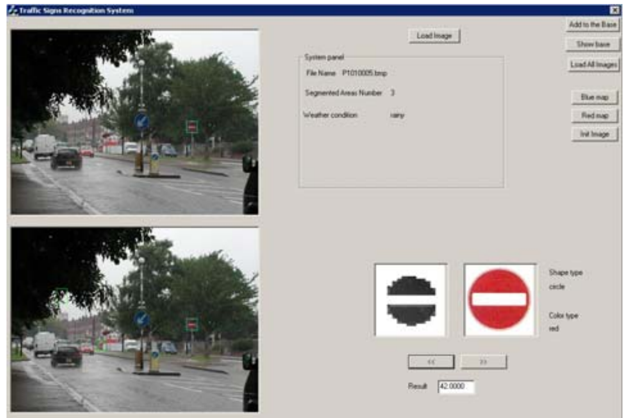


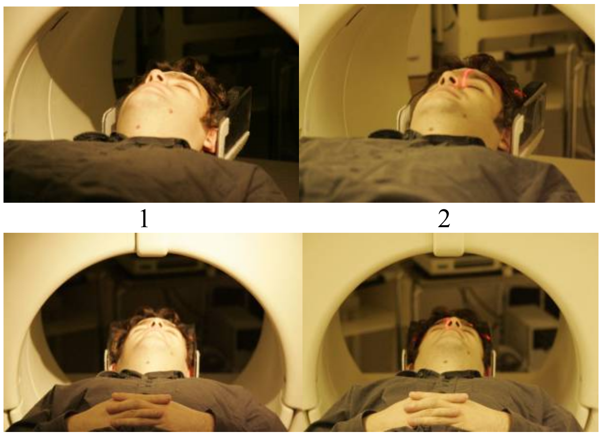
Fig. 2. Program realization of the FOSTS

The basic modifications of the BMV for traffic sign recognition include changing the structure of the IW. In particular, the size of the IW was increased into two times and an image was described just by one fixation of the IW in the centre of traffic sign informative part. The fixation point was chosen in the geometrical centre of colour elements describing sign contour.

The results of the computer experiments have shown that the algorithms developed provide recognition of traffic sign images invariantly to weather and viewing conditions and with the recognition rate of about 97%. Recognition errors appear for traffic sign images obtained from larger distances (more than 50m) and/or with essential shielding (more than 50%) of informative part of a sign.

The next application of the BMV was detection of head movement on the pictures recorded via digital cameras monitoring the scanning PET procedure [13] (Fig. 3). The BMV was used to detect local facial landmarks.

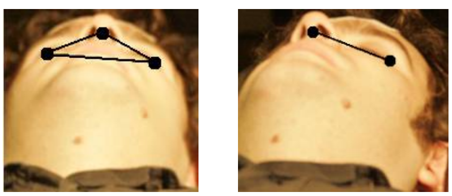
The developed algorithms were evaluated on the pictures (n=12) monitoring a subject's head while simulating PET scanning captured by two calibrated cameras (located in front and at the left of a subject). It has been shown that centres of chosen facial landmarks (eye corners and middle point of nose basement) have been detected with very high precision (1±0.64 pixels) on faces segmented by CIECAM97 [9].



1 2

3 4

Fig. 3. Examples of PET brain scanning. Viewing parameters: 1) left camera, distance 1.175 m, lightness 353 cd/m2; 2) left camera, distance 1.175 m, lightness 32.7 cd/m2; 3) front camera, distance 2.350 m, lightness 353 cd/m2; 4) front camera, distance 2.350 m, lightness 32.7 cd/m2.



(a) (b)

Fig. 4. Examples of detection of facial landmarks on the pictures captured by the front (a) and left side (b) cameras. Identified landmarks are marked by black spots.

Three landmarks on the frontal picture and two on the left side picture have been identified (Fig. 4). Preliminary results for 2D images with known movement parameters suggest that parameters of rotations and translations along X, Y, and Z directions can be obtained very accurately by the described methods.

Conclusion

The foveal models (in particular, the BMV) can be successfully applied to the solution of various tasks in image processing and recognition. In future, we plan to adjust the developed algorithms and methods for processing the real-time images and video sequences.

This work is supported in part by Russian Foundation for Basic Research, Grant No. 05-01-00689.

References

1. Ahrns, I., Neumann, H. Space-Variant Image Processing. // Proc. in Artificial Intelligence. - 2000. – No. 9. – P. 203-206.
2. Bandera C., Scott P. Machine vision for pursuit. The foveal alternative. // Journal of Electronic Defense. - 1991. - Vol. 14, Iss. 10. - P.61-66.
3. Hecht-Nielsen R., Zhow Y.T. VARTAC: A Foveal Active Vision ART System. // Neural Networks. – 1995. - Vol. 8, No. 7/8. – P. 1309-1321.
4. Salgian A.G., Ballard D. Visual Routines for Vehicle Control. // The Confluence of Vision and Control. Springer Verlag, 1998.
5. Samarin A.I. Bionic model of a “eye-hand” system for robot visual-motor coordination. // Bionica-78. Moscow: USSR Academia of Science – 1978. – Vol.1. (in Russian).
6. Schwartz E.L., Greve D.N., Bonmassar G. Space-variant Active Vision: Definition, Overview and Examples. // Neural Networks. – 1995. - Vol. 8, No. 7/8. – P. 1297-1308.
7. Yarbus A.L. Eye movements and vision. Plenum Press. New York, 1967.
8. Rybak I., Gusakova V., Golovan A., Podladchikova L., Shevtsova N. Attention-guided recognition based on "What"; and "Where" representations: A behavioral model. // Neurobiology of Attention, Elsevier, Academic Press. – 2005. - P. 663-670.
9. Gusakova V.I., Podladchikova L.N., Shaposhnikov D.G., Markin S.N., Golovan A.V., Lee S.W. () Advanced biologically plausible algorithms for low-level image processing. // In proc. of SPIE. – 1999. - Vol. 3837. - P. 377-386.
10. Petrushan M.V., Samarin A.I., Shaposhnikov D.G. FOSFI: A system for face image recognition. // Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications. -2005. -Vol. 15, No. 2. - P. 425-427.
11. Gao X., Podladchikova L., Shaposhnikov D. Application of Vision Models to Traffic Sign Recognition. // Lecture Notes in Computer Science (LNCS). - Springer-Verlag, 2003. – Vol. 2714. - P. 1100-1105.
12. Gao X., Shaposhnikov D.G., Anishenko S., Batty S. Human face determinants to detect head movements parameters during PET. // In proc. of All-Russian conf. «Neuroinformatic-2007». – 2007. – Vol.1. - P. 102-109.
13. Bloomfield P.M., Spinks T.J., Reed J., Schnoor L., Westrip A.M., Livieratos L., Fulton R., Jones T. The Development of a Motion Correction System in Neurological PET Using List Mode Data Acquisition and a Polaris Tracking System. // Neuroimage. – 2002. – No. 16. – P. 15.