DETECTION OF OBJECT SALIENT FEATURES BASED ON THE OBSERVER SCANPATH ANALYSIS

P. Augustyniak*, Z. Mikrut *

* Biocybernetic Laboratory, University of Science and Technology, Kraków, Poland

{august, zibi}@agh.edu.pl

Abstract: In this paper the applied methods and obtained results of scanpath analysis are presented. The volunteers' goal was to recognize handwritten digits gradually appearing from the background. Two parameters: the focus time and the eye returns count have been used to salient features detection. The paper presents the scanpath processing algorithm identifying the focus attention regions and detecting the episodes of returning attention. The quantitative comparison of scanpath parameters was particularly investigated in case of handwritten digits recognition for "1" and "7", very similar to each other.

Introduction

Partial models of human visual system are used to investigate many aspects covering the area from the receptors' topology on the retina [1] or the analysis of visual nerve data flow to the general studies of scene understanding, pattern recognition and cognitive identification [2]. The scanpath analysis [3][4] has currently a growing impact on the design of visual information media as well as on the applications of search strategies in the artificial intelligence-based visual systems.

This paper presents the method of scanpath analysis aimed at indentification of salient features of presented objects. The human gaze is recorded during a visual experiment consisting of presentation of frames sequence displaying an object gradually emerging from the background. At certain point the observer recognizes the object. The recorded eye trajectories processed in context of the image reveal observer's search strategy and scene conspicuity.

Materials and methods

Nine young volunteers performed a series of two visual experiments aiming at handwritten digits recognition. The visual tasks of such kind are often performed by the human. Therefore it is believed that the recognition is immediate, at a glance, so particular presentation method was designed to motivate the observers to perform the active search. One experiment used a pixelated images sequence of increasing resolution and the other used a texture background with gradually contrasting shape of a digit (see figure 1). The presentation time for each image varied from 5s for very blurry frames to 2s for clear ones.

The infra-red reflection-based eyetracker OBER2 [5] recorded two-dimensional eye positition independently at a rate of 750Hz with 12-bit precision. The computer image presentation and data acquisition were controlled by a custom-developed software supporting the static calibration of the eyetracker. The precise calibration is the crucial condition for accurate positionning of a scanpath layer over the background image. The calibration scanpath was also used for automatic determination of the dominant eye [6].

The search strategy was represented by a sequence of focus areas derived from the eyeglobe position by a custom-developed software. Two parameters of each focus point were investigated as representative for the salience of the scene region presented in the background: the focus time and the eye returns count.

The algorithm of focus point detection is explained hereafter using the left eye trace from observer 4 as an example (Figure 2). The algorithm consists of the following stages:

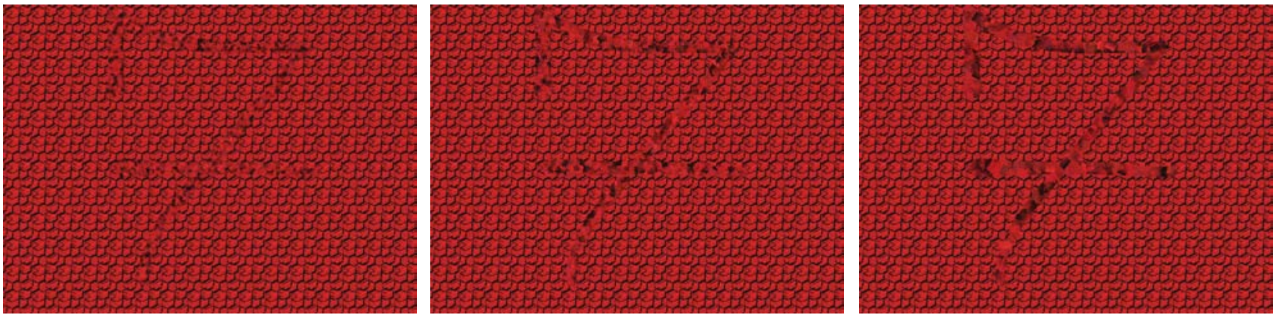

Figure 1. The sequence of three frames (digit "7") presented to the volunteers. Exposure periods were (from left to right): 4, 3 and 3 [s] . The digit is gradually appearing from the background

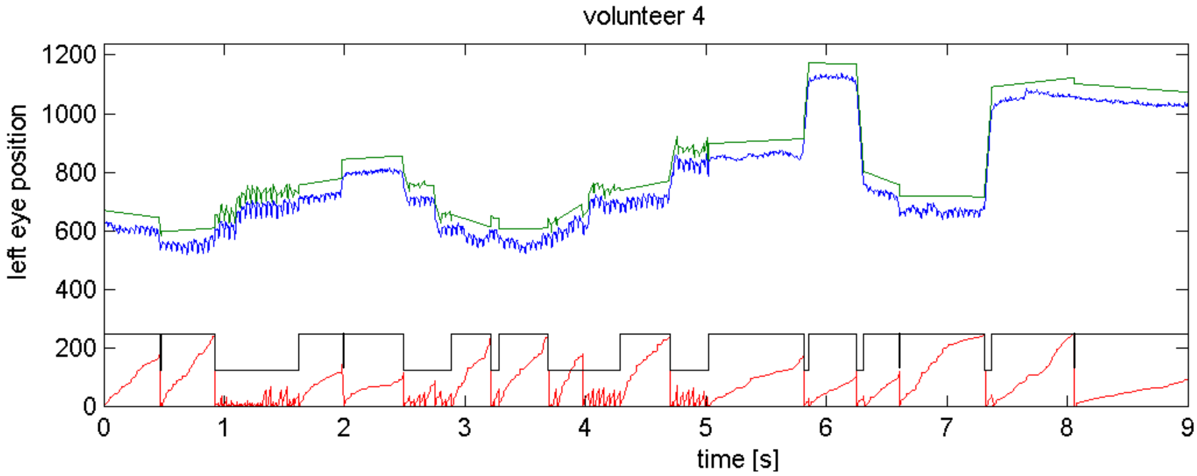

Figure 2. Analysis of the eye position data: source data after calibration (blue), line segment approximation (green – shifted up), approximation error (red), fixation areas detected (black – the higher level).

- subsequent sections of the recorded eye trace are approximated with the best fitted line; the approximation error is calculated for each line separately; the current approximation is discontinued and a new line is started if the approximation error exceeds a given threshold value or if the error difference for subsequent points exceeds the threshold value (see figure 2 red line). The latter case detects an abrupt change of the function slope representing fast eye movement,
- the fixation areas are detected as signal sections of the weak inclination (almost horizontal) of the length exceeding a given threshold value (see figure 2 black line). In the given example 13 fixation regions were detected and they are represented in figure 2 by upper positions of black line,
- the fixation times are computed and the longest fixation event is determined.

The second part of the algorithm provides the maximum number of returns that the eyeglobe performs to each given foveation region:

- each foveation region is converted to a separate image, similar to the image presented in figure 3,
- a triple morphological dilation transform is performed for each such image, for the reason the focus points for a given foveation may not be coherent,
- each foveation image is compared to the others with use of the logical function AND; the non-empty result of such operation indicates two foveations targeting a partly common region. Such cases are binary marked in a square matrix of a size equal to the total number of foveation points,
- the marks in the matrix are cumulated row-wise; the biggest value of the sum detects the fixation areas most attracting the observer returns,
- the fixation areas are put in a sequence by the returns number order; for the visualization clarity, the foveation regions are bordered by a rectangular frame (Figure 3).

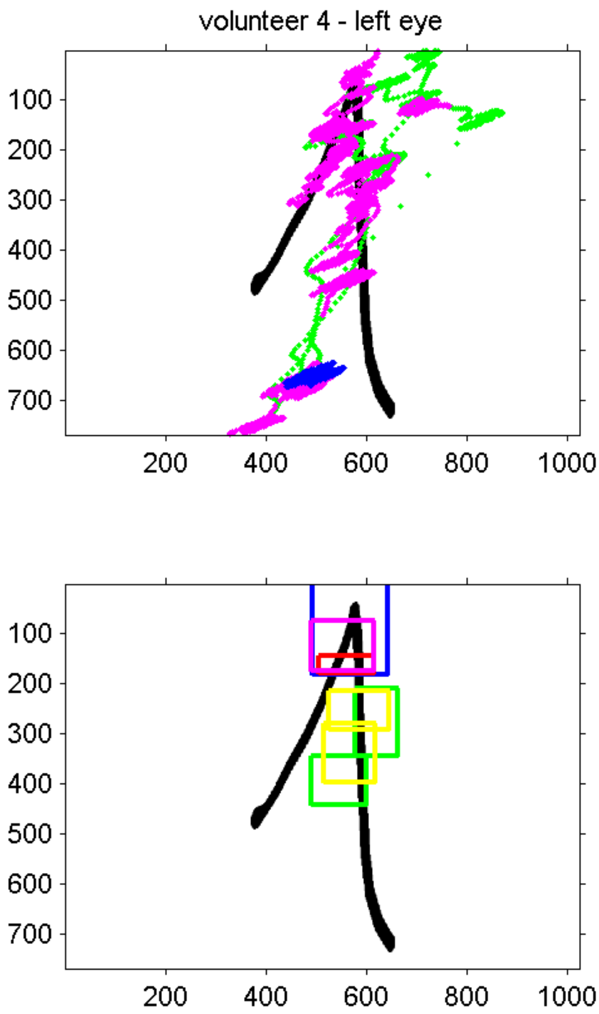

Figure 3: Upper: Original scanpath (green), fixation areas (magenta), maximal fixation (blue). Lower: two areas of maximal returns count

ISSN: 1727-1983 © 2005 IFMBE

Results

The results were processed for each experiment and for each of presented digit separately. As particularly interesting we found the perception comparison of very similar digits “1” and “7”.

The preliminary analysis of the fixation areas for these two digits led to determining of three regions most attracting the observers attention (Figure 4). Tables 1 and 2 summarize the results of the analysis.

The recognition of the digit „1” (Table 1) involves most of the observer’s attention in regions B (sloppy dash in left upper part of the sign) and A (central lower part of the sign).

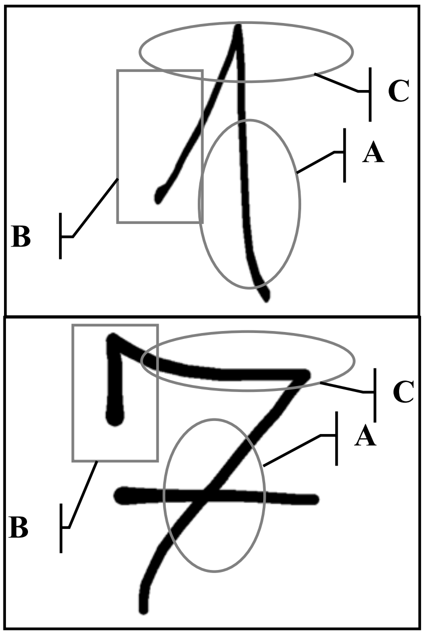

Figure 4. Two images of handwritten digits: the codes and locations of defined regions to be analysed

These results come from the detection of regions of the longest fixation time (for 3 observers A, for 5 observers B) and is also confirmed by the analysis of most frequently re-foveated regions (5 observers returned their attention to regions A and B). In 7 cases the longest fixation regions matched the most frequently re-foveated regions. Among these cases 3 observers preferred region A, while 3 others – region B.

For the recognition of the digit „7” the observers focused mainly on the central part of the digit (region A including a horizontal dash - see table 2). In seven observers (of the total of nine) the longest fixation region was A. In 9 cases (of the total of 11) region A was also the most frequently re-foveated region. Moreover, in all 7 cases region A was detected as the longest fixation area, it was also revealed as the most frequently re-foveated region. Region B (left upper part of the sign) was significantly less attractive for the observers.

Conclusions

The detection of salient features in the scene can be done properly using OBER2 eyetracker and the presented methodology of scanpath analysis. The parameters proposed in the paper: the focus time and the eye returns count indicate accurately the position of the most informative elements of the scene.

The analysis provides detection of the most significant features of selected handwritten digits. In case of digit “1” these features are: region B (sloppy dash in left upper part of the sign) and A (central lower part of the sign). In case of digit “7” these features were detected in the central part including a horizontal dash described as region A.

The next step is the identification and description of individual visual search strategy. We hope that the different strategies separated during our study will contribute to understanding of human image perception and therefore will assist the research on biologically inspired visual systems.

Acknowledgements

This work was supported by the University of Science and Technology under grant No.11.11.120.249. Authors wish to thank students: Bartlomiej Kluza and Adrian Korczynski for their help in carrying out the experiments.

Table 1: Results of experiments: recognition of handwritten digit “1”

No.	Recognition: image no.	– time [ms]	Dominant eye	Number of fixation areas	Criterion 1 Max. fixation time [ms] - zone				Criterion 2 Max. returns count - zones				Criterion 1 AND criterion 2			
1.	2	9000	right	17	661	A			3	A	B		1	A		
2.	3	12000	left	15	2877		B		5		B	C	1		B	
3.	2	9000	left	17	896	A			6	A			1	A		
4.	2	9000	left	13	940	A			2	A		C	1	A		
5.	1	5000	right	10	1080		B		3		B	C	1		B	
6.	2	9000	left	21	801			C	3		B	C	1			C
7.	2	9000	left	13	949		B		5		B		1		B	
8.	2	9000	left	12	568		B		4	A			0			
9.	3	12000	left	17	971		B		5	A			0			
				areas – total:		3	5	1		5	5	4		3	3	1

ISSN: 1727-1983

Table 2: Results of experiments: recognition of handwritten digit “7”

No.	**Recognition: image no. – time [ms]**		**Dominant eye**	**Number of fixation areas**	**Criterion 1: Max. fixation time [ms] - zone**			**Criterion 2: Max. returns count - zones**				**Criterion 1 AND criterion 2**	
1.	3	12000	left	21	784	A		5	A			1	A
2.	2	9000	left	12	1543	A		9	A			1	A
3.	2	9000	left	14	856	A		6	A			1	A
4.	2	9000	right	14	1049		B	8	A			0	
5.	1	5000	right	11	1123	A		2	A		C	1	A
6.	2	9000	left	19	805	A		3	A			1	A
7.	2	9000	left	13	756		B	5	A			0	
8.	2	9000	left	14	469	A		2	A			1	A
9.	2	12000	left	13	728	A		2	A	B		1	A
				areas – total:		7	2		9	1	1		7

References

[1] SCHWARTZ E.L. (1977): ‘Spatial mapping in the primate sensory projection: analytic structure and relevance to perception’, *Biological Cybernetics,* **25**, pp. 181-194

[2] TADEUSIEWICZ R., OGIELA M. R. (2004): ‘Medical Image Understanding Technology’, Springer, Berlin

[3] AUGUSTYNIAK P., MIKRUT Z. (2002): ‘Correlating the degree of observer's preoccupation and the observation time: visual tasks with OBER2 eyetracker’, *J. Medical Informatics & Technologies,* vol. 4, pp. MT3-MT10

[4] MIKRUT Z., AUGUSTYNIAK P. (2002): ‘Estimation of execution times for tasks of objects counting and localization using the OBER2 device’, Proc. 12th Nordic Baltic Conf. on Biomedical Engineering and Medical Physics, Reykjavik, pp. 144-145

[5] OBER J., HAJDA J., LOSKA J., JAMICKI M. (1997): ‘Application of eye movement measuring system OBER2 to medicine and technology’, *Proc. of SPIE, Infrared technology and applications,* , vol. 3061, part 1, pp. 327-333, Orlando, USA

[6] MIKRUT Z., AUGUSTYNIAK P. (2002): ‘Dominant eye recognition based on calibration of the OBER2 eyetracker’, *Proc. IFMBE* Vol. 3, pp. 394-395.

ISSN: 1727-1983 © 2005 IFMBE