8th. World Congress on Computational Mechanics (WCCM8)
5th. European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2008)
June 30 – July 5, 2008
Venice, Italy

USING WEIGHTED SALIENCY MAP IN VISUAL ATTENTION FOR OBJECT RECOGNITION

***H. Bahmani[1], A. M. Nasrabadi[2] and M. R. Hashemi Gholpayeghani[3]**

[1] Biomedical Engineering Department, Science and Research University
hbahmani@gmail.com

[2] Biomedical Engineering Department, Shahed University
nasrabadi@shahed.ac.ir

[3] Biomedical Engineering Department, Amirkabir University of Technology
hashemi_g@morva.net

Key Words: *visual attention, saliency map, data fusion, object recognition*

ABSTRACT

Primates use saliency-based visual attention to detect conspicuous objects in cluttered visual environments. The neuronal representation of the visual world is enhanced within the restricted area of focus of attention (FOA), or spotlight. This bottom-up pre-processing would allow the higher levels of the cortical hierarchy to perform a semantic visual processing in a top-down manner.

Several models have been proposed to functionally account for visual attention. All these models share similar general architecture. Multi-scale topographic “feature maps” detect local spatial discontinuities in intensity, color and orientation as basic early visual features. These maps from different visual modalities with different dynamic ranges are then combined into a final “saliency map”, containing spotlights.

Different strategies of combining feature maps to form a saliency map are addressed in this paper. Traditional methods of making saliency map in visual attention are to combine feature maps in a bottom-up manner. This consists of normalizing these maps by a global nonlinear function and then superimpose to each other. These methods however, can neither involve the top-down role of brain in biasing attention to a specific target, nor omitting distracters intelligently from the scene to be revealed.

Here, we proposed some modifications in the basic visual attention model to maximally mimic the biological behaviour of vision. For the application of object recognition, in order to detect a specific conspicuous target, the model should consider the learned weight of each feature map to form the saliency map for vision system. In this regard, we combined a weighted saliency map, with former normalizing methods for better performance in object recognition. Weight of each feature map is proportional to its influence on the final saliency map to make the desired target more conspicuous.

The comparing results suggest that the best candidate model to successfully find target over distracters in a 32 scenes datasets of emergency triangles, is a combination of weighted saliency map, followed by a global nonlinear normalization function. Since we are using visual attention for content-based object recognition, the number of successfully detecting the target at the first hit has been also considered in this evaluation, as well as average of false detections over the scenes. As a heuristic

experiment, a multiplicative product of feature maps has revealed surprising results, which would be a good starting point of using nonlinear data fusion strategies as our future work.

Table 1. Average number of false detection before target found for different combination strategies.

Method	Mean	Std. deviation	No. of zeros*
Naive (linear) normalization	2.1563	3.2240	12
Global nonlinear normalization	2.1875	2.3201	9
Weighting feature maps	1.2188	2.5493	23
Weighted maps with nonlinear normalization	1.1250	2.1515	22
Multiplicative saliency map	0.5172	1.0563	21

* zero counts stand for precisely detecting the target at first hit

Figure 1. visual attention scanpaths of different combination strategies, and FOA circles indicating number of false detections before the last circle overlap the target . From left to right: Naive, Nonlinear, Weighting, and Combined method. Top: Image containing emergency triangle to be detected, bottom: relevant saliency map.

REFERENCES

[1] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis", *IEEE Trans PAMI* 20(11), pp. 1254-1259, 1998.

[2] D. Walter, L. Itti, M. Riesenhuber, T. Poggio, and C. Koch, "Attentional selection for object recognition – a gentle way", *Lecture Notes in Computer Science*, Vol. 2525, pp. 472-479, 2002.

[3] L. Itti, and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention", *Vision Research* 40(10-12), pp. 1489-1506, 2000.

[4] L. Itti, and C. Koch, "A comparison of feature combination for saliency-based visual attention systems", *SPIE human vision and electronic imaging IV (HVEI'99)*, San Jose, CA, pp. 473–482, 1999.

[5] L. Itti, and C. Koch, "Feature combination strategies for saliency-based visual attention systems", *Journal of Electronic Imaging*, 10(1), pp. 161–169, 2001.