Detecting Object of Interest from a Noisy Image Using Human Visual Attention

Kyung-Joo Cheoi*
School of Electrical & Computer Engineering
Chungbuk National University, Cheongju, Korea

ABSTRACT

This paper describes a new mechanism of detecting object of interest from a noisy image, without using any a-priori knowledge about the target. It employs a parallel set of filters inspired upon biological findings of mammalian vision. In our proposed system, several basic features are extracted directly from original input visual stimuli, and these features are integrated based on their local competitive relations and statistical information. Through integration process, unnecessary features for detecting the target are spontaneously decreased, while useful features are enhanced. Experiments have been performed on a set of computer generated and real images corrupted with noise.

***Keywords**: Target, Mammalian Vision, Noisy Image*

1. INTRODUCTION

In the most of the classical methods of image analysis, the high-level features in the whole image are computed first, and then each of the local set of features is compared with the collection of stored prototypes defining known objects in a serial manner. Those methods are extensively slow and too application specific [5]. By the way, biological system appear to employ a serial strategy by which an attentional spotlight rapidly selects circumscribed regions in the scene, rather than attempting to fully interpret visual scenes in a parallel manner, for further analysis [1]. This function of visual attention enables the Primates to interpret natural scenes in real time, despite of the limited speed of the biological architecture available for such tasks.

Although many effective and fast computer vision systems for target detection exist, they were finely tuned to specific targets, and thus typically fail in detecting other types of objects. This makes the system hardly be extended to other applications. Also, most of them fail when severe noise exists in the image. Because of these reasons, we now propose a target('salient object') detection system which can be used for general purpose, and which also successfully works even in noisy images. The proposed system is based on bottom-up processing of above mentioned visual attention mechanism.

In engineering, visual attention mechanism can be very efficiently applied to the problems of background/foreground separation, object recognition, and etc., but is not utilized enough yet. A fundamental problem to be solved in particular is that of how to focus the attention of a system. What features are important for focusing the attention?

A number of paradigms describing human focus of attention have been developed over the years by researchers in psychology [2,3,6,7,13]. Among them, most literature agrees that the attention selection mechanism consists of two functionally hierarchical stages, pre-attentive and attentive stage. In early pre-attentive stage, all visual stimuli in the entire visual field are processed in parallel without capacity limitation and in limited capacity attentive stage, only one item or at best a few items are processed at a time. One of the most popular models on visual attention is the one proposed by Treisman. He suggested a feature integration theory [13], that views the perception of objects on the basis of above two-stage metaphor as a process. It says that the basic 'features' like edges, orientation, width, size, color, brightness, etc. were detected in the pre-attentive stage, and these basic features are 'integrated' in the attentive stage in order to be perceived as objects in the world.

Some biologically plausible systems for target detection in computer vision have been proposed [8,9,11,12,14]. As the systems in [12] and [14] were applied only to synthetic images or other simple images containing characters, it hard to be extended to applications which use complex natural color images as a input. And in [11], even if they tried to evaluate the system with real images, they presented rare experimental results of natural color images, still less noisy images. Also, the performance of the system in [8] and [9] needs to be analyzed with more noisy images.

Our system proposed here, finds the target in noisy image only by the properties of input visual stimuli. Very simple feature maps such as color and intensity contrast are extracted directly from visual stimuli, based on the color-opponency filter of the human visual system and the gray-level information

* *Corresponding author. E-mail: kjcheoi@chungbuk.ac.kr*
Manuscript received Sep. 15, 2005 ; accepted Mar. 13, 2006

of the input visual stimuli. These independent feature maps are then processed by oriented ON-center and OFF-surround operator in order to extract orientations from each feature maps and also to enhance the regions of pixels largely differing from their surround. Thereafter, these computed features are integrated based on their local competitive relations and statistical information. Through integration process, unnecessary features for detecting the target are spontaneously decreased while useful features are added and enhanced. And the system selects the most different features among other features as the target. We have tested our system with variety of noisy images, and the performance of the system proved very robust to noise irrespective of its properties : its distribution form or its color value. Also, by using only bottom-up cue of input image, it can be used for general purpose without major changes of the architecture.

In Section 2, the proposed methodology is explained. Experimental results are shown in Section 3, and concluding remarks are made in Section 4.

2. THE METHODOLOGY

In our system, the raw image is initially decomposed into four independent feature maps. Each map represents the value of a certain attribute computed on a set of low-level features. Here, two achromatic feature maps and two chromatic feature maps are generated. As two achromatic feature maps, F^1 and F^2 are generated by ON and OFF intensity information of input image, respectively. As In the nature, as well as in human made environments, color can distinguish objects and signal(e.g., traffic signs or warning signs in industrial environments). So, two color feature maps, F^3 and F^4 are generated. F^3 and F^4 are modeled with the two types of color opponency exhibited by the cells with homogeneous type of receptive fields in visual cortex, which respond very strong to color contrast. F^3 is generated to account for 'red/green' color opponency and F^4 for 'blue/yellow' color opponency. These four generated feature maps are then normalized in the range of 0~1 in order to eliminate the differences due to dissimilar feature extraction mechanisms, and to simplify further processing of the feature maps.

At the second stage of the processing, each of the early-vision feature maps, F^k (k=1~4), is convolved with the oriented ON-center, OFF-surround operators($\theta \in \{0, \pi/8, 2\pi/8, \cdots, 7\pi/8\}$) by

$$I^k_{x,y} = \sum_{\theta}\left(\sum_{m,n} F^k_{m,n} \cdot T\right)^2$$
$$T = \left|K_1 \cdot G_{x-m,y-n}(\sigma, r_1 \cdot \sigma, \theta) - K_2 \cdot G_{x-m,y-n}(r_2 \cdot \sigma, r_1 \cdot r_2 \cdot \sigma, \theta)\right| \quad (1)$$

where K_1, K_2 denote positive constants, r_1, denote the eccentricities of the two Gaussians, r_2, denote the ratio between the widths of the ON and OFF Gaussians, and $G_{x,y}(\cdot,\cdot,\cdot)$, 2-D oriented Gaussian functions. And the resulting maps are squared to enhance the contrast, and multiple oriented measures(θ) are integrated to provide a single measure of interest of each location. Through this process, reorganized importance map I^k extracts additional feature, orientation, and enhances the regions of pixels whose values are largely different from their surroundings'.

Although many features that influence visual attention have been identified, little quantitative data exist regarding the exact weighting of the different features and their relationship. Some features clearly have very high importance, but it is difficult to define exactly that how much the one feature is more important than another. A particular feature may be more important than another in one image, while the opposite may be true in another image. We integrated four importance maps that provide various measures of interest for each location into one single saliency map in order to have global measure of interest by

$$S^k_{x,y} = \frac{SI^k_{x,y} - MinSI}{MaxSI - MinSI} \quad (2)$$

where

$$SI^k_{x,y} = I^k_{x,y} \times \left(MaxI^k - AveI^k\right)^2,$$
$$MaxI^k = \max(I^k_{x,y}),$$
$$AveI^k = \frac{1}{N-1}\left[\left(\sum_{x,y} I^k_{x,y}\right) - MaxI^k\right] \quad (3)$$
$$MaxSI = \max\left(SI^1_{x,y}, SI^2_{x,y}, SI^3_{x,y}, SI^4_{x,y}\right),$$
$$MinSI = \min\left(SI^1_{x,y}, SI^2_{x,y}, SI^3_{x,y}, SI^4_{x,y}\right)$$

In Ex. 3, N denotes the size of I^k map. In making SI^k map, we used the main idea proposed by Itti and Koch [9]. This processing enhances the values associated with strong peak activities in the map while suppressing uniform peak activities by the statistical information of the pixels in the map. Also, comparing the map with other maps enables to retain relative importance of a feature map with respect to other ones. And irrelevant information extracted from ineffective feature map would be singled out. So, we just sum four S^k maps to make saliency map.

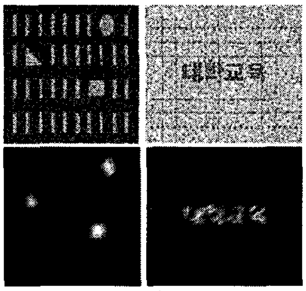
3. EXPERIMENTAL RESULTS

We tested our system with various images range from simple synthetic images to complex real images of natural environment. Each of the which has a severe noise (40%~100%), and the noise has some properties : it has gaussian or uniform distribution, and has color values or only intensity values.

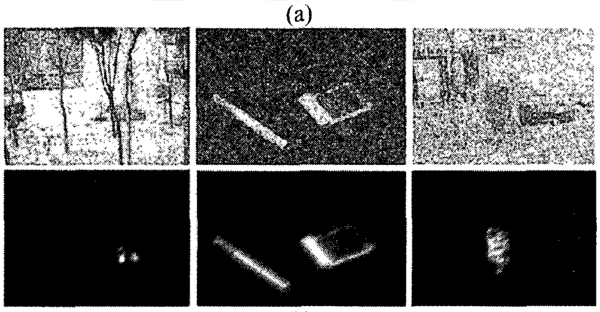
As synthetic images, we used various images of which the 'targets' are differed in orientations, in colors, in sizes, and in intensity contrast, with a set of 'distractors'. Also the system was tested with the images of which the background is lighter than the targets, the reverse images. The system detected the targets immediately, and some results about these tasks were shown in Fig. 1(a). As real images, each containing the target object such as signboard, signal lamp, traffic sign, mailbox, placard, etc., as well as the distractors such as strong local variations in illumination, textures. Fig. 1(b) ~ Fig. 2 shows some results of real images obtained through our system.

4. CONCLUSIONS

We reported in this paper a new mechanism of detecting object of interest from a noisy image, without using any a-priori knowledge about the target. The method proposed herein, uses only bottom-up cues of input visual stimuli, and does not requires any a-priori knowledge. The proposed system is composed of two main stages, feature extraction and integration stage. Several basic features are extracted directly from visual stimuli first, and these extracted features are integrated based on their local competitive relations and statistical information. We have studied the overall behaviors of the model with input from two large classes of noisy images. The first are visual scenes constructed analogously to the stimuli typically presented in psychophysical studies of visual search. The second class of input is color images of natural environment taken from different domains. As shown in experimental results, the performance of the system was good, and it shows the promise that it could be successfully used as a target detector in complex real images for general purpose.

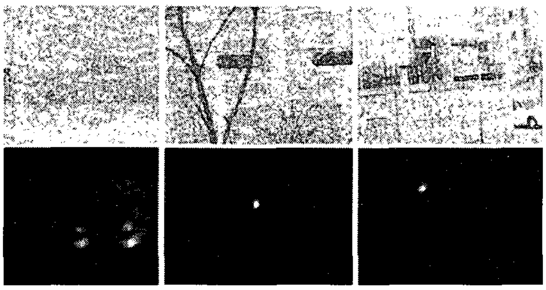


(a)

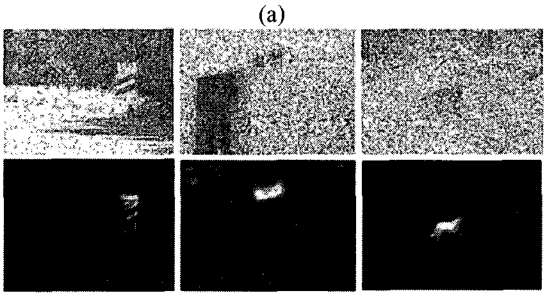


(b)

Fig. 1. Some results of noisy images which has a color noise: (a) *gaussian distribution* : (left) 50% noise : detecting three different shaped objects, (right) 70% noise : detecting blue characters. (b) *uniform distribution* : (left) 30% noise : detecting red basket, (center) 100% noise : detecting cigarette and the light, (right) 90% noise : detecting red mail box



(a)



(b)

Fig. 2. Some results of noisy images which has an intensity noise : (a) 40% noise with *gaussian distribution* : the target is (left) red emergency triangle, (center) green light, (right) red light. (b) 60% noise with *uniform distribution* : the target is (left) yellow traffic sign, (center) green emergency lamp, (right) yellow dog

REFERENCES

[1] M. Bear, B. Connors, and M. Paradiso, **Neuroscience exploring the brain**, Williams and Wilkins, USA, 1996.

[2] J. Duncan and J. Humphreys, "Visual search and stimulus similarity", **Psychological Reviews 96**, 1989, pp.433-458.

[3] K. Cave and J. Wolfe, "Modeling the Role of Parallel Processing in Visual Search", **Cognitive Psychology**, Vol.22, 1990, pp.225-271.

[4] D. Chapman, **Vision, Instruction, and Action**, Ph.D Thesis, Massachusetts Institute of Technology, 1990.

[5] O. Gallat, P. Gaussier, and J. Cocquerez, "A Model of the Visual Attention to Speed up Image Analysis", **Proc. of Intl. Conf. on Image Processing**, Vol.1, 1998, pp.246-250.

[6] J. Jonides, "Further toward a model of the mind's eye's movement", **Bulletin of the Psychonomic Society**, Vol.21, No.4, 1983, pp.247-450.

[7] B. Julesz and J. Bergen, "Textons, the fundamental Elements in Preattentive Vision and Perception of Textures", **Bell Systems Technical Jornal**, Vol.62, 1983, pp.1619-1643.

[8] L.Itti, C.Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention", **Vision Research,** Vol.40, No.10, 2000, pp.1489-1506.

[9] L.Itti, C. Gold, C.Koch, Itti, L., "Visual Attention and Target Detection in Cluttered Natural Scenes", **Optical Engineering**, Vol.40, No.9, 2001, pp.1784-1793.

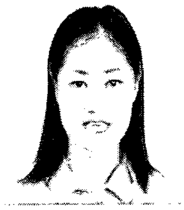
[10] C.Koch, S.Ullman, "Shifts in Selective Visual Attention : Towards the Underlying Neural Circuitry", **Human Neurobiology**, Vol.4, 1985, pp.219-227.

[11] R.Milanese, H.Wechsler, S.Gil, J.Bost, T.Pun, "Integration of Bottom-up and Top-down Cues for Visual Attention Using Non-Linear Relaxation", **Proc. of IEEE Conf. on Computer Vision and Pattern Recognition,** 1994, pp.781-785.

[12] M.Mozer, **The Perception of Multiple Objects : a Connectionist Approach** , MIT Press, Cambridge, MA, 1991.

[13] A.Treisman, G. Gelade, "A Feature-integration theory of Attention", **Cognitive Psychology**, Vol.12, No.1, 1980, pp.97-136.

[14] P.Sandon, "Simulating Visual Attention", **Journal of Cognitive NeuroScience**, Vol.2, No.3, 1990, pp.213-231.



Kyung-Joo Cheoi

She received the B.S. in computer science from Chungbuk National University, Korea in 1996 and also received M.S., Ph.D. in Computer Science from Yonsei University, Korea in 1998, 2002 respectively. From 2002 to 2005 she was with LG CNS R&D center, where she researched web service computing, CMS, Biometrics. Currently, she is a full-time lecturer in school of Electrical&Computer Engineering at Chungbuk National University. Her main research interests include computer vision, bio computing, and ubiquitous computing.