The Fifteenth International Symposium on Artificial Life and Robotics 2010 (AROB 15th '10),
B-Con Plaza, Beppu,Oita, Japan, February 4-6, 2010

Visual attention model involving feature-based inhibition of return

Shinji Hotta, Shigeyuki Oba, and Shin Ishii

Graduate School of infomatics, Kyoto University, Gokasho, Uji, Kyoto 619-0011, Japan
(Tel : 0774-38-3938; Fax : 0774-38-3941)
(hotta-s@sys.i.kyoto-u.ac.jp)

Abstract: Visual attention tends to avoid locations where past visual attention has once focused. This phenomenon is called inhibition of return (IOR) and is known as one of important dynamic properties of visual attention. Recently, several studies have reported that IOR occurs not only on locations but also on visual features. In this study, we propose a visual attention model that involves the feature-based IOR by extending a recent model of `Saliency Map'. Our model is demonstrated by a computer simulation and its neuronal basis is also discussed.

Keywords: visual attention, feature-based IOR, saliency map

I. INTRODUCTION

Visual attention puts focus on a part of a large amount of visual information coming into human brain simultaneously so as to process the essential information intensively, because precise processing of whole information is impossible in a limited time.

Bottom-up visual saliency is a well-known factor that influences attentional control. Visual stimuli that stand out from their surroundings are said salient. The more salient a stimulus is, the more easily it deploys visual attention. Itti *et al*. formerly proposed a computational procedure that constracts saliency map in a way incorporating biological findings in the early visual system [1]. Although this procedure could reproduce locations of instantaneous bottom-up attention, dynamic properties of visual attention in the level of saliency maps were not considered sufficiently.

Inhibition of return (IOR) is an important dynamic property of visual attention [2]. Directions of visual attention tend to avoid locations on which past visual attention has once focused and eye movement does not return to the inhibited areas in a short while. This property allows us to find out desired things lying among many other salient objects in an efficient manner. Recently, E. Shin *et al*. reported that IOR occurs not only on locations but also on visual features [3]. Its mechanism, however, is not understood well.

In order to understand the mechanism underlying the feature-based IOR, in this study, we propose a new model of visual attention, which combines the saliency map and the feature-based IOR. A computer simulation was performed to know the basic character of our new model. Also its neural basis is discussed.

II. VISUAL ATTENTION MODEL

1. Saliency map

In the saliency map[1], visual saliency level of each spatial location in a current input image is calculated by a linear summation of multiple topographic feature maps which are obtained from the input image. Detail processes are as follows.

First, static color image is transformed to multi-scale images in a form of dyadic Gaussian Pyramids. The Gaussian Pyramid $P(\sigma)$, $\sigma=1,\ldots,8$, is a pile of natural images created by low-pass filtering and subsampling of the original input image P, where σ indicates the scale; P(0) is the original image and P(8) is an image reduced into 1/256 size.

Second, seven feature maps are extracted from the Gaussian Pyramid $P(\sigma)$; they are intensity map I, two color opponency maps R/G, B/Y, and four orientation maps with angles 0˚, 45˚, 90˚, 135˚.

Third, the difference between center and surround scale maps is computed for each of the seven feature maps. The difference between two different scales of a feature map highlights salient areas with respect to the feature, because those areas have different feature values from their surrounding areas; such a difference map is called a Center-Surround map.

Fourth, the Center-Surround difference maps are integrated over different scales and over different features in each modality. In each integration, summation of normalized maps of different scales and features over the modality is calculated, where the normalization operator *N* emphasizes such a map that involves single or small number of peak salient areas and degrades such a map that involves many peaks. As a result, three conspicuity maps $N(I)$, $N(C)$, and $N(O)$, corresponding to three modalities, intensity, color opponency, and orientations, respectively, are obtained.

Last, the three conspicuity maps are again normalized and linearly summed into the consequent saliency map,

$$S = \frac{1}{3}(N(I) + N(C) + N(O)).$$

2. Inhibition of return

In the procedure above, we obtain the saliency map $S_j(t)$ based on scale-integrated feature maps $F_{ij}(t)$ at time t=1,2,⋯, where i=1,…,N and j=1,…,M are indices of feature types and spatial locations in each map, respectively.

The location of attention $j^*(t)$ is determined such to maximize the current saliency $S_j(t)$ in the simplest model. In this simple model, however, the attention cannot be directed to second and third peaks on the same saliency map. Thus, the idea of IOR is proposed.

Using the IOR, we assume that location of visual attention is determined by the modulated saliency

$$MS_j(t) = IOR_j(t) S_j(t),$$

where $IOR_j(t)$ belongs to [0, 1] and denotes the strength of feature-based IOR at location j at time t, and the location of attention $j^*(t)$ is determined to maximize the modulated saliency $MS_j(t)$. The effect of IOR is determined based on the history of locations of focused attention, so that the dynamics of $IOR_j(t)$ is formulated as

$$\Delta M_j(t) = K_j(t) + \eta M_j(t),$$

where $M_j(t)=1-IOR_j(t)$, $\Delta M_j(t)= M_j(t+1)-M_j(t)$, and the constant η is a decaying coefficient. This expression means that the IOR effect is smoothed over the current and past instantaneous IOR. $K_j(t)$ is a map of inhibition at time t; location j is inhibited if $K_j(t)$=1 and not inhibited if $K_j(t)$=0.

We compare location-based IOR [1] and a newly proposed feature-based IOR. The difference between them is reflected in the map $K_j(t)$. In the location-based IOR [1], locations near $j^*(t)$ are inhibited after visiting $j^*(t)$ so that visual attention will not return to the area

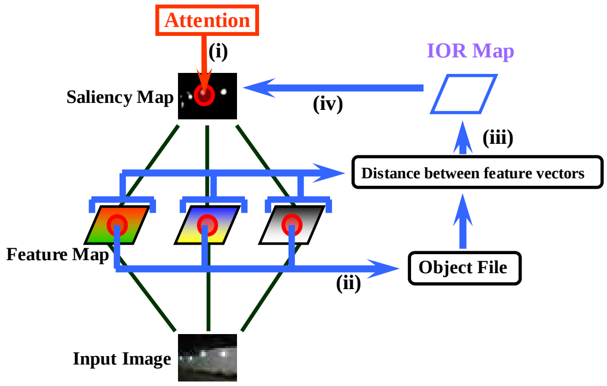

Fig.1. Proposed model.

(i)The location of focused attention is determined by the modulated saliency map. (ii)The feature vector of that location is taken into the object file module. (iii)The IOR map is created based on the Euclidian distance in the feature space. (iv)The modulated saliency map at the next time is calculated by the convolution of the saliency map and the IOR map. Attention is carried out according to the modulated saliency map.

around $j^*(t)$ in a short while. Namely, we set $K_j(t)$ =1 for all j such that

$$\| j^*(t) - j \| < \gamma$$

holds, where γ is a given constant and $\|\cdot\|$ is the Euclidean distance. In the proposed feature-based IOR, on the other hand, $K_j(t)$ is defined by means of the distance in the N-dimentional feature space so that $K_j(t)$=1 for all j such that

$$\| \mathbf{F}_{j^*(t)} - \mathbf{F}_j \| < \lambda$$

holds and $K_j(t)$=0 otherwise, where λ is a given constant, $\boldsymbol{F}_j = \{F_{1j}, \cdots, F_{Nj}\}$ is a feature vector at location j, and $\boldsymbol{F}_{j^*(t)}$ is a feature vector at the attended location $j^*(t)$.

III. RESULTS

We demonstrated the behaviors of the proposed feature-based IOR on several natural images, and two of them are shown in Fig. 2. In the left example image (A), there are five objects; four objects, D_1, D_2, D_3, D_4, are street lights which are significantly salient because of the high intensity, and have similar feature vectors to each other, while the target T_A is a fire hydrant whose saliency is high, but lower in total than those of the other four objects. In the right example image (B), there is a can, target T_B, within grasses. Some areas of the

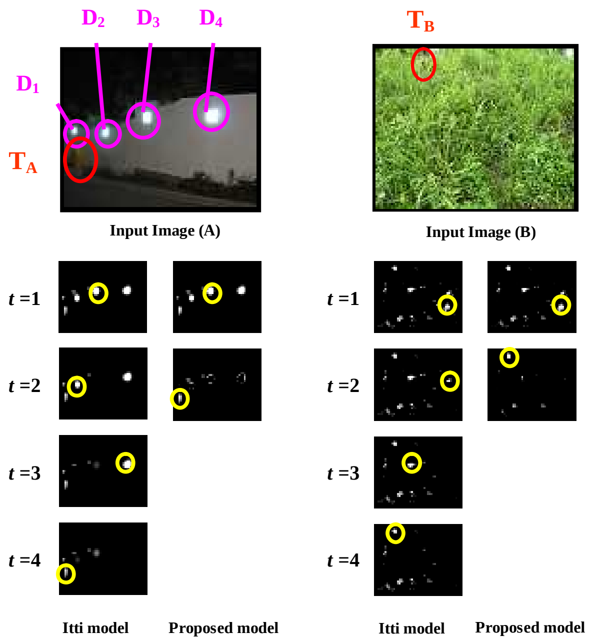

Fig. 2. Simulation results.
Two examples of natural images are shown in the top panels. In the panels below, the saliency maps with IOR convolved and the locations of focused attention (circles) at time *t*=1, 2, 3, 4 are shown. (Itti model) Itti *et al.*'s location-based IOR. (Proposed model) Our feature-based IOR.

grasses have high saliency because of the high intensity of green channel. The can has a different feature from grasses and its saliency is somewhat high, but lower than those of some grasses.

By the previous model based on location-based IOR, the attention focus tended to move between obstacles, the street lights in (A) or the grasses in (B), and to take long time to discover the target, T_A or T_B, respectively. By the proposed model based on feature-based IOR, on the other hand, attention rapidly moved onto the target. We also confirmed the efficient search ability based on the proposed feature-based IOR, through simulations using other natural images.

IV. DISCUSSION

We showed through computer simulations that the modulated saliency map with feature-based IOR leads to efficient search ever in cluttered situations. Thus we naturally think that the feature-based IOR is crucial for efficient visual search in human behaviors. In this section, we discuss neural mechanism underlying feature-based IOR.

In our model, feature-based inhibitory signals are assumed to be involved after calculating saliency map. This assumption is based on a biological finding by Shin *et al.*'s work on feature-based IOR [3]. They investigated modulation process in attentional control to generate distractor previewing effect (DPE) using an electrophysiological method. Repeating presentations of distractors composed of target-irrelevant features increases search time in subsequent visual search tasks, which is called DPE [4]. Shin *et al.* listed following four processing levels to generate the DPE; (a) preattentive perceptual processes, (b) preset attentional biases, (c) the ability to shift attention, and, (d) the weights to activate responses. They observed the event-related potential (ERP) corresponding to each processing level and achieved an evidence that DPE can directly affect (c) *the ability to shift attention toward the target*. This meant that IOR occurs on visual features. In the current study, we adopted their hypothesis in the proposed model by regarding processing level (a) as adaptation in feature maps, level (b) as weighting each feature map, and level (c) as determination of the location to be attended.

On the other hand, almost nothing has been known about neuronal basis of inhibitory signals applied on saliency map. In what follows, we discuss a possible mechanism that generates such signals in the neural system. The proposed model of modulated saliency map requires the following processes.

(I). Determine the attended location by the saliency map.
(II). Extract the feature vector in the attended location.
(III). Identify the areas whose feature vectors are similar to that of the attended location.
(IV). Provide inhibitional signals to the saliency map.

The saliency map (I) has been recognized as a plausible model of neural mechanism of visual attention control. In the visual pathway, especially early part, V1 [5], V4 [6] and so on, have been considered as neural bases of feature maps, because the retinotopic structure and simple visual features, such as color opponency and orientation, have been found in these areas. Treisman and Gerade had proposed the psychological notion `master map' which operates visual attention [7]. Koch *et al.* extended it to the saliency map in the computational viewpoint [8]. Recently, neural activities corresponding to the saliency are reported in V1 [9], the posterior parietal cortex [10], the frontal eye field [11], the superior colliculus [12], and so on.

For process (II), we assumed the `object file' module in the early visual system which memorizes the activities of feature neurons representing the feature vector at the attended location. `Feature Integration Theory' hypothesized that each feature information at the attended location is integrated into the object file, and transported to higher modules [7]. This theory may provide a mechanism to our idea.

Process (III) needs a `detector of synchronized firing neurons' module in the early visual system which detects population of firing neurons that are syncronized with those in the `object file module' indirectly. Based on such a synchronization mechanism, the detector specifies location j to be inhibited by matching feature vectors $\boldsymbol{F}_{j^*(t)}$ and $\boldsymbol{F}_j$. Since our feature-based IOR is represented as a weight value (Eq.(2)), it could be implemented by a probabilistic read-out, depending on $\|\boldsymbol{F}_{j^*(t)} - \boldsymbol{F}_j\|$.

In order for the processes (I), (II), and (III) to be linked to the process (IV), we assume five particular neural connctions between the modules; from `feature maps' to `object file', from `feature maps' and `object file' to `detector of synchronized firing neurons module', from `detector of synchronized firing neurons module' to `IOR signal generator module', and from `IOR signal generator module' to `the modulated saliency map'. In specific, the connection from V1 to the superior colliculus was found anatomically [13]. This connection suggests that the `detector of synchronized firing neurons module' in V1 can influence the IOR signals in the superior colliculus. Although the huge amount of anatomical data have suggested that the visual system is abundant in feedforward and feedback connections between various modules, further details of the implementation of the feature-based IOR is unclear.

V. Summary

We proposed a new model of visual attention involving the feature-based IOR which allows multiple salient objects to be inhibited when they have similar features. Simulation results showed that saliency of visually similar stimuli degenerated by the effect of the feature-based IOR, so that an efficient search for a target object was realized. We also discussed possible neural bases for the feature-based IOR.

REFERENCES

[1] Itti L, Koch C, Neibur E (1998), A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20, pp.1254-1259

[2] Posner MI, Cohen Y (1984), Components of Visual Orientating. Attention and Performance, 10, pp.531-556

[3] Shin E, Wan XI, Fabiani M, Gratton E, *et al.* (2008), Electrophysiological evidence of feature-based inhibition of focused attention across consecutive trials. Psychophysiology, 45, pp.804-811

[4] Goolsby BA, Grabowecky M, Suzuki S (2005), Adaptive modulation of color salience contingent upon global form coding and task relevance. Vision Research, 45, pp.901-930

[5] Knierim JJ, Van Essen DC (1992), Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. Journal of Neurophysiology, 67, pp.961-980

[6] Schein SJ, Desimone R (1990), Spectral properties of V4 neurons in the macaque. The Journal of Neuroscience, 10, pp.3369-3389

[7] Treisman AM, Gerade G (1980), A Feature-Integration Theory of Attention. Cognitive Psychology, 12, pp.97-136

[8] Koch C, Ullman S (1985) Shifts in selective visual attention: towards the underlying neural circuitr. Human Neurobiology, 4, pp.219-227

[9] Lamme VA (1995), The neurophysiology of figure-ground segregation in primary visual cortex. The Journal of Neuroscience, 15, pp.1605-1615

[10] Gottlieb JP, Kusunoki M, Goldberg ME (1998), The representation of visual salience in the monkey parietal cortex. Nature, 391, pp.481-484

[11] Schall JD, Hanes DP, Thompsons KG, *et al.* (1995), Saccade target selection in frontal eye field of macaque: I. Visual and premovement activation. The Journal of Neuroscience, 15, pp.6905-6918

[12] McPeek RM, Keller EL (2002), Saccade target selection in the superior colliculus during a visual search task. Journal of Neurophysiology, 88, pp.2019-2034

[13] Lund JS, Lund RD, Hendrickson AE, *et al.* (1975), The origin of efferent pathways from the primary visual cortex, area 17, of the macaque monkey as shown by retrograde transport of horseradish peroxidase. The Journal of comparative neurology, 164, pp.287-304