International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.7 (2016), pp.205-218
http://dx.doi.org/10.14257/ijmue.2016.11.7.21

Efficient Target Object Localization Model Based on Biologically Motivated Visual Selective Attention

Jaeho Oh, Chang-Beom Kwon and Sang-Woo Ban*

Department of Information and Communication Engineering, Dongguk University,
123 Dongdae-ro, Gyeongju, Gyeongbuk 780-714, Republic of Korea
**Corresponding author: swban@dongguk.ac.kr*

Abstract

This work proposes a novel biologically motivated visual selective attention model for efficient visual searching, which is implemented by integrating three attention mechanisms: bottom-up attention, top-down attention, and spatial attention. Bottom-up attention generates salient locations by reflecting top-down biases as well as three primitive visual features: intensity, edge and color. Prototype-based object perception is proposed for top-down attention, in which a 3-D color histogram is applied to generate a prototype of the target object. And experience based spatial attention can determine acceleration to localize a target object, which is modeled by using memorized spatial location information updated by object-localization experience. In order to verify the performance of the proposed visual selective attention model, we apply the proposed model to a real application in pedestrian traffic signal detection, to be utilized as part of a blind guide system. The proposed selective attention model shows plausible performance in terms of accuracy and computation time while efficiently localizing pedestrian traffic signals.

Keywords*: Visual selective attention, Saliency map, Bottom-up attention, Top-down attention, Spatial attention, The blind guide system*

1. Introduction

The human vision system efficiently conducts searches of complex visual scenes, in which visual selective attention plays an important role, without fully searching the current visual field [1-7]. Spatial cues predicting the probable location of a target are commonly used as an operational manipulation of covert visual attention [2]. Human's visual selective attention is a very complex process where many factors, such as visual saliency, goals, intentions, affection, experience, *etc.*, are involved directly and/or indirectly [8-18]. However, it is obvious that the human brain utilizes not only visual features of an input scene but also memorized spatial information obtained from previous experience with efficient visual search. A strong relationship between visual working memory and selective attention has been revealed, in which attention is biased by what is currently on our mind [8].

Many visual selective attention models have been introduced for efficient processing of complex visual scenes [9-17]. A big challenge is the degree to which a visual attention model agrees with biological findings [1]. In the context of attention, biologically inspired models have resulted in higher accuracies in some cases [1]. In Desimone and Duncan's model, the biased competition view of a visual search proposes two general sources for the control of attention, in which bottom-up sources arise from sensory stimuli present in a scene, and top-down sources arise from current behavioral goals [9]. Itti *et al*. proposed a brain-like model to generate a saliency map (SM) [10], which has been considered a representative engineering model of biologically motivated visual selective attention.

ISSN: 1975-0080 IJMUE

Navalpakkam and Itti proposed a top-down attention model that has a biasing mechanism for a salient map based on the signal-to-noise ratio of color and orientation feature generated by a bottom-up process [11]. Park *et al*. introduced a bottom-up saliency map model considering mutual information minimization mechanism [12]. Walther and Koch also proposed a top-down attention model having a bias based on features generated from a bottom-up process [13]. Ban *et al*. proposed top-down attention reflecting a human affective factor based on a psychological distance mechanism [14]. Carmi and Itti proposed an attention model that considered seven dynamic features in MTV-style video clips [15]. Kim *et al*.'s top-down attention model reflected independent feature-based object perception and incremental knowledge generation [16]. All these proposed attention models have not considered spatial cue information for generating selective attention. Torralba *et al*. proposed a top-down attention model using spatial information [17], which may show poor performance for objects located in an unusual place even though it considered spatial information and it does not consider bottom-up attention.

Most of the previous visual attention modeling research has been focused on the bottom-up component of visual attention [1]. However the field of visual attention still lacks computational principles for task-driven attention [1]. Thus, this work proposes a novel integrated visual selective attention model that can efficiently localize a target object by considering experience based spatial attention as well as top-down biased bottom-up attention and object perception based top-down attention altogether.

The proposed visual selective attention model is developed by understanding and mimicking of the human brain's mechanism for visual selective attention. The proposed model consists of four parts. Bottom-up attention generates a saliency map by integrating primitive visual features, such as edge, intensity and color opponency, which are affected by top-down bias generated from target object perception. Thus, this bottom-up attention efficiently generates candidate areas for a target object. Top-down attention localizes target- object areas using a prototype-based pattern matching process. For target object perception, a 3-D color histogram is applied to represent a target object and Euclidean distance is utilized to measure similarity when deciding on the target object area [18]. Also, memory-based spatial attention plays a role by utilizing spatial cue relevant to target object location obtained from object localization experience. Memorized spatial attention can enhance the computation time for a visual search. Thus, the proposed model can enhance the performance of the visual search process and speed up processing time. These three attention components interactively work to generate the final selective attention, which is done by an integration and control aspect of attention. The proposed visual selective attention model shows plausible selective attention with high efficiency in terms of accuracy and computation time for target localization.

This paper is organized as follows. Section 2 describes the proposed visual selective attention model, integrating bottom-up saliency, top-down object perception and spatial attention for an efficient visual search. The experimental results and conclusions follow in Section 3 and Section 4, respectively.

2. Proposed Visual Selective Attention Model

Figure 1 shows the brain areas and their functions related to visual selective attention as considered in the proposed model. The retina cells can extract edge and intensity information as well as color opponency as primitive visual features. The relativity of these extracted primitive visual features is extracted by on-center and off-surround mechanism of the lateral geniculate nucleus (LGN). These processed features go to the visual cortex (V1, V2, V4, IT) for further complex processing [4-9]. In general, the secondary visual areas play a role in form and color perception of an object, plus 3-D position and motion perception. The infero-temporal (IT) area located in a secondary visual area contains complex shape coding information, and generates corresponding activity according to

object form information [4-9]. The neurons in area V4 respond best to specific colors of objects, irrespective of lighting conditions [4-9]. Therefore, it is natural to assume that the IT and V4 areas play an important role in the detection and recognition of objects based on the pre-processed visual information in the primary visual cortex. It is well known that the hippocampus mainly works on memory formation about perceived objects. As well, the hippocampus is part of a network that develops associations between representations in different areas [8]. Many parietal areas, including the lateral intra-parietal (LIP), medial temporal (MT), intra-parietal superior (IPS), as well as temporal parietal junction (TPJ) contribute to integration and control of attention factors generated in many different areas, which are also very closely related to motor area for action generation [8].

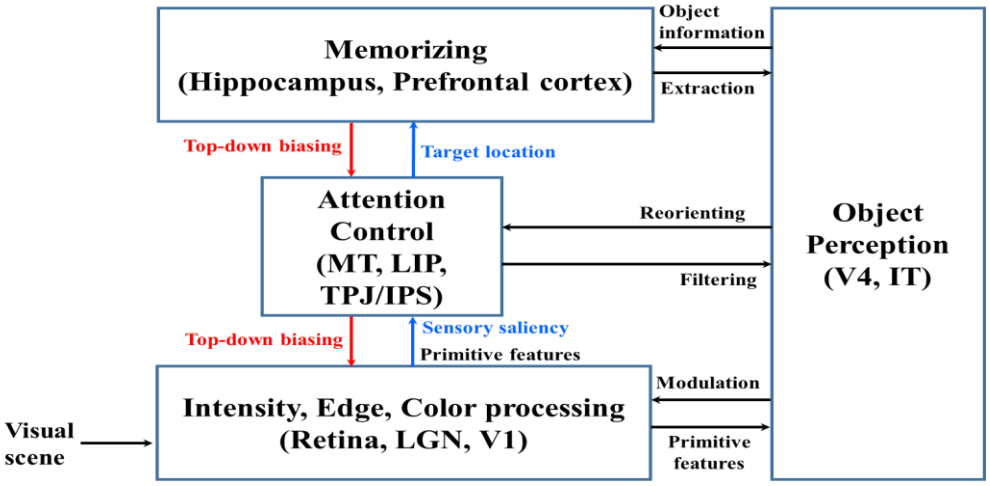


Figure 1. Visual Selective Attention Control Process of the Human Brain

Based on understanding biological mechanisms of the human brain, we propose a novel visual selective attention model, shown in Figure 2, that mimics the role of each area of the brain related to visual selective attention, as shown in Figure 1. In Figure 2, bottom-up processing mimics the roles of the retina, LGN and V1 areas in Figure 1. Saliency information based on relative primitive visual features plays a role in bottom-up attention generation. And the object perception region mimics the roles of the V4/IT area. Object perception results are utilized for top-down attention generation by reorienting the attention area. The process for memorizing spatial target location mimics the roles of the hippocampus and prefrontal cortex, which memorize characteristics and locations of objects. As well, the attention integration and control process mimics the roles of the MT, LIP, IPS areas and TPJ, which integrate every feature related to attention generation interacting with other attention factors to generate the final attention in a sequence. In Figure 2, R, G and B represent three color components of red, green and blue. And r, g, b and y denotes four real color components calculated using R, G and B preferred in [11, 13]. As well, directed lines in Figure 2 show feed-forward paths for visual information processing in order to generate visual selective attention. Instead, directed dotted lines in Figure 2 present feedback paths for top-down biasing while generating visual selective attention.

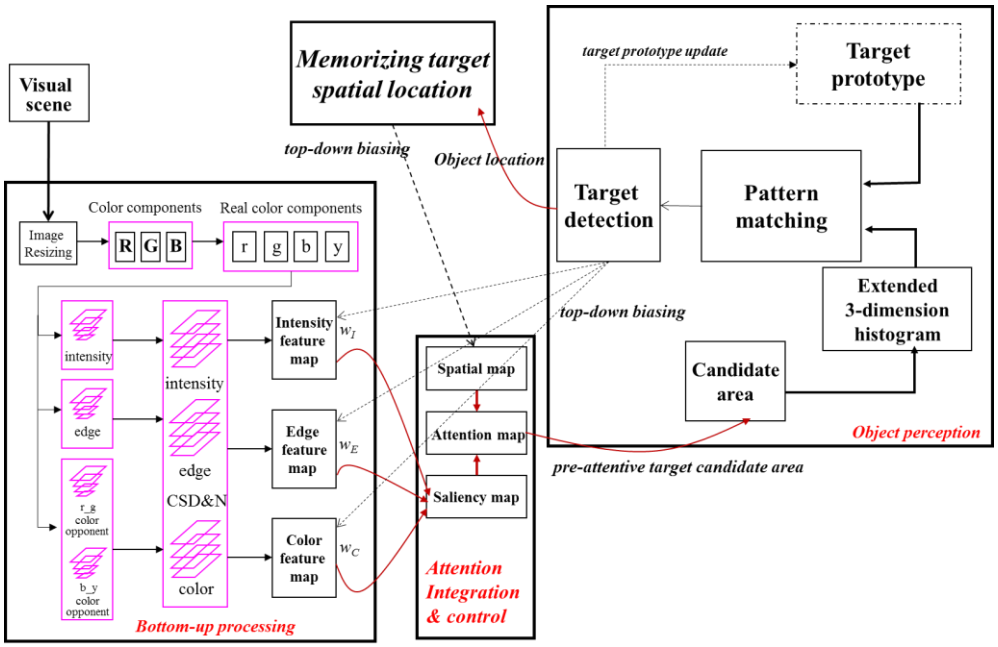


Figure 2. Proposed Biologically Motivated Visual Selective Attention Model

2.1. Bottom-Up Attention with Top-Down Bias

In order to implement a human-like visual attention function, we consider the modification of a simplified bottom-up SM model [10, 12]. In our approach, we use the SM model that reflects the functions of the retina cells, the LGN and the visual cortex. Since the retina cells can extract edge and intensity information as well as color opponency, we use these factors as the basic features of the SM model [10, 12].

The function of the LGN and ganglion cell is implemented by the on-center and off-surround operation via Gaussian pyramid images with different scales from 0 to the *n*-th level, whereby each level is made by the sub-sampling of 2^n. Thus the operation is able to construct four basic features, such as intensity (*I*), edge (*E*), and color (*RG* and *BY*) [10, 12]. This reflects the non-uniform distribution of the retina-topic structure. Then, the center-surround mechanism is implemented in the model as a difference operation between the fine and coarse scales of the Gaussian pyramid images [10, 12]. Consequently, three feature maps, such as intensity feature map ($\bar{I}$), edge feature map ($\bar{E}$) and color feature map ($\bar{C}$), can be obtained by the center-surround difference algorithm [10, 12]. For a detailed description on obtaining the three feature maps from an input image, refer to Itti *et al*. [10] and Par *et al*. [12].

Even in a bottom-up process, we can consider top-down bias since it has been revealed that modulation effects in the sensory pathway occur at all cortical levels and even in the thalamus [8]. Top-down attention derived by object detection can be utilized as top-down modulation information in order to make more relevant features of the localized object be considered more dominant than others. Top-down bias serves to enhance primitive feature based bottom-up attention by reflecting coincidence of top-down attention and features of the corresponding attention area. Therefore, if some primitive features at the location of attention coincidentally dominate when attention occurs at the location, then those features can be considered as important in forming attention. This mechanism can be modeled by increasing bias for the corresponding dominant feature, which might be implemented by a Hebbian learning mechanism. A top-down bias is calculated by Eq. (1) based on a Hebbian learning rule, where each bias for a corresponding feature map has a larger value if the corresponding feature map has larger values when target object

detection occurs. In Eq. (1) x_i and x_j are target object detection occurrence and dominant contributive feature activity, respectively, where x_i is 0 or 1, x_j is the average amplitude of a *j*-feature map at a target-object area and γ is the learning rate. Therefore a more dominant contributive feature can have larger bias in order to obtain more plausible feature-based bottom-up attention. In the proposed model, the weight value of each feature map is updated by Eq. (2), where β is a scaling factor. Accordingly, the weight cannot change if $bias_j(t)$ is zero, which means that there is no co-occurrence of target object localization. And a bottom-up SM is generated by the summation of these three feature maps multiplied by corresponding weights via Eq. (3). Each weight combination (w_I, w_E, w_C) for the corresponding three feature maps plays a role in bias reflecting the characteristics of a target object.

$$bias_j(t) = \gamma \cdot x_j \cdot x_i \quad \text{, where } j = I, E, \text{ or } C \tag{1}$$

$$w_j(t+1) = w_j(t) + \frac{1 - e^{-\beta \cdot bias_j(t)}}{1 + e^{-\beta \cdot bias_j(t)}} \quad \text{, where } j = I, E, \text{ or } C \tag{2}$$

$$SM = w_I \cdot \bar{I} + w_E \cdot \bar{E} + w_C \cdot \bar{C} \tag{3}$$

2.2. Prototype-Based Object Perception for Top-Down Attention

In the proposed model, we applied a 3-D color histogram prototype method to represent a target object, and Euclidean distance to measure similarity when deciding on a target-object area. An object perception model can vary according to the target application for more efficient target object dependent localization. The target application of the proposed model is pedestrian traffic light signals and sign board detection for pedestrians who are visually impaired, in which spatially distributive color features are dominant for object localization. This target application implicitly affected selection of the object perception model in this work. In order to generate a 3-D color histogram as a prototype of an object, the 3-D RGB space is divided into 3-D small voxels and a histogram is generated by concatenating the number of pixels in each voxel [18]. Each dimension is divided into five levels at the value range of each dimension, and then 125 (5x5x5) voxels are generated and a histogram of the object is generated by concatenating the number of pixels obtained from each voxel. Accordingly, a 3-D color histogram can plausibly represent color characteristics of an object. As well, the proposed model introduces a modified histogram method in order to reflect spatial characteristics of the color features of a target object. The object area can be divided into sub-areas, and each histogram can be obtained from each sub-area. Then, with Eq. (4), each histogram is concatenated to generate an extended 3-D histogram for the object, which is then normalized for scale-invariant representation. By considering the extended 3-D histogram approach, we can overcome a weak point in histogram-based approaches that cannot reflect spatial distribution of features. Moreover, a prototype representation, $hist_{prototype}$, for an object is also generated by Eq. (4) and trained by updating its histogram reflecting the characteristics of a newly detected object, as shown in Eq. (5), where η is the training rate.

$$hist_{object} = \lfloor hist_{subarea\_1};\ \cdots\cdots\ ; hist_{subarea\_n} \rfloor \tag{4}$$

$$hist_{prototype}(t+1) = \eta \cdot hist_{prototype}(t) + (1-\eta) \cdot hist_{localized\_object} \tag{5}$$

2.3. Spatial Attention

Humans can utilize spatial location information, when trying to understand a visual scene, in order to efficiently process a complex visual scene. In a specific case, spatial location of an object is not so varied, but almost static, in the visual field. In such a case, spatial location information is very important for efficient visual target localization. That spatial location information can be obtained from repeatable attention experiences during localization of an object. By considering already experienced location information of an object, humans can generate a kind of virtual map to represent spatial information. Such a spatial map might be generated by accumulating the frequency of an object's appearance in each area of a visual field. An area more frequently localized can have higher attractiveness than less frequent ones. Thus a spatial map can be defined by a function of target object occurrence and previously trained spatial information for target object localization as shown in Eq. (6). A spatial map value for each location is calculated with Eqs. (7) to (9), where spatial map values of neighborhood locations centered at the target location are only increased. In Eq. (8), σ is determined by the size of the neighborhood of a target location. This experience-based memorized spatial location information enhances object search time by reducing the candidate areas for localized target objects.

$$spatial\_map(x,y)_{t+1} = f(spatial\_map(x,y)_t, target\_occurrence(x,y)_t) \tag{6}$$

$$spatial\_map(x',y')_{t+1} = spatial\_map(x',y')_t + \Delta c(x',y')_t \tag{7}$$

$$\Delta c(x',y')_t = target\_occurrence(x,y)_t \cdot e^{\frac{d^2((x',y'),(x,y))}{\sigma^2}} \quad where\ (x',y')\ is\ a\ neighbor\ location\ centered\ at\ (x,y) \tag{8}$$

$$d((x',y'),(x,y)) = \sqrt{(x'-x)^2 + (y'-y)^2} \tag{9}$$

2.4. Integrated Selective Attention for Efficient Target Object Localization

The human visual processing system can plausibly utilize three different mechanisms to efficiently localize a target object: primitive feature-based bottom-up attention, top-down object related biasing, and spatial attention. In the proposed model, these three different functions are properly integrated in order to provide more efficient target object localization. Candidates for object detection are obtained by both bottom-up saliency and spatial attention. Spatial attention increasingly plays an important role in efficient target localization. Finally, target object localization is achieved by applying an object perception process in the localized candidate areas. Therefore candidate-area localization can be expressed by a function, $g(\cdot)$, of a bottom-up saliency map and a spatial attention map, as expressed in Eq. (10). A location having a high value in both an SM and a spatial map will become the highest priority candidate for target object localization. As well, the final attention can be described by a function, $h(\cdot)$, of localized candidate areas and an object perception process as expressed in Eq. (11). If a target object is localized at a

selected candidate location by the object perception process, the selected candidate location of the attention map has a high value. The result of $h(\cdot)$ generates an attention map, which is utilized for deciding the final selective attention.

$$candidates(x,y)_t = g(SM(x,y)_t,\ spatial\_map(x,y)_t) \tag{10}$$

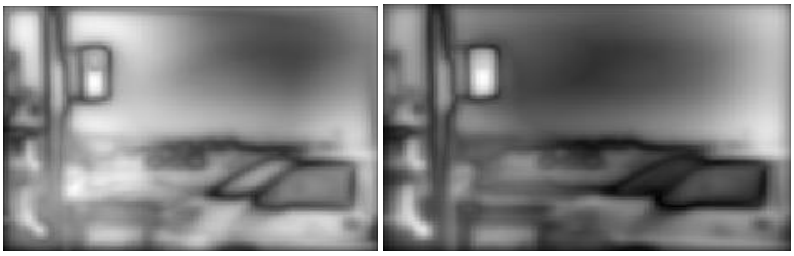
$$attention\_map(x,y)_t = h(candidates(x,y)_t, object\_perception(x,y)_t) \tag{11}$$

3. Experimental Results

To verify the proposed visual selective attention model, we applied the proposed model to localization of pedestrian traffic signals since that is very important to the visually impaired. The proposed model is supposed to be utilized as a part of a blind guide system, as well. In order to set up the simulation, we obtained a road image database (DB) including pedestrian traffic signals captured during the day from 1 p.m. to 4 p.m., which is the target time of the guide system. In that DB, 55 images show red traffic signals and 44 images show green traffic signals. Figure 3 and Figure 4 shows experimental results of the bottom-up SM generation part. Figures 3 (a) to (c) are three features maps ($\bar{I}$, $\bar{E}$, $\bar{C}$) obtained from an input image. Figures 4 (a) and (b), relatively, show an SM obtained by integration of non top-down biased feature maps and an SM from integration of top-down biased feature maps. As shown in Figure 4 (b), the top-down biased SM shows better performance in generating candidates of the target object, since the traffic signal area becomes more salient. But a non-traffic signal area becomes less salient, compared with those areas of non top-down biased SM shown in Figure 4 (a). Each area in a selective candidate area is represented by an extended 3-D color histogram. In this experiment, each area was divided into two sub-areas to reflect spatial characteristics of traffic signals, since the upper part and the lower part of a traffic signal typically have different features. Therefore, each area is represented by a 250-dimension histogram vector, with 125 dimensions from the upper part and another 125 dimensions from the lower part of each area, since each dimension of a three dimensional RGB space is divided into five levels. Thus each histogram is generated by concatenation of the number of pixels in each of the 125 voxels.

(a) Intensity Feature Map (b) Edge Feature Map (c) Color Feature Map

Figure 3. Three Bottom-Up Feature Maps



(a) Non Top-Down Biased SM (b) Top-Down Biased SM

Figure 4. Comparison between Non Top-down Biased Saliency Map and Top-down Biased Saliency Map

Figure 5 shows example experimental results from localizing red pedestrian traffic signals and green ones, in which yellow boxes are the candidate areas of the traffic signals localized by using a spatial attention map and a bottom-up saliency map. An extended 3-D color histogram obtained from each candidate is compared with the prototype extended 3-D color histogram obtained from the training target objects. If the similarity between two compared histograms is high, the selected candidate is classified as a target object area. As well, red and green boxes in Figure 5 are the finally localized pedestrian traffic signal areas via the proposed selective attention model. The proposed model shows successful localization of pedestrian traffic signals, even though there are many distracters having characteristics similar to traffic signals in the visual scenes.

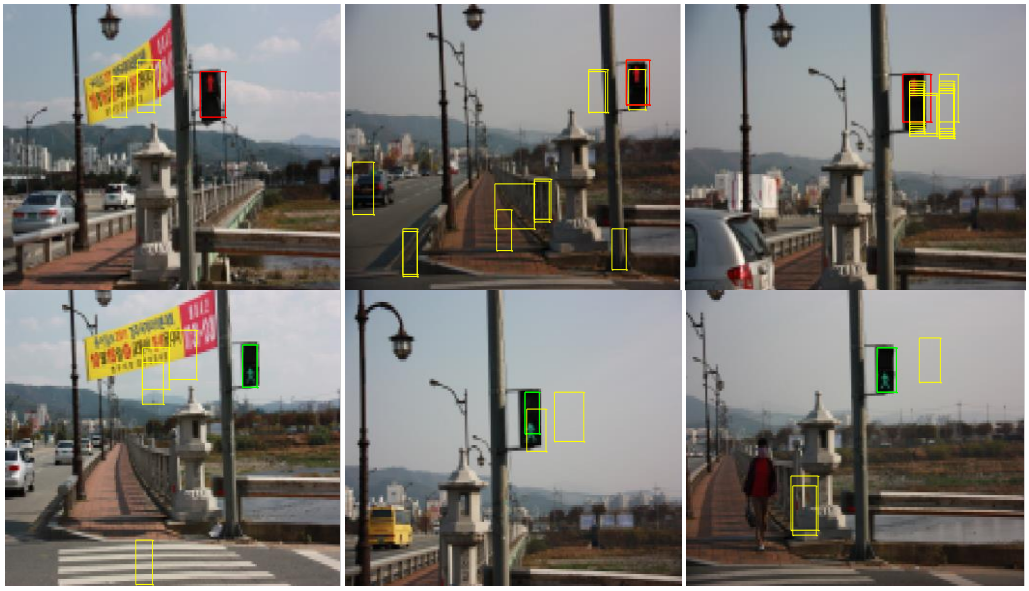


Figure 5. Examples of Pedestrian Traffic Signal Detection

Table 1 shows that the proposed visual selective attention model generates better performance for both accurate recognition of pedestrian traffic signals and computation time. We compared the performance of the proposed model against a model without a bottom-up SM process, and against a model with only a bottom-up SM process without a spatial attention process.

Table 1. Performance of Three Visual Selective Attention Models for Pedestrian Traffic Signal Detection

Correct target localization performance (# of images)				
Using the same top-down object perception method for each model		Model without bottom-up attention	Model with top-down biased bottom-up attention without spatial attention	(Proposed model) Model with both top-down biased bottom-up attention & spatial attention
True Positive	Red Signals	98.18 % (54)	100 % (55)	100 % (55)
	Green Signals	100 % (44)	100 % (44)	100 % (44)
True Negative	Red Signals	1.82 % (1)	0 % (0)	0 % (0)
	Green Signals	0 % (0)	0 % (0)	0 % (0)
False Positive	Red Signals	0 % (0)	0 % (0)	0 % (0)
	Green Signals	0 % (0)	0 % (0)	0 % (0)
Average computa-tion time	SM generation time	-	t ~ 0.02sec	t ~ 0.02 sec
	Object perception time	0.273960 sec	0.147582 sec	0.086263 sec
	Post-processing time	0.003808 sec	0.003932 sec	0.004978 sec
	Total target localization time	0.277768 sec	0.151513 + t sec	0.091141 + t sec

Three different models utilized the same prototype-based object perception model. The proposed model shows 100% accurate localization of pedestrian traffic signals and greater enhancement of computation time by 0.06 seconds per image. In Table 1, the saliency generation time t is about 0.02 seconds. In order to compare computation time under the same conditions, saliency generation time t is denoted separately. In the proposed model, the computation time can be reduced by considering a weighting mechanism for generating an SM, since the candidate areas were decided using SM. As shown in Figure 4 (b), the SM weighted by top-down bias can inhibit non-target areas, which not only reduced candidate areas but also enhanced accurate localization of target objects by removing distracters by much more. Although the three different models use the same prototype based top-down object perception method, they show different computation times for object perception (0.274 sec, 0.148 sec, and 0.086 sec), since computation time for object perception depends on the amount of the candidate area to be perceived for target localization. This result suggests that the proposed model reduces candidate areas more than the other two models. Post-processing includes a decision on the final attention area based on comparison of the degree of similarity for each candidate area. As well, by considering a spatial attention process, the proposed model also enhances computation time much better as shown by the total target localization time in Table1, which means that the spatial attention process properly contributes to reducing the candidate areas.

Moreover, we applied the proposed model for traffic light signal detection to a video taken with a smart phone camera when a pedestrian crossed a road during the day from 1 p.m. to 4 p.m. The proposed model successfully localized the traffic light signals under various situations with complex backgrounds when many people were crossing the road and when many cars were on the road. The proposed model successfully localized the traffic light signals with good localized performance, except for occluded situations. Even though the proposed model failed to localize a traffic signal owing to occlusion by other pedestrians, the proposed model showed plausible performance for guiding a pedestrian. Figure 6 shows example frames for green pedestrian traffic light detection during crossing of the road and for red traffic light detection while a person stands in a crosswalk waiting area. The proposed model considering spatial attention took about .12 seconds per frame on average, and the attention model without spatial attention took about 0.23seconds per frame on average. Thus, computation time was enhanced by about 50%, from which we can conclude that spatial attention plays an important role in reducing computation time.

(a) Green Signal Detection (b) Red Signal Detection

Figure 6. Pedestrian Traffic Signal Detection during Crossing a Road

We also applied the proposed model to localize two pedestrian sign boards on the pavement. By considering pedestrian sign board detection, the proposed model can guide the blind walking on the pavement until they arrive at the crosswalk. After successfully arriving at the crosswalk, the proposed traffic light localization model can guide people crossing the road. The proposed model shows plausible performance in pedestrian sign board detection, even though only two sign boards were considered. Figure 7 shows examples of pedestrian sign board detection. The proposed model enhanced computation time for localizing the pedestrian sign boards as shown in Table 2. Pedestrian sign board detection is important for assisting the visually impaired while walking on the pavement. In this work, even though we only considered two pedestrian sign boards, they are important sign boards. One indicates crossing caution, and the other is indicating crossing guidance. Fifty crossing caution sign boards and 50 crossing guide sign boards were utilized for this performance evaluation. The proposed model showed localizing time enhancement by 33%, on average, for computation time, with the same correct localization performance. Even though the object perception process of the proposed model was not optimized, the proposed model shows good performance for properly guiding the blind while walking on the pavement.

(a) Pedestrian Caution Sign Board (b) Pedestrian Guide Sign Board

Figure 7. Examples of Pedestrian Sign Board Detection

Table 2. Performance of Proposed Visual Selective Attention Model for Pedestrian Sign Board Detection

Correct target localization performance (# of images)			
Using the same top-down object perception method for each model		Model with top-down biased bottom-up attention without spatial attention	(Proposed model) Model with both top-down biased bottom-up attention & spatial attention
True Positive	Caution sign boards	100 % (50)	100 % (50)
	Guide sign boards	100 % (50)	100 % (50)
True Negative	Caution sign boards	0 % (0)	0 % (0)
	Guide sign boards	0 % (0)	0 % (0)
False Positive	Caution sign boards	0 % (0)	0 % (0)
	Guide sign boards	0 % (0)	0 % (0)
Average computation time	SM generation time	0.193458 sec	0.180287 sec
	Object perception time	1.736394 sec	1.113093 sec
	Post-processing time	0.001041 sec	0.000965 sec
	Total target localization time	1.930893 sec	1.294345 sec

4. Conclusion and Future Works

A novel, biologically motivated, visual selective attention model for efficient visual searching is presented in this paper, which is aimed to be utilized as part of a blind guide system. The proposed visual selective attention model can efficiently localize a target

object by considering experience-based spatial attention as well as top-down biased bottom-up attention and object perception based top-down attention altogether. By considering experience-based spatial attention together with bottom-up saliency attention, the proposed model efficiently localizes candidate areas for target object localization, which provides performance enhancement. As well, the proposed model applied an extended 3-D color histogram as an object feature for object perception, which reflects both statistical property of color features and spatial distribution of color feature of the target object. Accordingly, the proposed model shows efficient target object localization performance in terms of both enhanced computation time and accuracy in target detection.

For further work, additional experiments with various image DBs should be considered in order to make the proposed model a more general one to efficiently localize general target objects. As well, a more general object perception model with a deep learning approach based on a convolution neural network (CNN) is also considering for providing a more biologically plausible model.

Acknowledgments

This research was supported by the Converging Research Center Program funded by the Ministry of Education, Science and Technology (2013034988). The authors appreciate Dong-Oh Kim(a former student involved in this research project) for allowing to use the previous version of the developed computer program.

References

[1] A. Borji and L. Itti, "State of the art in visual attention modeling", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 35, no. 1, **(2013)**, pp. 185-207.
[2] M. P. Eckstein, S. C. Mack, D. B. Liston, L. Bogush, R. Menzel and R. J. Krauzlis, "Rethinking human visual attention: Spatial cueing effects and optimality of decisions by honeybees, monkeys and humans", Vision Research, vol. 85, **(2013)**, pp. 5-19.
[3] T. Z. Luo and J. H. R. Maunsell, "Neuronal modulations in visual cortex are associated with only one of multiple components of attention", Neuron, vol. 86, **(2015)**, pp. 1182-1186.
[4] E. B. Goldstein, "Sensation and perception", 4th ed., an international Thomson publishing company, USA, **(1996)**.
[5] A. M. Treisman, "Features and objects in visual processing", Scientific American, vol. 255, **(1986)**, pp. 114B-125B.
[6] A. M. Treisman and G. Gelde, "A feature–integration theory of attention", Cognitive Psychology, vol. 12, **(1980)**, pp. 97-136.
[7] J. A. Mazer and J. L. Gallant, "Goal –related activity in V4 during free viewing visual search; evidence for a ventral stream visual saliency map", Neuron, vol. 40, **(2003)**, pp. 1241-1250.
[8] C. N. Olivers, J. Peters, R. Houtkamp and P. R. Roelfsema, "Different states in visual working memory: When it guides attention and when it does not", Trends in Cognitive Sciences, vol. 15, **(2011)**, pp. 327–334.
[9] R. Desimone and J. Duncan, "Neural mechanisms of selective visual attention", Annual Review of Neuroscience, vol. 18, **(1995)**, pp. 193-222.
[10] L. Itti, C. Koch and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis", IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 20, no. 11, **(1998)**, pp. 1254-11259.
[11] V. Navalpakkam and L. Itti, "An integrated model of top-down and bottom-up attention for optimal object detection", Proceedings of Computer Vision and Pattern Recognition, **(2006)**, pp. 2049-2056.
[12] S. J. Park, K. H. Ahn and M. Lee, "Saliency map model with adaptive masking based on independent component analysis", Neurocomputing, vol. 49, **(2002)**, pp. 417-422.
[13] D. Walther and C. Koch, "Modeling attention to salient proto-objects", Neural Networks, vol. 19, no. 9, **(2006)**, pp. 1395-1407.
[14] S. W. Ban, Y. M. Jang and M. Lee, "Affective saliency map considering psychological distance", Neurocomputing, vol. 74, **(2011)**, pp. 1916-1925.
[15] R. Carmi and L. Itti, "Visual causes versus correlates of attentional selection in dynamic scenes, Vision Search", vol. 46, no. 26, **(2006)**, pp. 4333-4345.
[16] B. Kim, S. W. Ban and M. Lee, "Top-down attention based on object representation and incremental memory for knowledge building and inference", Neural Networks, vol. 46, **(2013)**, pp. 9-22.
[17] A. Torralba, A. Oliva, M. Castelhano and J. M. Henderson, "Contextual guidance of attention in natural scenes: The role of global features on object search", Psychological Review, vol. 113, no. 4, **(2006)**, pp. 766-786.

[18] J. Oh, D. O. Kim, C. B. Kwon and S. W. Ban, "Biologically inspired selective attention model for efficient visual target detection", Proceedings of Conference of IEIE (In Korean), Jeju, Korea, **(2014)**, pp. 2086-2087.

Authors

Jaeho Oh, is currently pursuing MS degree in the Department of Information & Communication Engineering, Dongguk University, Gyeongju, Gyeongbuk, Korea. His research interests include biologically motivated vision system, deep learning, pattern recognition, and intelligent signal processing.

Chang-Beom Kwon, is currently pursuing MS degree in the Department of Information & Communication Engineering, Dongguk University, Gyeongju, Gyeongbuk, Korea. His research interests include pattern recognition, eye-tracking, intelligent signal processing, u-health care system.

Sang-Woo Ban, received the Ph.D. degree in electrical engineering from Kyungpook National University in 2006 and is currently an associate professor in the Department of Information & Communication Engineering, Dongguk University, Gyeongju, Gyeongbuk, Korea. His research interests include biologically motivated vision system, brain science and engineering, pattern recognition, deep learning, neural networks, intelligent signal processing, and intelligent sensor system.