Biologically Inspired Selective Attention Model Using Human Interest

Le Dong[1], Sang-Woo Ban[2] and Minho Lee[1]

[1]School of Electrical Engineering and Computer Science, Kyungpook National University
1370 Sankyuk-Dong, Puk-Gu, Taegu 702-701, Korea

[2]Department of Information & Communication Engineering, Dongguk University
707 Seokjang-Dong, Gyeongju, Gyeongbuk, 780-714, Korea

ledong@ee.knu.ac.kr, swban@dongguk.ac.kr, mholee@knu.ac.kr

Abstract

A new selective attention model is proposed in this paper, which integrates a top-down attention mechanism into a bottom-up saliency map model to generate salient areas related with human interest. Human selects the certain area from natural scene and decides whether the selected area is preference or refusal. The fuzzy adaptive resonance theory (ART) network trains and memorizes the characteristic of that area, also generates a refusal or a preference signal so that the sequence of test areas is modified to be a desired scan path. The proposed model generates a plausible scan path based on human interest by endowing weight values to feature maps in a course of constructing the saliency map.

Keyword: Knowledge representation, selective attention model, autonomous mental development, self organizing feature map, adaptive resonance theory network

I. Introduction

We can obtain a sequence of salient areas according to visual stimuli using the bottom-up saliency map model. However, the bottom-up saliency map model may select unwanted area and generate unreasonable scan path because it just generates the salient sequence based on the primitive features such as intensity, edge, color and symmetry. On the other hand, human being can learn and memorize the characteristics of the unwanted area, and also inhibits or reinforces attention to that area in subsequent visual search. As a previous work, Itti and Koch introduced brain-like model to generate the saliency map. Based on the Treisman's result [1], they use three types of bases such as intensity, orientation and color information, to construct a saliency map in a natural scene [2]. Koike and Saiki proposed that a stochastic WTA enables the saliency-based search model to cause the variation of the relative saliency to change search efficiency, due to stochastic shifts of attention [3]. In a hierarchical selectivity mechanism, Sun and Fisher integrated visual salience from bottom-up groupings and the top-down attention setting [4]. Ramström and Christensen calculated saliency with respect to a given task using a multi-scale pyramid and multiple cues. Their saliency computations were based on game theory concepts, specifically a coalitional game [5]. However, the weight values of these feature maps for constructing the saliency map are still determined artificially. Moreover, all of these models are non-interactive with environment and human being, and resultantly it is insufficient to give confidence of the selected salient area whether the selected area is interesting. Furthermore, the previous training selective attention model was based on the salient area generated by bottom-up process, thus the interesting area to the human supervisor in natural scene may not be reflected in the plausible scan path.

In this paper, we propose a new selective attention model to mimic such a human-like selective attention mechanism not only with truly bottom-up process but also with interactive process to skip a refusal area with salient feature and to pay attention to a preference area according to human inclination in subsequent visual search process. In order to implement the intentional selective attention model, we use the fuzzy adaptive resonant theory (Fuzzy ART) networks in conjunction with the bottom-up saliency map (SM) model. It is well known that the Fuzzy ART model maintains the plasticity required to learn new patterns, while preventing the modification of patterns that have been learned previously [6]. Thus, the characteristics of feature information contained in a selected area from the natural image are used as an input data of Fuzzy ART models that is to learn and generalize the characteristics of feature information of the selected area in natural scene. In training process, the Fuzzy ART networks learn about feature characteristics of the areas that are decided by human supervisor interactively, which is different from the conventional Fuzzy ART network [7]. In test mode, the vigilance parameter in the Fuzzy ART networks determines whether the new input area is interesting or not, because the Fuzzy ART networks memorize the characteristics of the interesting areas. If the vigilance value is larger than a threshold, the Fuzzy ART networks restores the corresponding area to match with new feature map for the saliency map process and modifies the scan path by generating the weight values so that the selected area is to be the most salient area and/or the least salient area according to human's preference and/or refusal, respectively. The proposed model can generate more intentional scan path like human visual attention system and reflects the relative importance of different feature maps in consideration of human supervisor interest.

The previously proposed bottom-up saliency map model and the intentional selective attention model considering human interest are explained in section II. Experimental results will be followed in this section and conclusion is made in the end.

II. Selective Attention Model with Preference Mechanism

A. Previously Proposed Bottom-up Saliency Map

In order to model the human-like bottom-up visual attention mechanism, four bases of intensity (I), edge (E), color (RG and BY) and symmetry information (S) are used as shown in Fig. 1, for which the roles of retina cells and lateral geniculate nucleus (LGN) are reflected in the previously proposed attention model [4]. The feature maps ($\overline{I}$,$\overline{E}$, $\overline{S}$, and $\overline{C}$) are constructed by center surround difference and normalization (CSD & N) of four bases, which mimics the on-center and off-surround mechanism in our brain, and then are integrated by the maximum of entropy process. The independent component analysis (ICA) algorithm is used to implement the maximum of entropy process that models the roles of the primary visual cortex for redundancy reduction. The reason why we use the ICA is that it is the best way to reduce redundancy [8, 9].

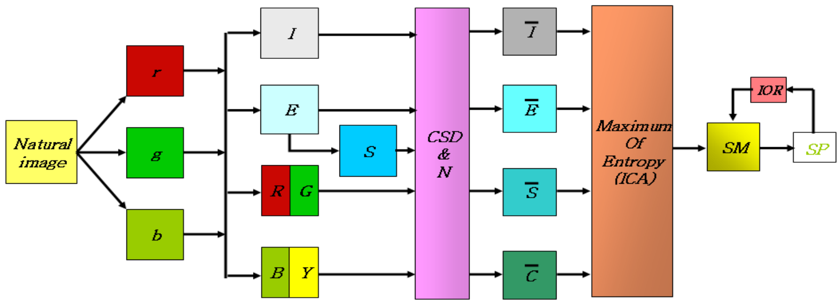


Fig. 1. The architecture of saliency map model (r : red, g : green, b : blue, I : intensity feature, E : edge feature, S : symmetry feature, RG : red-green opponent coding feature, BY : blue-yellow opponent coding feature, *CSD & N* : center-surround difference and normalization, $\overline{I}$: intensity feature map, $\overline{E}$: edge feature map, $\overline{S}$: symmetry feature map, $\overline{C}$: color feature map, ICA : independent component analysis, SM : saliency map, SP : saliency point, IOR : inhibition of return).

In the course of preprocessing, we used a Gaussian pyramid with different scales from 0 to n level, in which each level is made by subsampling of 2^n, thus constructing five feature maps. It is to reflect the non-uniform distribution of retinotopic structure. Then, the center-surround mechanism is implemented in the model as the difference between the fine and coarse scales of Gaussian pyramid images. Consequently, five feature maps are obtained by the following equations.

$$I(c, s) = |I(c) \otimes I(s)| . \quad (1)$$

$$E(c, s) = |E(c) \otimes E(s)| . \quad (2)$$

$$S(c,s)=|S(c) \otimes S(s)| . \quad (3)$$

$$RG(c,s)=|R(c) - G(c)| \otimes |G(s) - R(s)| . \quad (4)$$

$$BY(c,s)=|B(c) - Y(c)| \otimes |Y(s) - B(s)| . \quad (5)$$

$$\overline{I} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} N(I(c,s)), \quad \overline{E} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} N(E(c,s)), \quad \overline{S} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} N(S(c,s))$$
$$\overline{C} = \bigoplus_{c=2}^{4} \bigoplus_{s=c+3}^{c+4} N(RG(c,s) + BY(c,s)) \ . \quad (6)$$

where "⊗" represents interpolation to the finer scale and point-by-point subtraction, N stands for normalization operation, c and s are indexes of the finer scale and the coarse scale, respectively. Totally, 30 feature maps are computed because the five feature maps individually have 6 different scales. Feature maps are combined into four "characteristic maps", as shown in Eq. (6), where $\overline{I}$, $\overline{E}$, $\overline{S}$ and $\overline{C}$ stand for intensity, edge, symmetry, and color opponency respectively and are obtained through across-scale addition "⊕". To obtain ICA filters, the four characteristic maps ($\overline{I}$, $\overline{E}$, $\overline{S}$, and $\overline{C}$) that are used for input patches of the ICA. In order to realize the ICA filter, 20,000 patches of $7\times7\times4$ pixels are selected from four characteristic maps at random. Each patch consists of a column in the input matrix of which the rows and columns are 196 and 20,000 respectively. The basis functions are determined using the extended infomax algorithm [9]. The learned basis functions are with a dimension of 196×196. Each row of the basis functions represents a independent filter and that is ordered according to the length of the filter vector.

We apply the obtained ICA filters to the four characteristic maps ($\overline{I}$, $\overline{E}$, $\overline{S}$, *and* $\overline{C}$) as shown in Eq. (7), and obtain saliency map according to Eq. (8).

$$E_{ri} = FM_r * ICs_{ri} \qquad for \ \ i = 1,\ldots,N, \quad r = 1,..,4 \ . \quad (7)$$

$$S(x,y) = \sum E_{ri}(x,y) \quad for \quad all \quad i \quad . \tag{8}$$

where N denotes the number of filters. The convolution result E_{ri} represents the influences of the four characteristic maps on each independent component and the most salient point $\boldsymbol{P}$ is computed by maximum operator, then an appropriate focus area centered by the most salient location is masked off and the next salient location in the input image is calculated using the saliency map model. The symmetrical property of the object may be used as a guideline to determine the suitable size and shape of the masking area due to an object with symmetrical property in general. This inhibition of return (IOR) mechanism indicates that previously selected salient location is not considered duplicate.

B. Intentional Selective Attention Model According to Human Interest

Human ignores unwanted area even if it has salient primitive features, and can memorize the characteristics of the unwanted area, also do not give an attention to a new area with similar characteristics of the previously learned unwanted area. Correspondingly, human can pay an attention to an interesting area even if it does not have salient primitive features, or it is less salient than any other area. The selective attention model is proposed to mimic such a human-like selective attention mechanism considering not only the primitive input features but also interactive property with environment. Moreover, human brain can learn and memorize a multitude of new things without catastrophic forgetting of existing ones. It is well known that the Fuzzy ART network can be easily trained for additional input pattern and also can solve the stability-plasticity dilemma in conventional multi-layer neural network [10]. Therefore, the Fuzzy ART network is employed with the bottom-up saliency map model to implement an intentional selective attention model that can interact with human supervisor. During the training process, the Fuzzy ART networks learn and memorize the characteristics of the refusal areas or preference areas selected by human supervisor from natural scene. After successful training of the Fuzzy ART networks, an unwanted salient area is refused and a desired area is preferred by the vigilance value of the Fuzzy ART network.

As shown in Fig. 2, any ART-type net can be characterized by its preprocessing, choice, match and adaptation rule, where choice and match define the search circuit for a fitting prototype. With Fuzzy ART, these rules are as follows [11].

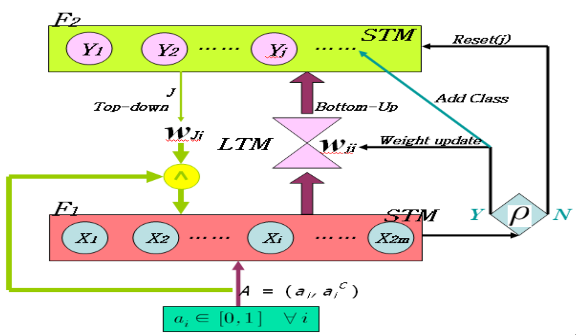


Fig. 2. Fuzzy ART network

▪Preprocessing

All values of an input pattern must fit into the interval [0, 1]

$$i_k \in [0,1] \quad \forall k \quad . \tag{9}$$

▪Choice

Net activities, leading to a preliminary choice of a prototype, are determined using the fuzzy conjunction ($\wedge$), which is defined by

$$\begin{aligned} x \wedge y &= \min\{x, y\} \\ X \wedge Y &= (x_1 \wedge y_1, \cdots, x_m \wedge y_m) \end{aligned} \quad . \tag{10}$$

A single net activity t_j can be seen as the degree of prototype W_j, being a fuzzy subset of input pattern I

$$t_j = \frac{|I \wedge W_j|}{\alpha + |W_j|} . \tag{11}$$

where Y is fuzzy subset of X if $X\&Y=Y$. This size of a vector ($|X|$) is determined by its L_1-norm, the sum of its components. The choice parameter alpha provides a floating point overflow, if $|W_j| \rightarrow 0$. Simulations in this paper are performed with a value $\alpha \approx 0$.

▪Match

The similarity of input I and current winning prototype W_J is measured by the degree of I being a fuzzy subset of W_J. Resonance and adaptation occurs, if

$$\rho \leq \frac{|I \wedge W_J|}{|I|} . \tag{12}$$

▪Adaptation

The winning prototype W_J is adapted by moving its values toward the common MIN vector of I and W_J

$$W_J^{(new)} = \eta(I \wedge W_J^{(old)}) + (1-\eta)W_J^{(old)} . \tag{13}$$

Carpenter and Grossberg mention a problem of cluster proliferation that can occur with Fuzzy ART. It is avoided by normalizing inputs to a constant vector length. One possible method is to use a Euclidean normalization to convert an input pattern A into a coded input I. But the main disadvantage of this method is the complete loss of any information stored in the vector length of an input pattern. Therefore, a modified normalization variant called complement coding is typically used to set all input patterns to a common vector length. An original vector A is coded into an input pattern I by adding the complements of its elements to the original vector as shown in Eq. (14) [12].

$$\begin{aligned} I &= (A, A^c) \\ &= (a_1, \cdots, a_k, 1-a_1, \cdots, 1-a_k) \quad a_i \in [0,1] \quad \forall i \end{aligned} . \tag{14}$$

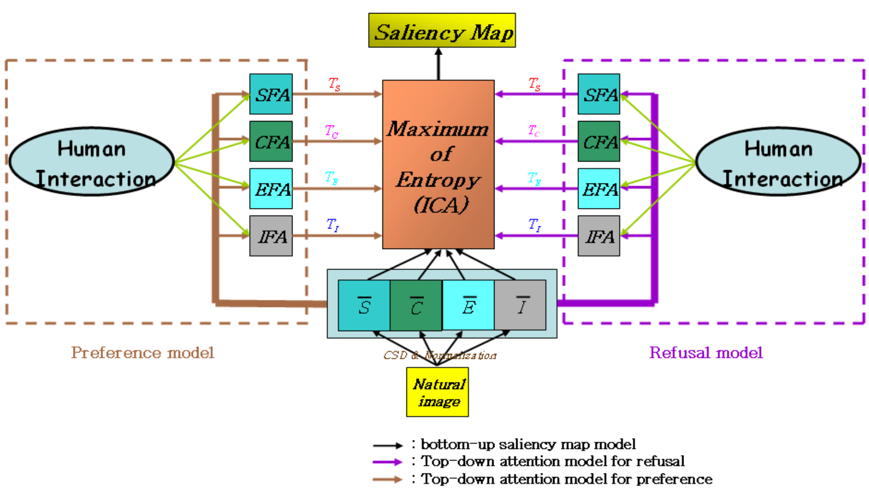
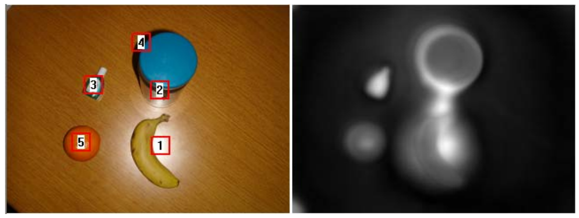


Fig. 3. The architecture of the proposed selective attention model using Fuzzy ART network ($\bar{I}$: intensity feature map, $\bar{E}$: edge feature map, $\bar{S}$: symmetry feature map, $\bar{C}$: color feature map, *SFA*: Fuzzy ART network for symmetry feature map, C*FA*: Fuzzy ART network for color feature map, E*FA*: Fuzzy ART network for edge feature map, *IFA*: Fuzzy ART network for intensity feature map, T_I: intensity feature weight, T_E: edge feature weight, T_S: symmetry feature weight, T_C: color feature weight).

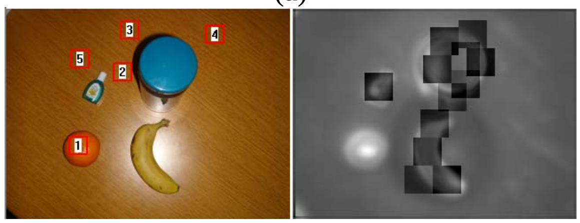
Fig. 3 shows the architecture of the proposed selective attention model for reflecting a human interest. In training mode, the inputs of the Fuzzy ART networks consist of the feature information such as intensity, color, edge and shape characteristics related with attention area selected by the supervisor, and then the supervisor decides whether it is a preference area or a refusal area. If the selected area is a refusal area, the refusal part of the Fuzzy ART model trains and memorizes characteristics of that area to be ignored in later processing. If some area is decided to be preference by the supervisor, that area is trained by the preference part of the Fuzzy ART model. After training process of the Fuzzy ART models is successfully finished, they memorize the characteristics of feature information of refusal areas and preference areas. If a test image contains an area of which the feature characteristics occur in the resonance in the Fuzzy ART memory for human preference, the selected area is intensified in the saliency map by endowing a positive weight value to the output of the Fuzzy ART network for preference to be selected as most salient areas. On the contrary, if the Fuzzy ART memory for human refusal has a resonance for an area in test image, the related area is deteriorated in the saliency map by endowing a negative weight value to the output of the Fuzzy ART network for refusal so as not to be selected as salient area. The resonance for both the Fuzzy ART networks for preference and refusal is controlled by setting a threshold for the vigilance value. Also, the weight values for intensifying and deteriorating the saliency map can be trained by a learning algorithm, which will be considered in later. In this paper, the weight values are fixed to a value of which the magnitude is obtained by scaling of the vigilance values in each feature map so that each feature characteristic can be considered respectively to reflect relative importance of the feature information to construct the intentional selective attention model.

C. Computer Experimental Results

Figs. 4, 5 and 6 show the simulation results of the proposed selective attention model that can reflect human preference and refusal. In Fig. 4, the orange color is decided as preference feature by human supervisor. The characteristic of orange color is trained in the Fuzzy ART network to memorize the color feature of orange as preference one. As shown in Fig. 4 (b), after training the preference for the orange color, the orange becomes the most salient area and several areas with similar orange color feature are selected as salient areas, which show the proposed model can appropriately reflect human preference. Fig. 5 shows the simulation result of the test image with different object sizes and variant luminance, which shows that the proposed model can give robust performance against environment changes. Furthermore, Fig. 6 shows two simulations results for skin color preference. As shown in the saliency map of the right image in Fig. 6 (b), two face areas with skin color becomes much more salient than any other area by preference and refusal mechanism. Moreover, Fig. 6 shows that the proposed model can be utilized to localize face candidate areas. According to the simulation results, after training process of the Fuzzy ART network is successfully finished, the proposed model can generate more plausible scan path like human visual attention system in consideration of human supervisor preference and refusal.



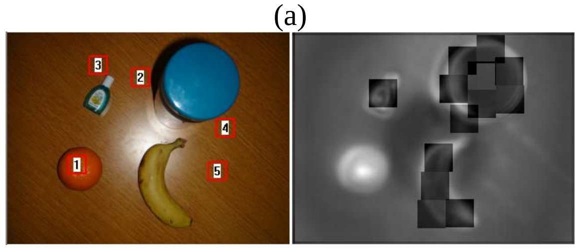
(a)



(b)

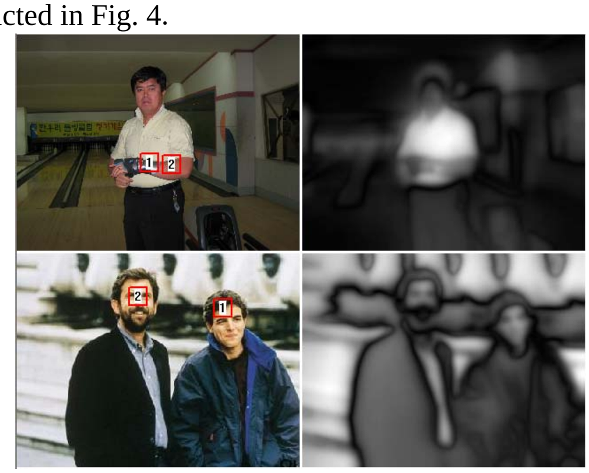
Fig. 4. Simulation results: (a) scan path (left part) and saliency map (right part) generated by the bottom-up saliency map model; (b) scan path (left part) and saliency map (right part) generated by the proposed intentional selective attention model after training for orange color preference.

(a)



(b)

Fig. 5. Simulation results under environment changes: (a) scan path (left part) and saliency map (right part) generated by the bottom-up saliency map model; (b) scan path (left part) and saliency map (right part) generated by the proposed intentional selective attention model after same training conducted in Fig. 4.



(a)

(b)

Fig. 6. Simulation results: (a) scan path (left part) and saliency map (right part) generated by the bottom-up saliency map model; (b) scan path (left part) and saliency map (right part) generated by the proposed intentional selective attention model with skin color preference mechanism.

III. Conclusion

We proposed a new selective attention model that can reflect human preference and refusal. The proposed model is implemented using a Fuzzy ART network in conjunction with a biologically motivated bottom-up saliency map model. The Fuzzy ART network generates weight values to the maximum of entropy process for reflecting preference or refusal of corresponding feature information whenever there is resonant characteristic information in a test image. Experimental results suggest that the proposed method fixes a reasonable intentional salient region, and also the scan path can be modified to a desired one through interaction with human. Human preference can be reflected through feature map domain in some specific application. Our proposed selective attention model can autonomously focus on a salient area based on intentional selection mechanism in natural scene and refuse to pay attention to an unwanted area or prefer a desired one potentially through incremental learning, which can play an important role for autonomous incremental intelligence that will be required for future humanoid system.

References

[1] A. M. Treisman, G. Gelde, “A feature-integrations theory of attention cognitive psychology,” vol. 12. 1, 1980, pp. 97-136.

[2] L. Itti, C. Koch, E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” in *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 20, no. 11, 1998, pp. 1254-1259.

[3] T. Koike, J. Saiki, “Stochastic guided search model for search asymmetries in visual search tasks,” in *Lecture Notes in Computer Science*, vol. 2525, Heidelberg: Springer-Verlag, 2002, pp. 408-417.

[4] Y. Sun, R. Fisher, “Hierarchical selectivity for object-based visual attention,” in *Lecture Notes in Computer Science*, vol. 2525, Heidelberg: Springer-Verlag, 2002, pp. 427-438.

[5] O. Ramström, H. I. Christensen, “Visual attention using game theory,” in *Lecture Notes in Computer Science*, vol. 2525, Heidelberg: Springer-Verlag, 2002, pp. 462-471.

[6] P. D. Wasserman, *Neural Computing Theory and Practice*. Australia, 1989.

[7] S. B. Choi, *Selective Attention System Using Inhibition and Reinforcement Function at the Saliency Area*. *Ph.D. diss.*, Dept. of Sensor Engineering, Kyungpook National Univ., 2004.

[8] A. J. Bell, T. J. Sejnowski, "The independent components of natural scenes are edge filters," in *Vision Research*, vol. 37, 1997, pp. 3327-3335.

[9] T. W. Lee, *Independent Component Analysis-Theory and Application*. USA, 1998.

[10] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, D. B. Rosen, "Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps," in *IEEE Trans. Neural Networks*, vol. 3, no. 5, 1992, pp. 698-713.

[11] T. Frank, K. F. Kraiss, T. Kuklen, "Comparative analysis of Fuzzy ART and ART-2A network clustering performance," in *IEEE Trans. Neural Networks*, vol. 9, no. 3, 1998, pp. 544-559.

[12] J. A. Bednar, R. Miikkulainen, "Self-organization of innate face preferences: could genetics be expressed through learning?" in *Proceeding of 17th National Conference on Artificial Intelligence*, 2000, pp. 117-122.

Le Dong is a Ph. D. candidate, School of Electrical Engineering & Computer Science, Kyungpook National University, Taegu, Korea. Her research interest includes brain science and engineering, knowledge representation, neural networks, pattern recognition techniques and biologically motivated active vision systems.

Sang-Woo Ban got a Ph. D. from School of Electrical Engineering & Computer Science, Kyungpook National University, Taegu, Korea in 2006. He is currently an assistant professor in Department of Information & Communication Engineering, Dongguk University, Gyeongju, Gyeongbuk, Korea. His research interest includes brain science and engineering, intelligent sensor systems, neural networks, pattern recognition techniques, and biologically motivated active vision systems.

Minho Lee graduated from Korea Advanced Institute of Science and Technology in 1995, and is currently a professor of School of Electrical Engineering & Computer Science, Kyungpook National University, Taegu, Korea. His research interests include active vision systems based on human eye movements, selective attention, independent component analysis, active noise control, and intelligent sensor systems. (Homepage: http://abr.knu.ac.kr)