A Method to Predict Human Emotion Using Sentimental Similarity

Hyeong-Joon Kwon and Kwang-Seok Hong

College of Information and Communication Enginnering, Sungkyunkwan University, 300, Chunchun-dong, Jangan-gu, Suwon, Kyungki-do, 440-746, South Korea
katsyuki@skku.edu and kshong@skku.ac.kr

Abstract

In this paper, we propose a method to predict human emotion based on a multi-dimensional emotion model using sentimental similarity. The proposed framework predicts the user's level of emotional response to an emotional stimulus, based on a sensitivity database that consists of self-assessment manikin-based integer-scale data, rated on various object stimulations by many users. We experimented on 72 users, with 1,073 stimulation objects, based on the International Affective Picture System, and used Thayer's arousal-valence 2-dimensional emotion model to verify the proposed framework. As a result, we have confirmed that the proposed framework can predict user emotion using an arousal-valence model.

Keywords: *Affective Computing, Emotion Prediction, Human-Computer Interaction*

1. Introduction

Affective computing is an emerging post-generation computing paradigm. Its concept is to make that a computer or machine can cognitive emotion of human. Related works are receiving careful study by various academic fields, which include electronic and electric engineering, cognitive science, human engineering, computer engineering and psychology [1].

A method of approach in electronic, electric and computer engineering fields consists of the emotion recognition of humans. These use facial images, speech and bio-signals that contain heartbeat, state of tension, skin temperature and circulation of the blood, to recognize human emotions [2, 3]. Above all, electroencephalogram (EEG) of brain wave was noted by many human-computer interaction (HCI) researchers [4], because the brain wave should have the inner working of a human. These kinds of existing works have a critical problem or limitation, from the viewpoint of realization for affective computing, which is hard pressed to quantitatively verify the emotion recognition method. Because we do not know the real emotion of a target user, it is difficult to find the basis of recognition success. A method of approach in cognitive science and psychology field consists in the emotional modeling of humans [5, 6]. Many psychology researchers have been studying visualization and the principles of emotion for a long time. In effect, various multi-dimensional emotion models have been proposed by Russell and Thayer, among others [7, 8].

Recent studies reproduce human sense organs such as eyes and ears by computer using pattern recognition and machine learning technology. In addition, more studies on human emotion recognition (called emotion computing) is in progress. The method based on sight is processed with a face visual to determine different features of emotion from one’s facial expressions, and is recognized by computer. Javier Movellan has announced that a computer can determine great and small changes from every part of a human face to recognize emotion such as anger, sadness, displeasure, pleasure, and more. The method based on hearing uses different human voice features from emotional changes. Speech signals differ for each individual for distinct articulator and speaking habits, so usual voice signals and other voice signals with features are detected to recognize emotion.

However, these approaching methods involve several limitations. First, visual and voice methods are useful to recognize outward appearance constituents. But emotion is the constituent not exposed by outward appearance, so there is always a limitation from using visual and voice methods. Second, facial expressions and human voices differ for all individuals on each emotion. Humans may express the exact facial expression and voice to represent different emotions, and emotions toward the same object may vary. This means that existing methods using visual and vocal features cannot find the sensibility of an individual, and have limitations in emotion recognition. Human emotions differ depending on individual sensibilities, so the computer must reflect individual sensibilities for more accurate emotion recognition. For example, an object may cause one person to feel sad, while another encounter with the same object may not cause sadness; this is because of the difference in sensibilities due to individual experience. Therefore, the study of emotion prediction or recognition methods based on the sensibilities of individuals is necessary.

In this paper, we propose a novel method that is based on psychological theory. The proposed method is a framework to predict human emotion before stimulation. We use Thayer's 2-dimensional model, which consists of arousal and valence. The arousal and valence rating of many users of various stimuli are used as a sensitivity database for each user. The proposed framework can be used for existing studies, with the purpose of improving the emotion recognition ratio and confidence of the emotion recognition result.

We give full details of the proposed framework in Section 2, and then in Section 3, we show experimental results that contain the prediction accuracy of an emotional response level, and a directional distance error between the real position and predicted position, based on the arousal-valence emotion model. Section 4 covers future works and conclusions of this study.

2. Proposed Framework of Emotion Prediction

Figure 1 shows the architecture of the proposed framework. It basically consists of 2 steps. In the first step, it predicts the level of emotional response (ER) of a target user, based on the multi-dimensional emotion model. Then it maps a combination of the predicted level of ER to the emotion model. The details are described below.

The sensitivity database consists of n-ER elements, which can contain emotional response elements, based on elements of the emotion model that are based on psychological principle. For example, Thayer's 2-dimensional emotion model consists of arousal and valence, and Russell's 2-dimensional emotion model consists of activation and pleasant [7, 8]. Content of the sensitivity database is constructed by integer rating of many users of various stimulations with objects. In psychology, the self-assessment manikin (SAM) is widely used as a rating method [9]. The SAM is useful, regardless of the nationality, ethnicity, language or illiteracy of a human individual. The rating scale of SAM is from 1 to 9, as integer number. A subject for constructing the database evaluates the response level, after stimulation with an object. Figure 2 shows the evaluation form of SAM, and the structure of a matrix for the emotional response element in the sensitivity database. The rating of each user is inputted to the emotional response. Namely, the rating of the mth user on the kth object is stored in a matrix of relevant emotional response elements.

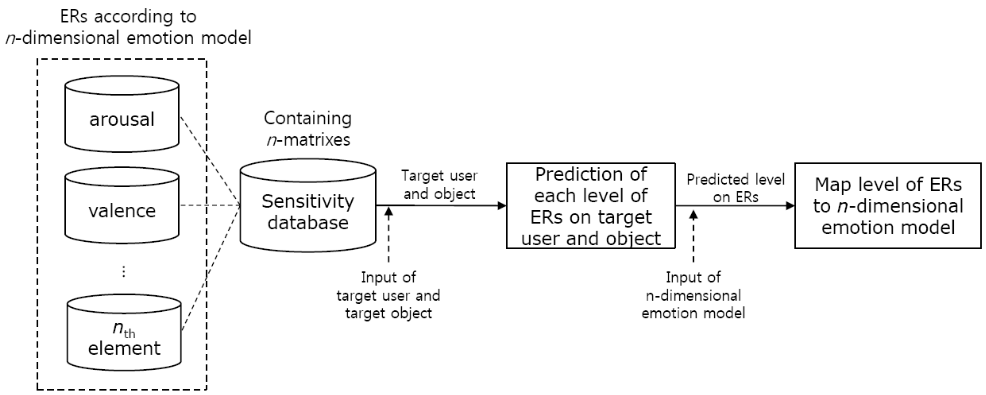

Figure 1. The architecture of the proposed framework

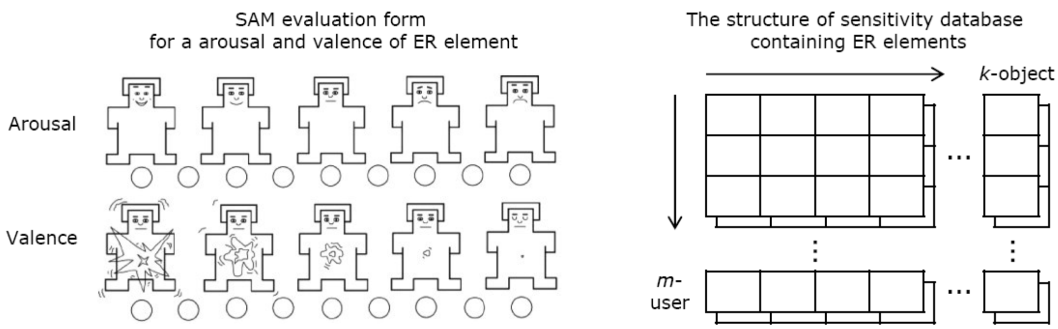

Figure 2. SAM evaluation form and the structure of a matrix in sensitivity database

A prediction of ERs on the target user and target object is calculated by object similarity in each ER. A similarity between the target object and each other object is used as a weight for prediction. We can consider the Pearson Correlation Coefficient (PCC) of Eq. (1), Cosine Coefficient (COS) of Eq. (2), Euclidean Distance (ED) of Eq. (3) as similarity methods of linear variables, and the weighted average (P) of Eq. (4) as a prediction [10]. In prediction, a threshold for neighbor selection is necessary, to predict the rating of a target user on a target object.

$$PCC(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \overline{X})(y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{Y})^2}} \tag{1}$$

$$COS(X,Y) = \frac{X \bullet Y}{\|X\|\|Y\|} = \frac{x_1 y_1 + x_2 y_2 \cdots x_n y_n}{\sqrt{x_1^2 + y_1^2}\sqrt{x_2^2 + y_2^2}\cdots\sqrt{x_n^2 + y_n^2}} \tag{2}$$

$$ED(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{3}$$

$$P(u_t, o_t) = \frac{\sum_{i=1}^{m} sim(o_t, o_i) r_{o_i}}{\sum_{i=1}^{m} r_{o_i}} \tag{4}$$

where, X and Y are the target object and another object, n is the number of objects rated by common users, x_i and y_i are rating scores rated by the same user, m is the threshold of the number of similar objects, $sim(o_t, o_i)$ is the similarity between the target object o_t and another object o_i, and $P(u_t,o_t)$ is the quantitative result of a prediction level on an ER element. The number of P accords with the number of ER elements in the emotion model. After that, every P is mapped to the multi-dimensional emotion model. The number of dimensions in the model is equal to the number of P. Each P is matched to a unique axis in the emotion model. A position in the emotion model is connected to an emotion that contains happy, sad, peaceful, and angry.

4. Experimental Results

We have collected subjective emotions of individual users using SAM over IAPS (International Affective Picture System) of an often-used emotion induction dataset from 72 users, to build individual emotional sensibility data. The IAPS is being developed to provide a set of normative emotional stimuli for experimental investigations of emotion and attention. The goal is to develop a large set of standardized, emotionally-evocative, internationally accessible, color photographs that includes contents across a wide range of semantic categories.

The IAPS (pronounced EYE-APS), along with the International Affective Digitized Sound system (IADS), the Affective Lexicon of English Words (ANEW), as well as other collections of affective stimuli, are being developed and distributed by the NIMH Center for Emotion and Attention (CSEA) at the University of Florida in order to provide standardized materials that are available to researchers in the study of emotion and attention. The existence of these collections of normatively rated affective stimuli should:

1) Allow better experimental control in the selection of emotional stimuli,
2) Facilitate the comparison of results across different studies conducted in the same or different laboratory,
3) Encourage and allow exact replications within and across research labs who are assessing basic and applied problems in psychological science.

The IAPS consists of just 1,200 photographs of commonly seen objects from human life [11]. Extremely suggestive or cruel photographs were excluded; 1,073 visual stimulation objects were composed randomly into 8:2 ratios, with 80 % of total objects as training data, and 20 % as test data. At the beginning of the rating, the participants performed a picture recognition task and two cognitive tasks: a perceptual speed task and a vocabulary task. After these tasks, the participants were asked to rate 1,073 pictures on valence and arousal. Thus, each participant rated only one half of the total picture set. In a first block, all 1,073 pictures were rated on valence; in a second block, the same 1,073 pictures were rated on arousal.

We then used PCC, COS, neighbors' average and Raw Moment-based Similarity (RMS) to predict 20 % test data ratings [12]. The prediction results are shown in Figure 3 and Figure 4. The MAE curve, the absolute average of the difference between the real rating and prediction rating, shows increasing neighbors, and rises at the optimal point. The experimental result shows a remarkable result in arousal and valence datasets. The PCC, COS and RMS showed greater prediction accuracy than the neighbors' average.

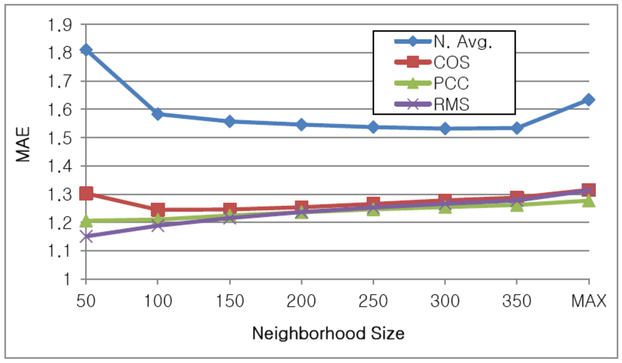

Figure 3. Experimental result of arousal prediction using object-based similarity

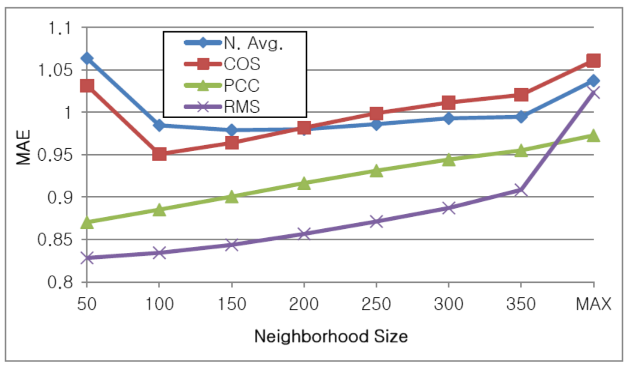

Figure 4. Experimental result of valence prediction using object-based similarity

Chief of all, the RMS algorithm gave matchless performance in valence prediction. The next experiment, in Table 1, is to observe the error distance of emotion model in Thayer's 2-dimensional emotion model, widely used by psychological works, the reason being that it is very intuitive, manageable, and useful for combinations of SAM evaluation. This experiment predicts arousal and valence ratings regarding target objects, and marks real values and predicted values on Thayer's 2-dimensional emotion model. We then measured the lineal absolute distance error between the two points. Euclidean distance was used as a distance measurement. Figures 5 and 6 shows experimental results using user-based similarity. In comparison with object-based similarity, we can confirm that the method using object-based similarity shows more performance than user-based similarity.

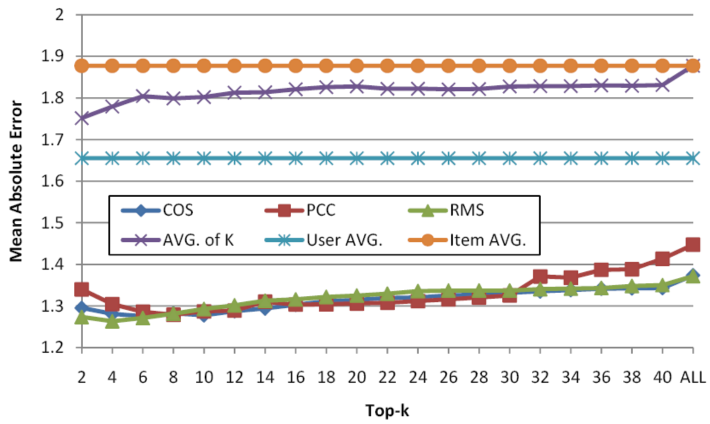

Figure 5. Experimental result of valence prediction using user-based similarity

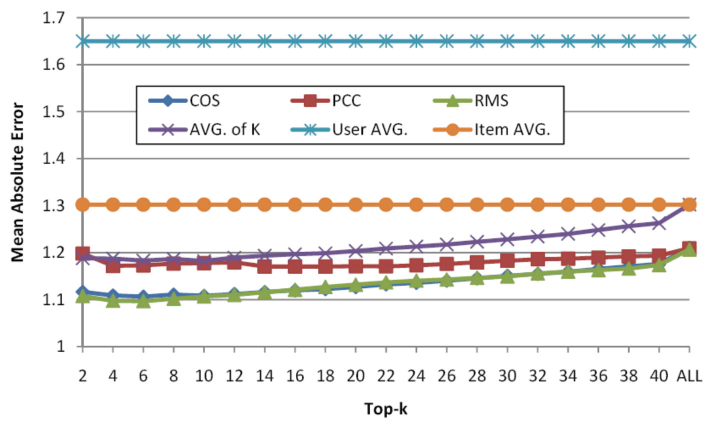

Figure 6. Experimental result of valence prediction using user-based similarity

Table 1 shows the results and experimental conditions, which contain the optimal neighborhood size per similarity method in Figures 3 and 4. As a result, we were able to confirm the effectiveness of the proposed framework, because the diagonal length in a 3 by 3 square is 2.449, and the diagonal length of one quadrant in the 2-dimensional emotion model is 2.828 on the 1-9 integer scale level, which is the standard type. Each quadrant is representative of human feelings, which contain joy, anger, sorrow or happiness.

Table 1. Experimental result of emotion prediction using object-based similarity

Performance Rank	Similarity Method	Optimal Neighborhood Size	Error Distance
3	COS	100	2.461
2	PCC	50	2.372
1	RMS	50	2.248
4	N. Avg.	150	2.682

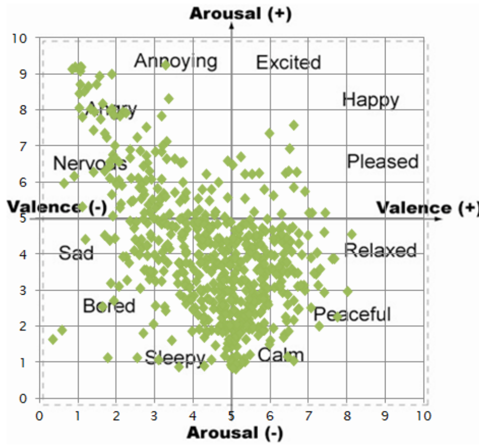

Figure 7. The prediction result distribution of the proposed method

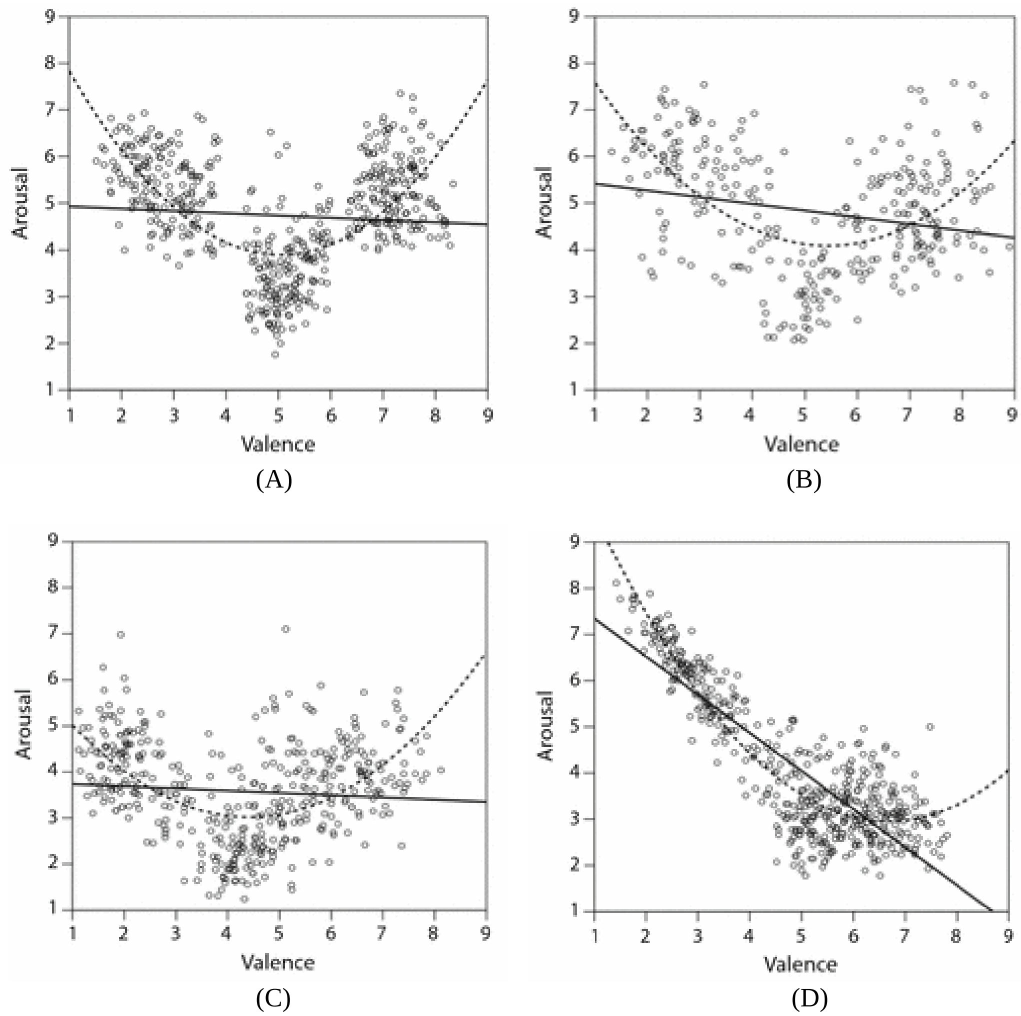

Figure 8. The prediction result distribution of existing studies

Figure 7 shows the prediction result distribution of the proposed method. Our the count of prediction is 214. To verify the generalizability of the obtained ratings, we compared our ratings of young and older adults with available ratings of young adults from previous studies (Ito, *et al.*, 1998; Lang, *et al.*, 1998; Libkuman, *et al.*, 2007; Ribeiro, *et al.*, 2005).2 Whereas the normative data by Lang and colleagues provide ratings for all 504 pictures, the sets used by Ito, *et al.*, Libkuman, *et al.*, and Ribeiro, *et al.*, overlapped with the present set of pictures for only 290 pictures (99 negative, 70 neutral, and 121 positive; categorization based on the normative ratings by Lang, *et al.*, 1998), 430 pictures (141 negative, 136 neutral, and 153 positive), and 426 pictures (139 negative, 135 neutral, and 152 positive), respectively.

5. Conclusion

We propose a novel framework for emotion prediction, based on a multi-dimensional emotion model using object similarity. The proposed framework predicts a level of emotional response element, based on a collective sensitivity database, which consists of multiple matrixes, according to the number of emotional response elements. From the experimental results, the proposed framework showed greater prediction accuracy. In the future, we will study convergence, fusion and combination of the proposed framework, and a multimodal emotion recognition method using facial expression or speech. This approach will be an optimal recognition method that uses external and internal human elements in affective computing.

Acknowledgements

This research was supported by MKE, Korea under ITRC NIPA-2011-(C1090-1121-0008) and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2010-0021411).

References

[1] R. W. Picard, Applications of Affective Computing in Human-Computer Interaction, vol. 59, **(2003)**, pp. 1-2.
[2] G. Castellano, L. Kessous and G. Caridakis, in Lecture Note in Computer Sciences, Edited C. Peter and R. Beale, Springer, Heidelberg, vol. 4868, **(2008)**, pp. 92-103.
[3] C. D. Katsis, N. Katertsidis, G. Ganiatsas and D. I. Fotiadis, IEEE Trans. on Systems, Man and Cybernetics Part A: Systems and Humans, vol. 38, no. 3, **(2008)**.
[4] B. Rachev and A. Smrikarov, (Eds.), "Emotion Recognition Using Brain Activity", Proceedings of the 9th International Conference on Computer Systems and Technologies, **(2008)** June 12-13; Gabrovo, Bulgaria.
[5] J. Edwards, H. J. Jackson and P. E. Pattison, Clinical Psychology Review, vol. 22, no. 6, **(2002)**.
[6] R. P. Hobson, J. Ouston and A. Lee, Psychological Medicine, vol. 18, no. 4, **(1988)**.
[7] R. E. Thayer, (Ed.), "The Biopsychology of Mood and Arousal", Oxford University Press, London, **(1989)**.
[8] J. A. Russell, "A Circumplex Model of Affect", vol. 39, no. 6, **(1980)**.
[9] M. M. Bradley and P. J. Lang, Journal of Behavior Therapy and Experimental Psychiatry, vol. 25, no. 1, **(1994)**.
[10] G. Adomavicius and A. Tuzhilin, IEEE Trans. on Knowledge and Data Engineering, vol. 17, no. 6, **(2005)**.
[11] P. J. Lang, M. M. Bradley and B. N. Cuthbert, "International Affective Picture System (IAPS): Technical Manual and Affective Ratings", NIMH Center for the Study of Emotion and Attention, **(1997)**.
[12] H. -J. Kwon and K. -S. Hong, IEEE Trans. on Consumer Electronics, vol. 57, no. 3, **(2011)**.

Authors

Hyeong-Joon Kwon

He received his B.S. in Computer Engineering at Seoul Health College in 2006 and his M.S. from the Department of Electrical and Computer Engineering at Sungkyunkwan University in 2008. He is presently a Ph.D. candidate at the College of Information and Communication Engineering, Sungkyunkwan University. His current research focuses on IT convergence, human-computer interactions and collective intelligence.

Kwang-Seok Hong

He received his B.S., M.S., and Ph.D. in Electronic Engineering from Sungkyunkwan University, Seoul, Korea, in 1985, 1988, and 1992, respectively. Since March 1995, he has been a professor at Sungkyunkwan University, Suwon, Korea. His current research focuses on HCI, five-sense recognition, interaction and representation. He became a Member of IEEE in 1995.