Design of Prototype-Based Emotion Recognizer Using Physiological Signals

Byoung-Jun Park, Eun-Hye Jang, Myung-Ae Chung, and Sang-Hyeob Kim

This study is related to the acquisition of physiological signals of human emotions and the recognition of human emotions using such physiological signals. To acquire physiological signals, seven emotions are evoked through stimuli. Regarding the induced emotions, the results of skin temperature, photoplethysmography, electrodermal activity, and an electrocardiogram are recorded and analyzed as physiological signals. The suitability and effectiveness of the stimuli are evaluated by the subjects themselves. To address the problem of the emotions not being recognized, we introduce a methodology for a recognizer using prototype-based learning and particle swarm optimization (PSO). The design involves two main phases: i) PSO selects the *P*% of the patterns to be treated as prototypes of the seven emotions; ii) PSO is instrumental in the formation of the core set of features. The experiments show that a suitable selection of prototypes and a substantial reduction of the feature space can be accomplished, and the recognizer formed in this manner is characterized by high recognition accuracy for the seven emotions using physiological signals.

Keywords: Emotion recognition, physiological signal, prototypes, feature selection, particle swarm optimization.

Manuscript received Nov. 13, 2012; revised Mar. 19, 2013; accepted Apr. 1, 2013.

This research was supported by the Converging Research Center Program through the Ministry of Science, ICT and Future Planning, Korea (2013K000329).

Byoung-Jun Park (phone: +82 42 860 1503, bj_park@etri.re.kr), Eun-Hye Jang (cleta4u@etri.re.kr), and Sang-Hyeob Kim (shk1028@etri.re.kr) are with the IT Convergence Technology Research Laboratory, ETRI, Daejeon, Rep. of Korea.

Myung-Ae Chung (machung@etri.re.kr) is with the Creative Future Research Laboratory, ETRI, Daejeon, Rep. of Korea.

http://dx.doi.org/10.4218/etrij.13.0112.0751

I. Introduction

Emotion plays an important role in the contextual understanding of messages from others in speech and visual forms. For effective communication between users and computers, how emotions can be recognized and expressed during human-computer interaction must be considered. In addition, emotion recognition is one of the key steps toward emotional intelligence in advanced human-machine interactions [1]. Emotions are displayed through visual, vocal, and physiological means. Psychologists and engineers have anyalyzed facial expressions, vocal expressions, gestures, and physiological signals in an attempt to understand and categorize emotions [2]-[4]. Physiological signals have been used to recognize human emotions and feelings, as signal acquisition by noninvasive sensors is relatively simple, physiological responses to emotions are relatively similar among different societies and cultures, and there is a strong relationship between physiological reactions and the emotional and affective states of humans [5].

Many previous studies on emotion have reported that there is a correlation between basic emotions and physiological responses [5]-[10]. Emotion recognition using physiological signals has been performed using various machine-learning algorithms, including the Fisher linear discriminant (FLD), the k nearest neighbor (k-NN) algorithm, and support vector machines (SVMs). To use machine-learning algorithms for the recognition problem, it is necessary to define a small and consistent subset of data to improve both the computing speed and method performance. In addition, feature selection constitutes a fundamental development phase of pattern recognition and predetermines the effectiveness of the overall recognition schemes to a significant extent [11]. It has become

apparent that this is essential both to reduce the overall computational overload and to possibly enhance the discriminatory capabilities of the reduced feature space. Any optimization of the feature subspaces quite often involves various mechanisms of evolutionary optimization, as evidenced in other pattern recognition research [12]-[16], including genetic algorithms, evolutionary algorithms, and particle swarm optimization (PSO).

The objective of this study is to acquire physiological signals for seven human emotions (happiness, sadness, anger, fear, disgust, surprise, and stress) and to develop a prototype-based recognizer driven by the instance-based learning paradigm [17], [18] and feature selection for emotion recognition. To obtain the physiological signals for the emotions, twelve subjects participate in the experiments, and ten stimuli sets for the seven emotions are used. For an induced emotion, each subject evaluates the suitability and effectiveness of the presented stimulus, and the physiological signals of each subject are measured for the induced emotion. Physiological signals are determined through skin temperature (SKT), photoplethysmography (PPG), electrodermal activity (EDA), and an electrocardiogram (ECG). From these signals, 28 features are extracted for emotion recognition. To efficiently develop a seven-emotion recognizer, we use one of the evolutionary optimization techniques, namely, PSO. PSO makes two formations in the optimization process. The first formation is realized from choosing the P% of the patterns, as a set of prototypes comes from patterns that contain all seven emotions. The second formation consists of a core set of features, which is a collection of the most meaningful and discriminative components of the original feature space. The design of the optimally reduced feature space is investigated in a parametric setting by varying the size of the prototype set (P%) and the size of the feature set (d%) used in the proposed construct. This study provides an algorithmic framework and demonstrates the effectiveness of this approach. To demonstrate the usefulness of the proposed recognizer for the seven emotions, we discuss the comparative results of emotion recognition using certain machine-learning algorithms, including neural networks (NNs), SVMs, and self-organizing maps (SOMs). It is demonstrated that a suitable selection of prototypes and a substantial reduction of the feature space can be accomplished, accompanied by a higher recognition accuracy using physiological signals for the seven emotions.

II. Induction of Emotions and Acquisition of Their Physiological Signals

In this section, we present the experiments for the induction of the seven emotions using stimuli and the acquisition of physiological signals of these emotions. Six male (20.8 years ± 1.26) and six female (21.2 years ± 2.70) college students participate in this study. Written consent is obtained from each subject prior to conducting the experiments. None of the subjects report any history of medical illness or the use of psychotropic or other medications that would affect their cardiovascular, respiratory, or central nervous systems.

The laboratory used for the experiments is a room 5 m × 2.5 m in size. The room is sound proof (lower than 35 dB), completely blocking all outside noise and artifacts. The laboratory contains a comfortable chair, a 38-inch monitor, an intercommunication device, and CCTV to observe and record the behaviors of each subject. We introduce the experimental procedures and their indications in detail to the subjects. Each subject has an adaptation time of about 30 minutes to feel comfortable in the laboratory environment, and electrodes are then attached to their wrists, fingers, and ankles for measurement of their physiological signals.

1. Emotional Stimuli and Induction of Emotions

In the experiments, we use ten sets of stimuli for each emotion, that is, 70 stimuli, to successfully induce the seven emotions (happiness, sadness, anger, fear, disgust, surprise, and stress). The stimuli are made up of two- to four-minute-long audio-visual film clips captured from movies, documentaries, and TV shows; stills of such clips are shown in Fig. 1.

Audio-visual film clips are widely used because they have the desirable properties of being readily standardized, non-deceptive, and dynamic rather than static. They also have a relatively high degree of ecological validity, in so far as emotions are often evoked by dynamic visual and auditory stimuli that are external to the individual [19], [20].

The suitability and effectiveness of the emotional stimuli are

Fig. 1. Examples of emotional stimuli: (a) happiness (victory, laughing, etc.), (b) sadness (death of parents, longing for mother, etc.), (c) anger (massacre, beating, etc.), (d) fear (ghosts, haunted house, etc.), (e) disgust (vomiting, etc.), (f) surprise (unexpected screams, etc.), and (g) stress (audio/visual noises on the screen, etc.).

Table 1. Suitability and effectualness of emotional stimuli.

		Emotion							Avg.
		Happiness	Sadness	Anger	Fear	Disgust	Surprise	Stress	
Set	1	100% (8.4)	92% (9.5)	75% (9.7)	75% (10)	75% (10.2)	75% (9.3)	92% (9.3)	83% (9.5)
	2	100% (8.9)	100% (9.1)	75% (9.9)	100% (9.9)	92% (10.8)	92% (9.7)	100% (9.1)	94% (9.6)
	3	100% (8.8)	100% (8.7)	75% (9.7)	83% (9.8)	92% (9.9)	100% (9.7)	100% (8.8)	93% (9.3)
	4	100% (9.6)	100% (9.7)	75% (9.5)	92% (9.6)	100% (10.4)	100% (9.9)	100% (8.9)	95% (9.7)
	5	100% (9.6)	100 (9.3)	92% (9.8)	92% (9.7)	92% (9.7)	83% (9.6)	100% (9.3)	94% (9.6)
	6	100% (9.3)	100% (9.3)	92% (9.4)	92% (9.7)	100% (10.3)	83% (9.6)	100% (8.8)	95% (9.5)
	7	100% (9.3)	75% (8.9)	92% (8.9)	83% (9.6)	100% (9.3)	100% (9.5)	92% (9.3)	92% (9.3)
	8	92% (8.0)	100% (9.0)	83% (9.2)	100% (9.3)	83% (10.2)	83% (9.4)	100% (9.3)	92% (9.2)
	9	100% (9.7)	100% (9.2)	92% (9.5)	100% (9.3)	100% (10.1)	83% (8.6)	100% (9.1)	96% (9.4)
	10	92% (8.8)	100% (9.3)	92% (9.7)	75% (8.7)	100% (10.1)	75% (10.3)	100% (9.3)	91% (9.5)
Avg.		98% (9.1)	96% (9.2)	84% (9.5)	89% (9.6)	98% (9.1)	89% (9.5)	98% (9.1)	93% (9.5)

examined in a preliminary study prior to executing the experiments. The suitability of the emotional stimuli is based on the consistency between the target emotions designed to induce each emotion and the categories of the experienced emotions of each subject. The effectiveness is determined by the intensity of the emotions reported and is rated by the subjects on a Likert-type 1 to 11 point scale (1 being the "least happy" or "not happy" and 11 being the "most happy"). Twenty-two college students and ten psychologists, different from the subjects used in the experiments, categorize their experienced emotions according to the seven emotions previously described and estimate the intensity of their categorized emotions on an emotional assessment scale after being shown each film clip. Regarding the suitability and effectiveness of the emotional stimuli, Table 1 shows the results obtained from the preliminary study. The results show that the emotional stimuli have a suitability of 93% and an effectiveness of 9.5 points on average for the 10 sets. The results also show that the suitability of each stimulus ranges from 75% to 100% and the effectiveness ranges from 8.4 points to 10.4 points.

The induction of emotion is carried out as follows. First, the subject is informed of the details of the experiment and fitted with electrodes. Second, for the baseline, physiological signals are measured for the 60 seconds preceding the presentation of a stimulus (film clip). Third, for the emotional state, physiological signals are measured during the presentation of the stimulus, which lasts anywhere between two to four minutes. Lastly, for the recovery, physiological signals are measured for the 60 seconds following the presentation of the stimulus, during which time the subject stabilizes. After a stimulus is presented, the subject rates the emotions they experienced during the presentation of the film clip according to the emotion assessment scale. This procedure is conducted ten times for each of the seven emotions.

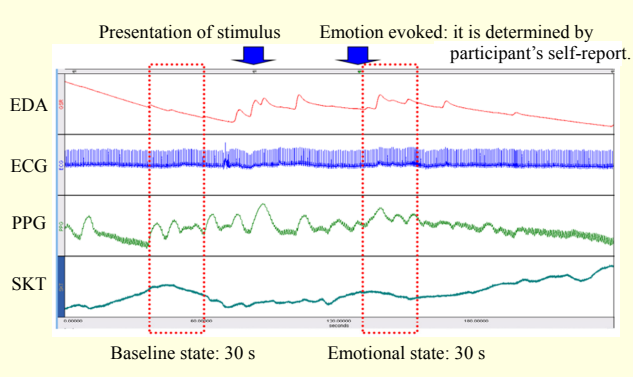

Fig. 2. Analysis of physiological signals.

2. Measurements of Physiological Signals

For an induced emotion, the physiological signals (SKT, PPG, EDA, and ECG [7], [21]) of the subject are measured to determine their baseline and emotional states. Electrodes are attached to the first joint of the ring finger and the last joint of the thumb finger of the non-dominant hand to measure the SKT and PPG, respectively. The EDA is measured through two Ag/AgCl electrodes attached to the middle joint of the index and middle fingers of the non-dominant hand. For the ECG, electrodes are placed on both wrists and the left ankle using a two-electrode method based on lead I. The electrode on the left ankle is used as a reference. The electrodes are filled with a 0.05 molar isotonic NaCl paste to provide a continuous connection between the electrodes and the skin.

The signals obtained are analyzed for 30 seconds from the baseline and emotional states, as shown in Fig. 2. The baseline is selected and analyzed for 30 seconds before an emotional stimulus is presented. The emotional states are determined according to the results of each subject's report regarding which emotion was most strongly expressed during the presentation of each stimulus. Figure 2 shows the analysis ranges of the physiological signals. We extract 28 features from the analysis of the signals within these ranges.

Table 2. Features resulting from physiological signals.

Physiological signals		Features	
EDA		SCL, NSCR, meanSCR	
SKT		meanSKT, maxSKT	
PPG		meanPPG	
ECG	Time domain	Statistical parameter	meanRRI, stdRRI, meanHR, RMSSD, NN50, pNN50
		Geometric parameter	SD1, SD2, CSI, CVI, RRtri, TINN
	Frequency domain	FFT	FFTapLF, FFTapHF, FTnLF, FFTnHF, FFTLF/HFratio
		AR	ARapLF, ARapHF, ARnLF, ARnHF, ARLF/HFratio

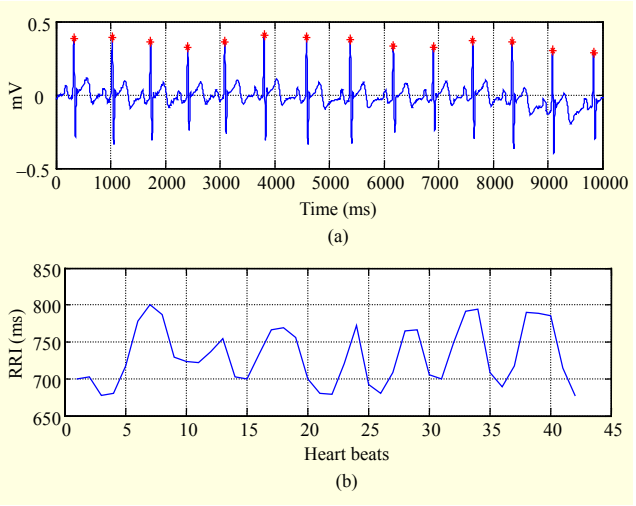

Fig. 3. ECG signal and RRI tachogram: (a) ECG signals for 10 s and (b) RRI tachogram for 30 s.

3. Feature Extraction from Physiological Signals

For the seven emotions, 28 features are extracted from the physiological signals and are summarized in Table 2. Skin conductance level (SCL), average skin conductance response (meanSCR), and number of skin conductance responses (NSCLs) are obtained from the EDA. The mean SKT (meanSKT), maximum SKT (maxSKT), and the mean amplitude of blood volume changes (meanPPG) are obtained from the SKT and PPG, respectively. The ECG signals are analyzed for the time and frequency domains. The analysis of the time domain is divided into statistical and geometric approaches, and the analysis of the frequency domain deals with FFT and AR [22], [23].

Figure 3 illustrates an ECG signal and R-R interval (RRI) tachogram. RRI is the interval time of the *R* peaks on the ECG signal. The RRI and hear rate (HR) offer the mean RRI (meanRRI), standard deviation (stdRRI), the mean hear rate (meanHR), RMSSD, NN50, and pNN50. RMSSD is the square root of the mean of the sum of the squares of differences between successive RRIs. NN50 is the number of RRIs of 50 ms or more, and the proportion of NN50 divided by the total number of RRIs is pNN50. In addition to these, an RRI triangular index (RRtri) and TINN are extracted from the histogram of the RRI density as a geometric parameter. RRtri is used to divide the entire number of RRIs by the magnitude of the histogram of the RRI density, and TINN is the width of the RRI histogram.

The relations between RRI(n) and RRI(n+1) are shown in Fig. 4 in a Poincaré plot. Here, n and n+1 are the n-th and (n+1)th values of the RRI, respectively. In the figure, L is the direction that is most efficient for representing the data, and T is the orthogonal direction of L. The standard deviations,

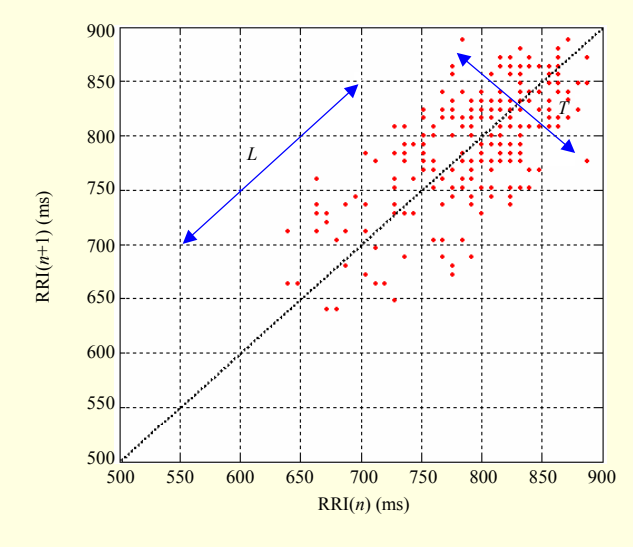

Fig. 4. Poincaré plot of RRI.

SD1 and SD2, are obtained for the T and L directions, respectively. The cardiac sympathetic index (CSI) is calculated by CSI = 4SD2/4SD1, and the cardiac vagal index (CVI) is obtained from CVI = log10(4SD1 * 4SD2) as an emotional feature. SD1, SD2, CSI, and CVI reflect the short-term heart rate variability (HRV), long-term HRV, sympathetic nerve activity, and parasympathetic activity, respectively.

To extract an emotional feature based on physiological signals, we use a fast Fourier transform (FFT) and auto regressive (AR) power spectrum. The low-frequency (LF) band is 0.04 Hz to 0.15 Hz, and the high-frequency (HF) band is 0.15 Hz to 0.4 Hz. The total spectral power between 0.04 Hz and 0.15 Hz is apLF, and the normalized power of apLF is nLF. The terms apHF and nHF indicate the total spectral powers

between 0.15 Hz and 0.4 Hz and the normalized power, respectively. L/Hratio indicates the ratio of LF power to HF power. These parameters result from the means of the FFT and AR. LF and HF are used as indexes of the sympathetic and vagus activities, respectively. L/Hratio reflects the global sympathovagal balance and can be used as a measure of this balance.

III. Design of Prototype-Based Recognizer Driven by PSO

1. Two-Phase Process for Prototype-Based Recognizer

The prototype-based recognizer is a type of instance-based learning that uses only specific instances to solve the recognition problem. This methodology involves two selection problems to recognize a new emotional pattern. One is the selection of emotional prototype patterns, and the other is the feature selection.

A. First Level of Optimization Process: Prototype Formation

We start by choosing the P% of the emotional patterns using PSO; these patterns should come from all seven categories of emotion and are called prototypes. The prototype-based recognizer does not use any model to fit and is only based on the distance between the emotional data (a new input pattern) of a system and prototypes in the memory of the system. Given a set of prototypes and an emotion pattern q without the emotional label as an input of a system, the recognizer finds the closest pattern of prototypes to q in the feature (input) space and assigns to it the emotional label of its nearest prototype. The underlying distance between a given input pattern and a prototype is measured using the weighted Euclidean distance, which is

$$\|\boldsymbol{x}-\boldsymbol{y}\|^2 = \sum_{i=1}^{n} \frac{(x_i - y_i)^2}{\sigma_i^2}, \quad (1)$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are the two patterns in n-dimensional space, and σ_i is the standard deviation of the i-th feature whose value is computed using the emotional prototype set.

B. Second Level of Optimization Process: Feature Selection

Once the prototypes have been formed, we reduce the overall feature space by choosing a core set of features. Often, their number can be quite limited in comparison with the dimensionality of the overall feature space. One can consider the d% of the total number of features, for example, 10% or 20%. Namely, d is the proportion of subfeatures to all features to be selected as the core set. Considering d% of features of the original feature space, we arrive at ${}_nC_{d\times n}$ of possible combinations of features that can be selected to build this core set. For instance, with $n = 28$ achieved from the physiological signals and $d = 20$% of features selected to form the core set, we are faced with ${}_{28}C_{0.2\times 28} = {}_{28}C_6 = 376{,}740$ combinations. This number goes up to 3.11×10^6 and 2.15×10^7 when the percentage of features to be used in the core set of features is equal to 30% and 40%, respectively.

The construction of the core set of features and the formation of the prototypes are computationally challenging and hence require the use of a suitable optimization mechanism that is capable of dealing with the combinatorial nature of the task. Given this, we consider one of the biologically inspired optimization tools, PSO.

2. Particle Swarm Optimization

The underlying principle of PSO involves a population-based search in which individuals representing possible solutions carry out a collective search by exchanging their individual findings while taking into consideration their own experience and evaluating their own performance. PSO is motivated by social behavior and cognition aspects. Regarding the social facet of the search, individuals in a neighborhood interact with one another, and the best behaviors among the individuals are ultimately adopted by all. The cognition aspect of the search underlies the importance of the individual experience in which the element of population is focused on its own history of performance and makes adjustments accordingly [12]-[14].

There are PSO and GAs (genetic algorithms) in a well-known class of biologically inspired optimization techniques. PSO and GAs are similar in the sense that both types of algorithm use a population-based search method and an iteration-based updating method for the optimal solution. In this study, we select PSO for the optimization of the prototype and the feature due to the advantages PSO has over GAs, which are summarized below [24]-[28].

a) PSO has the control parameter for the balance between the global and local exploration of the search space. This feature enhances the search capability of PSO.
b) PSO has memory, namely, information of good solutions is retained and shared by all particles, whereas, in GAs, previous knowledge is destroyed once the population changes.
c) PSO exhibits algorithmic simplicity. A GA consists of three major operators: selection, crossover, and mutation. However, PSO comes with a single operation of velocity.
d) PSO has a simple implementation, and this induces the reduction of computation and eliminates the necessity to select the best operator for a given optimization.

e) Quite often, PSO is superior to other biologically inspired techniques in terms of convergence, speed, and accuracy.

PSO is conceptually simple, easy to implement, and computationally efficient. Unlike many other heuristic techniques, PSO has a flexible and well-balanced mechanism to enhance the global and local exploration abilities [24]. The basic elements of the PSO are briefly introduced below.

Performance index (fitness) — Each particle is characterized by some value of the underlying performance index or fitness. The fitness is reflective of the nature of the problem for which an optimal solution is being sought.

Particles — The vectors of the variables in the n-dimensional search space are denoted by $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_N$. In the search, a swarm is composed of N particles involved, leading to the concept of a swarm. The performance of each particle is described by a fitness function.

Best particles — As a particle wanders through the search space, we compare its fitness at the current position using the best fitness value it has thus far attained. This is done for each element in the swarm. The location of the particle at which it has attained the best fitness is denoted by ***pbest***. Similarly, by ***gbest***, we denote the best location attained among all values of ***pbest***.

Velocity — The particle is moving in the search space with a certain velocity, which plays a pivotal role in the search process. Denote the velocity of the i-th particle by $\boldsymbol{v}_i$. From iteration to iteration, the velocity is governed by the following expression.

$$\boldsymbol{v}_i = w\boldsymbol{v}_i + c_1 r_1 (\boldsymbol{pbest}_i - \boldsymbol{p}_i) + c_2 r_2 (\boldsymbol{gbest} - \boldsymbol{p}_i), \quad (2)$$

where $i = 1, 2, \ldots, N$, c_1 and c_2 are positive constants, and r_1 and r_2 are uniformly distributed random values in [0,1]. Constants c_1 and c_2 are called acceleration constants and are referred as a cognitive parameter and a social parameter, respectively. As the above expression shows, c_1 and c_2 reflect the weighting of the stochastic acceleration terms that pull the i-th particle toward the $\boldsymbol{pbest}_i$ and ***gbest*** positions. Low values allow particles to roam far from the target regions before being pulled back. High values of c_1 and c_2 result in an abrupt movement toward or past the target regions. The inertia factor, w, is a control parameter used to establish the impact of the previous velocity on the current velocity. Hence, it influences the tradeoff between the global and local exploration abilities of the particles.

Overall, the algorithm can be outlined as the following sequence of steps.

Step 1. Randomly generate N particles, $\boldsymbol{p}_i$, and their velocities, $\boldsymbol{v}_i$. Each particle in the initial swarm is evaluated using the fitness function. For each particle, set $\boldsymbol{pbest}_i = \boldsymbol{p}_i$ and search the best particle of ***pbest***. Set the best particle associated with the global best, ***gbest***.

Step 2. Adjust the inertia weight, w. Typically, its values decrease linearly over the search time. We start with $w_{\max} = 0.9$ at the beginning of the search and move down to $w_{\min} = 0.4$ at the end of the iterative process.

$$w(t) = w_{\max} - \frac{w_{\max} - w_{\min}}{iter_{\max}} \times t\ , \quad (3)$$

where $iter_{\max}$ denotes the maximum number of iterations of the search, and t stands for the current index of the iteration.

Step 3. Given the current values of ***gbest*** and $\boldsymbol{pbest}_i$, the velocity of the i-th particle is adjusted according to (2).

Step 4. Based on the updated velocities, each particle changes its position using (4).

$$p_{ik} = v_{ik} + p_{ik}. \quad (4)$$

Step 5. Move the particles in the search space and evaluate their fitness in terms of both $\boldsymbol{pbest}_i$ and ***gbest***.

Step 6. Repeat Steps 2 through 5 until the termination criterion has not been met. Otherwise, return ***gbest*** as the found solution.

3. Prototypes and Features versus PSO

As a generic search strategy, the PSO must be adjusted to solve a given optimization problem for emotion recognition.

Fitness function — Given the prototype formation and feature selection, we consider the minimization of the recognition error to be a suitable fitness measure.

Particles — The elements of a vector of a particle consist of the number of patterns and number of features for the optimal subsets of prototypes and features, as shown in Fig. 5. To solve the combinatorial problem, there can be different search spaces to choose from, considering that an optimal collection of prototypes and features can be represented in several different ways. Here, we adopt a representation scheme of the search space in the form of a (N+n)-dimensional unit hypercube (N is the number of patterns, whereas n denotes the number of features of the emotional dataset). The content of the particle is ranked in such a way that each value in this vector is associated with an index that the given value assumes within the ordered sequence of all values encountered in the vector. Considering the P% of all patterns and d% of all features, the first $P \times N$ ($0<P<1$) entries of the vector of the pattern search space are selected for the prototypes and the first $d \times n$ ($0<d<1$) entries of the vector of the feature search space are chosen as a core set of features. The employed vectors as a set of prototypes and a core set of features are concerned with defining the training consistent subset and forming the reduced feature space. This mechanism of the particle formation of the prototype and feature spaces is portrayed in Fig. 5.

Stop condition — There are two stopping conditions for recognition of the seven emotions: the algorithm terminates if

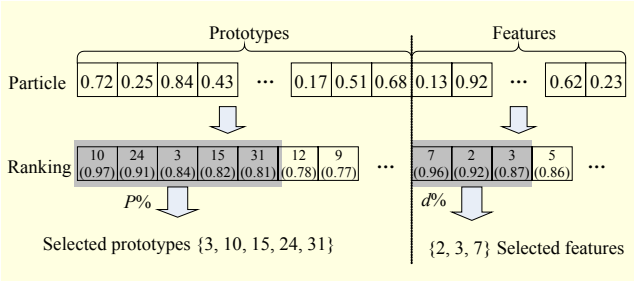

Fig. 5. Particle formation of selected prototypes and reduced feature.

the fitness function does not improve during the last 100 generations; otherwise, it terminates after 500 iterations. The size of the population is related to the dimensionality of the search space.

IV. Results of Emotion Recognition

The primary objective is to illustrate the recognition performance of the proposed recognizer. The two essential parameters that we use in the assessment of the performance of the prototype and feature selection are the percentage of features (denoted by d) forming the core of the reduced feature space, and the percentage of data forming the prototype set (P) optimized by the PSO. The results are reported for the testing data sets for various values of P and d.

The numeric values of the parameters of the PSO are either predetermined (considering some existing guidelines available in the literature [27] and [28]) or are selected through experimentation. More specifically, we use the following values of the parameters: a maximum number of generations of 500; a swarm size of 150; maximal velocity, v_{max}, of 20% of the range of corresponding variables; w_{min} of 0.4; w_{max} of 0.9; and acceleration constants c_1 and c_2 of 2.0. The inertia weight factor, w, is regarded as a linearly decreasing function of iterations (generations), as presented in (3).

The proposed methodology is applied to the recognition of the seven emotions. As mentioned in the previous section, the dataset has 730 instances with 28 features achieved through an analysis of the physiological signals, as shown in Table 2. The seven emotions are happiness, sadness, anger, fear, disgust, surprise, and stress.

The proposed prototype-based recognizer is evaluated using a tenfold validation (repeated random subsampling validation) for recognition of the seven emotions. The dataset is split into P% training and testing subsets, namely, P% of the whole patterns are selected randomly for training, and the remaining patterns are used for testing purposes. The results are reported by presenting the average and standard deviations of the recognition accuracy obtained over ten repetitions of the

Table 3. Accuracy of emotion recognition regarded as function of "d" and "P" for emotion dataset used.

d% (No. of F)	P%			AVG±STD over P
	30	50	70	
10 (3)	62.7 ± 1.41	88 ± 1.84	96.4 ± 1.13	82.3 ± 14.63
20 (6)	60.6 ± 3.25	85.1 ± 2.86	96.7 ± 1.78	80.8 ± 15.49
30 (8)	58.3 ± 3.89	79.5 ± 10.93	90.3 ± 6.36	76 ± 15.42
40 (11)	53.8 ± 3.86	69.5 ± 4.5	81 ± 4.21	68.1 ± 12.04
50 (14)	45.5 ± 4.15	54.1 ± 5.04	72.4 ± 4.39	57.3 ± 12.24
60 (17)	36.5 ± 4.44	46.6 ± 2.13	49.4 ± 8.18	44.2 ± 7.74
70 (20)	35.5 ± 1.92	39.9 ± 3.41	47.3 ± 3.74	40.9 ± 5.81
80 (22)	35.8 ± 3.1	39.9 ± 2.31	49.6 ± 1.93	41.8 ± 6.38
90 (25)	34.4 ± 2.53	40.7 ± 2	48.1 ± 4.42	41.1 ± 6.46
100 (28)	35.5 ± 1.86	39.4 ± 1.9	49.9 ± 2.15	41.6 ± 6.46

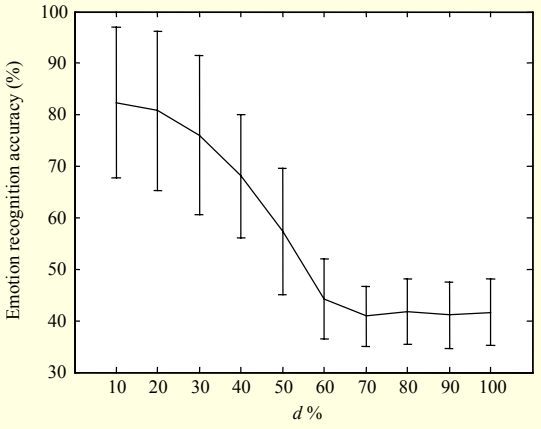

Fig. 6. Recognition accuracy regarded as function of d over P.

experiment for the test dataset. When reporting the results, we concentrate on a determination of the relationships between the collections of features and the obtained recognition rates. We also look at the optimal subsets of features constructed with the use of the proposed method.

For the dataset of seven emotions, the relationship among the percentage of features used in the PSO, the values of d, and the resulting recognition accuracy are presented in Table 3. Here, No. of F indicates the number of selected features for d% of all features, and AVG and STD indicate the average and standard deviations, respectively.

With an increase in the value of d, the recognition accuracy decreases substantially; in the case of $P = 30$%, it drops from 62.7 to 35.5 when increasing the number of features from 10% to 70%. A similar downward tendency occurs when dealing with any P% and considering the same increase in the

percentage of features. Conclusively, the use of all features diminishes the accuracy of recognition of the seven emotions. On the contrary, changes in the value of P have the opposite effect of changes in the value of d on the recognition rate. For any d%, the improvement in the recognition accuracy is about 40% over the value of P varying from 30% to 70%.

Figure 6 illustrates the recognition accuracy versus the percentage of the feature space used. The plot is the result of the average and standard deviations of all P and a tenfold run of the experiments. The results reveal notable dependencies related to the discriminatory characteristic of the feature space. There is an evident effect of the optimal subset of the feature space: a subset of the original features clearly leads to better recognition results when compared with the outcomes of the recognizer operating on the entire feature space. In all cases, we observe that there is an optimal number of features leading to the highest value of recognition accuracy. Depending on the value of d, the dimensionality of the reduced feature space varies between 3 and 6.

We report the number of occurrences of features for all experiments. This indicator becomes more illustrative and offers an interesting view of the suitability of the features when forming various reduced feature spaces and using different prototype set sizes. The results obtained in this case are illustrated in Fig. 7. The number of occurrences of a given feature is computed across all values of P and d. Interestingly, there are several dominant features, such as FFTnHF (feature 24), FFTLF/HFratio (feature 25), ARnLF (feature 26), ARnHF (feature 27), and ARLF/HFratio (feature 28). The values of meanPPG (feature 6), meanHR (feature 9), and CSI (feature 15) are of lowest relevance.

The prototypes are chosen by the PSO as a supervisor of the recognizer for each emotional category and given patterns are assigned into a category through these prototypes and selected features. In the case of $d = 20$%, namely, when the number of selected features is 6, we obtain a recognition accuracy of 80.8 for all P. Therefore, we can consider that the number of core features is six and features 23, 24, 25, 26, 27, and 28 are started in advance, as shown in Fig. 7. For $d = 30$% and $P = 70$%, an example set of prototypes (supervisor) and patterns (pupil) is shown over the specific features in Fig. 8.

As Table 4 demonstrates, the quality of the reduced feature space is quantified with the use of the recognition accuracy produced by the prototype-based recognizer on the testing set. The results are reported for selected scenarios of the experiments. First, the best recognition accuracy obtained across all combinations of d and P is given. For a prototype set size of 50% of the overall data, the recognition accuracy attains lower values, which is not surprising given that the sizes of the prototype sets for the best recognition accuracy are typically higher in the range of 70%. To assess the effect of the size of

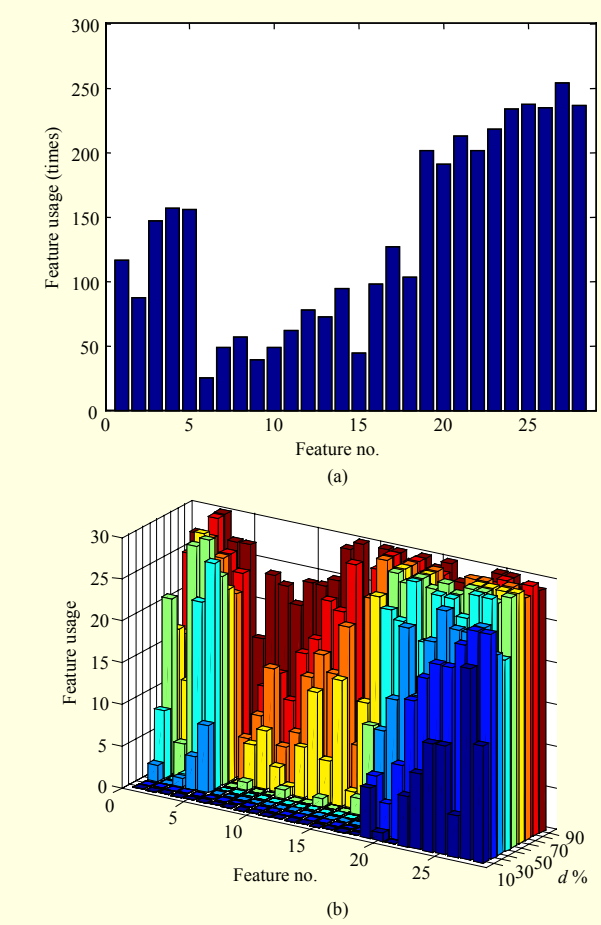

Fig. 7. Cumulative number of occurrences of individual features for seven types of emotional data: feature usage index over (a) all values of d and P, and (b) all values of P for each d.

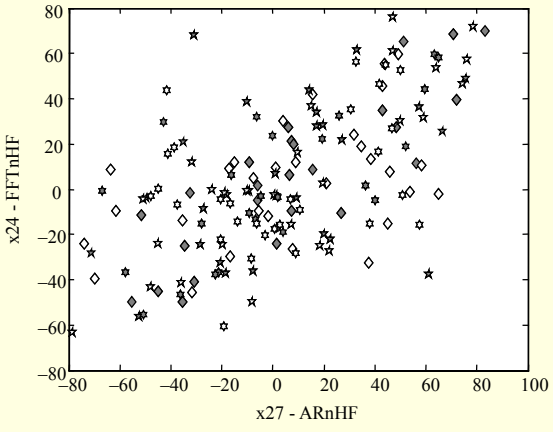

Fig. 8. Example set of prototypes (gray) and patterns (white) for sadness (♦), fear (★), and stress (✱), with features 24 and 27 ($d = 30$% and $P = 70$%).

Table 4. Comparison of recognition accuracy reported in experiments.

Method	Prototype-based recognizer		
Dimensionality of reduced space and reduction rate (%)	3 (89)	6 (79)	8 (71)
Max. accuracy (%) and assoc. value of P	96.4 ± 1.13 (70)	96.7 ± 01.78 (70)	90.3 ± 6.36 (60)
Accuracy (%) at P = 50%	88 ± 1.84	85.1 ± 2.86	79.5 ± 10.93
Avg. accuracy (%) over P	82.3 ± 14.63	80.8 ± 15.49	76 ± 15.42

Table 5. Emotion recognition accuracy of well-known models.

Method		Accuracy (%)
Classification and regression trees (CART)		21.7 ± 3.6
C4.5		16.1 ± 1.2
k-NN		45.3 ± 2.3
Fuzzy *k*-NN		41.2 ± 4.0
Principal component analysis (PCA)-*k*-NN		43.9 ± 3.4
FLD		20.3 ± 1.8
Quadratic discriminant analysis (QDF)		14.8 ± 2.8
NN		18.0 ± 1.0
Probabilistic NN (PNN)		16.3 ± 1.9
Radial basis function (RBF) networks		17.4 ± 1.2
SOM	Supervised	16.3 ± 1.9
	Unsupervised	15.4 ± 3.2
Naïve Bayes	General	17.8 ± 1.7
	Kernel density	17.9 ± 1.9
SVM		17.8 ± 3.1

the prototype set, we compute the average recognition taken over the range of sizes of the prototype sets used in the experiments. This provides us with a better view of the diversity of the results implied by the size of the prototype set.

In this study, we carry out a recognition of the seven emotions using some well-known machine-learning and artificial intelligence algorithms: CART, C4.5, *k*-NN, Fuzzy *k*-NN, PCA, FLD, QDF, NN, PNN, SOM, RBF networks, Naïve Bayes, and SVM [29]-[33]. These machine-learning algorithms are evaluated using ten repeated random subsampling validations. Seventy percent of all emotional patterns are selected randomly for the training, and the remaining patterns are used for testing purposes. The results shown in Table 5 are reported by presenting the average and standard deviations of the recognition accuracy obtained over ten repetitions of the experiment.

The values of the testing performance are good indicators of the generalization capabilities of the constructed models. When selecting a model, if only the approximation capability of the trained model is considered, the selected model has great recognition accuracy; however, it has a deteriorated generalization (prediction) capability and cannot apply to a real system. In particular, this is conspicuous in a nonlinear problem. Another important question also arises in this study regarding the selection of a proper structure of the emotion recognition. Even if the constructed model using a training set with physiological signals has 100% accuracy for the results of the recognition of the seven emotions for approximation capability, its generalization is 15% to 40%, as shown in Table 5. This means that some algorithms shown in the table are not useful as a recognizer for the recognition of the seven emotions. In addition, these results indicate that the given dataset generated by the physiological signals for the seven emotions has a high nonlinear characteristic and reflects the individual physiological property of each subject for the seven emotions.

The proposed method leads to better recognition results on the reduced feature space than those obtained from the others, as shown in Tables 4 and 5. We achieve substantially higher recognition accuracy with a lower dimensionality.

V. Conclusion

In this study, we discussed the acquisition of physiological signals using emotional stimuli and introduced a prototype-based recognizer with feature selection learned by a PSO mechanism for the recognition of seven emotions (happiness, sadness, anger, fear, disgust, surprise, and stress). The appropriateness and effectiveness of emotion stimuli were evaluated by assessors of the stimuli in the preliminary study, executed prior to the experiment conducted to evoke emotion. The results showed that the stimuli have a suitability of 93% and effectiveness of 9.5 points on average. Twenty-eight features were extracted using statistical and geometric approaches in the time and frequency domains from physiological signals, such as EDA, SKT, PPG, and ECG. For the recognition of the seven emotions using physiological signals, we proposed a prototype-based recognizer learned through PSO. PSO consists of a two-phase optimization process used to form a set of prototypes based on emotional patterns and a core set of features. These are a collection of the most meaningful and discriminating components, and their discriminatory capabilities emerge through their co-occurrence in the set. The proposed recognition scheme might be the simplest scheme that can be created in emotional pattern recognition. That the reduction of the feature space is related to

the corresponding recognition rate is evident. Therefore, the use of the prototype-based emotion recognizer becomes legitimate considering this recognition scheme, and the feature selection is justifiable considering the results of the feature reduction. While the experiment results provide sound evidence behind the selection process showing that the reduced feature spaces lead to better recognition results than those obtained in other methods, they are also quite revealing in showing that the reduction of the feature space can exhibit different levels of effectiveness. The proposed recognizer will lead to a better chance to recognize human emotions using physiological signals in the emotional interactions between humans and machines.

References

[1] J. Wagner, J. Kim, and E. Andre, "From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification," *IEEE Int. Conf. Multimedia Expo*, Amsterdam, July 2005, pp. 940-943.

[2] N. Sebe, I. Cohen, and T.S. Huang, "Multimodal Emotion Recognition," *Handbook of Pattern Recognition and Computer Vision*, Amsterdam: Publications of the Universiteit van Amsterdam, 2005.

[3] L. Li and J. Chen, "Emotion Recognition Using Physiological Signals," *ICAT*, 2006, pp. 437-446.

[4] E.L. Meerwijk and S.J. Weiss, "Toward a Unifying Definition of Psychological Pain," *J. Loss Trauma*, vol. 16, no. 5, 2011, pp. 402-412.

[5] S.D. Kreibig, "Autonomic Nervous System Activity in Emotion: A Review," *Bio. Psychology*, vol. 84, 2010, pp. 394-421.

[6] T. Verhoef et al., "Bio-sensing for Emotional Characterization without Word Labels," *13th Int. Conf. Human-Comput. Interaction: Ambient, Ubiquitous, Intell. Interaction*, San Diego, CA, USA, July 2009, pp. 693-702.

[7] C. Maaoui and A. Pruski, "Emotion Recognition through Physiological Signals for Human-Machine Communication," *Cutting Edge Robotics 2010*, Vedran Kordic, Ed., 2010, pp. 317-322.

[8] C.L. Stephens, I.C. Christie, and B.H. Friedman, "Autonomic Specificity of Basic Emotions: Evidence from Pattern Classification and Cluster Analysis," *Bio. Psychology*, vol. 84, 2010, pp. 463-473.

[9] E.-H. Jang et al., "Emotion Recognition by Machine Learning Algorithms Using Psychophysiological Signals," *Int. J. Eng. Industries*, vol. 3, no. 1, 2012, pp. 55-66.

[10] B.-J. Park et al., "Seven Emotion Recognition by Means of Particle Swarm Optimization on Physiological Signals," *Proc. 9th IEEE Int. Conf. Netw., Sensing, Control*, 2012, pp. 277-282.

[11] C.M. Bishop, *Neural Networks for Pattern Recognition*, New York: Oxford University Press Inc., 1995.

[12] J. Kennedy, "The Particle Swarm: Social Adaptation of Knowledge," *Proc. IEEE Int. Conf. Evolutionary Comput.*, 1997, pp. 303-308.

[13] K.E. Parsopoulos and M.N. Vrahatis, "On the Computation of All Global Minimizers through Particle Swarm Optimization," *IEEE Trans. Evolutionary Comput.*, vol. 8, 2004, pp. 211-224.

[14] C.L. Huang and C.J. Wang, "A GA-Based Feature Selection and Parameters Optimization for Support Vector Machines," *Expert Syst. Appl.*, vol. 31, 2006, pp. 231-240.

[15] H. Cheng et al., "Conditional Mutual Information-Based Feature Selection Analyzing for Synergy and Redundancy," *ETRI J.*, vol. 33, no. 2, Apr. 2011, pp. 210-218.

[16] J.X. Huang et al., "Feature-Based Relation Classification Using Quantified Relatedness Information," *ETRI J.*, vol. 32, no. 3, June 2010, pp. 482-485.

[17] D.W. Aha, D. Kibler, and M.K. Albert, "Instance-Based Learning Algorithms," *Machine Learning*, vol. 6, 1991, pp. 37-66.

[18] C. Gonzalez and V. Dutt, "Instance-Based Learning: Integrating Sampling and Repeated Decisions from Experience," *Psychological Rev.*, vol. 118, 2011, pp. 523-551.

[19] D. Palomba et al., "Cardiac Responses Associated with Affective Processing of Unpleasant Film Stimului," *Int. J. Psychophysiology*, vol. 36, 2000, pp. 45-57.

[20] J.J. Gross and R.W. Levenson, "Emotion Elicitation Using Films," *Cognition Emotion*, vol. 9, 1995, pp. 87-108.

[21] A. Gorini and G. Riva, "The Potential of Virtual Reality as Anxiety Management Tool: A Randomized Controlled Study in a Sample of Patients Affected by Generalized Anxiety Disorder," *Trials*, vol. 9, 2008, pp. 1-9.

[22] J. Arroyo-Palacios and D.M. Romano, "Towards a Standardization in the Use of Physiological Signals for Affective Recognition Systems," *Proc. Meas. Behavior*, Aug. 2008, pp. 121-124.

[23] M. Toichi et al., "The Influence of Psychotic States on the Autonomic Nervous System in Schizophrenia," *Int. J. Psychophysiology*, vol. 31, 1999, pp. 147-154.

[24] M.A. Abido, "Optimal Design of Power-System Stabilizers Using Particle Swarm Optimization," *IEEE Trans. Energy Conversion*, vol. 17, 2002, pp. 406-413.

[25] B. Liu et al., "Improved Particle Swarm Optimization Combined with Chaos," *Chaos, Solitons, Fractals*, vol. 25, 2005, pp. 1261-1271.

[26] A.E. Yilmaz and M. Kuzuoglu, "Calculation of Optimized Parameters of Rectangular Microstrip Patch Antenna Using Particle Swarm Optimization," *Microw. Opt. Technol. Lett.*, vol. 49, 2007, pp. 2905-2907.

[27] Z.L. Gaing, "A Particle Swarm Optimization Approach for Optimum Design of PID Controller in AVR System," *IEEE Trans. Energy Conversion*, vol. 19, 2004, pp. 384-391.

[28] R. Eberhart and Y. Shi, "Particle Swarm Optimization: Developments, Applications and Resources," *Proc. IEEE Int. Conf. Evolutionary Comput.*, 2001, pp. 81-86.
[29] W.-Y. Loh, "Classification and Regression Trees," *Wiley Interdisciplinary Rev: Data Mining Knowl. Discovery*, vol. 1, 2011, pp. 14-23.
[30] J.R. Quinlan, *C4.5: Programs for Machine Learning*, San Mateo, CA: Morgan Kaufmann, 1992.
[31] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, 2nd ed., New York: Wiley-Interscience, 2000.
[32] J.M. Keller, M.R. Gray, and J.A. Givens, Jr., "A Fuzzy K-Nearest Neighbor Algorithm," *IEEE Trans. Syst., Man, Cybern.*, vol. 15, no. 4, 1985, pp. 580-585.
[33] M.M Van Hulle, "Self-Organizing Maps," *Handbook of Natural Computing: Theory, Experiments, and Applications*, Springer-Verlag, 2010.

Byoung-Jun Park received his BS, MS, and PhD in control and instrumentation engineering from Wonkwang University, Iksan, Rep. of Korea, in 1998, 2000, and 2003, respectively. From 2005 to 2006, he was a postdoctoral fellow in the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, Canada. Since 2008, he has worked as a senior researcher at ETRI, Daejeon, Rep. of Korea. His research interests encompass computational intelligence, pattern recognition, granular and relational computing, and IT convergence.

Eun-Hye Jang received her BA, MA, and PhD in experimental and biological psychology from Chungnam National University, Daejeon, Rep. of Korea, in 2000, 2002, and 2009, respectively. Since May 2009, she has worked as a researcher and senior researcher at ETRI, Daejeon, Rep. of Korea. Her research interests include psychophysiology of emotion, emotion recognition, and IT-cognition/emotion convergence.

Myung-Ae Chung received her BS and MS from the Department of Chemistry at Ewah Womans University, Seoul, Rep. of Korea, in 1986 and 1988, respectively, and her PhD in polymer physical chemistry from the Clausthal University of Technology, Clausthal-Zellerfeld, Lower Saxony, Germany, in 1997. From 1997 to 1998, she was a postdoctoral researcher at the Clausthal University of Technology. From 1998 to 1999, she worked as a research fellow at Max Planck Institute for Polymer Research, Mainz, Germany. In 2000, she joined ETRI, Daejeon, Rep. of Korea, and she is currently the managing director of the IT Convergence Services Future Technology Research Department. Her research interests include ICT convergence technology, cognitive systems, biomimicry, creative ideas, and future technology.

Sang-Hyeob Kim received his MS and BS in material science from Chonbuk National University, Rep. of Korea, in 1984 and 1986, respectively, and his PhD in material science and engineering from Tohoku University, Japan, in 1994. From 1994 to 1997, he worked as a postdoctoral researcher at KRISS, Rep. of Korea. Since 2000, he has worked as a senior researcher and principal researcher at ETRI, Daejeon, Rep. of Korea. His research interests include organic/inorganic hybrid devices, IT-BT-NT-ET convergence devices, and the synthesis of nano-structured materials. He has authored or co-authored over 50 papers and holds 10 US patents as well as 30 Korea patents.