Improved Two-Phase Framework for Facial Emotion Recognition

Hyunjin Yoon, Sangwook Park, Yongkwi Lee, Mikyong Han, and Jong-Hyun Jang

Automatic emotion recognition based on facial cues, such as facial action units (AUs), has received huge attention in the last decade due to its wide variety of applications. Current computer-based automated two-phase facial emotion recognition procedures first detect AUs from input images and then infer target emotions from the detected AUs. However, more robust AU detection and AU-to-emotion mapping methods are required to deal with the error accumulation problem inherent in the multiphase scheme. Motivated by our key observation that a single AU detector does not perform equally well for all AUs, we propose a novel two-phase facial emotion recognition framework, where the presence of AUs is detected by group decisions of multiple AU detectors and a target emotion is inferred from the combined AU detection decisions. Our emotion recognition framework consists of three major components — multiple AU detection, AU detection fusion, and AU-to-emotion mapping. The experimental results on two real-world face databases demonstrate an improved performance over the previous two-phase method using a single AU detector in terms of both AU detection accuracy and correct emotion recognition rate.

Keywords: Automatic emotion recognition, facial action unit detection, decision fusion, facial expression analysis.
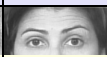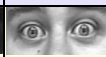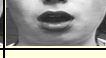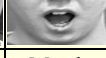
I. Introduction

Automatic recognition of human behaviors in general and emotions in particular has a wide variety of applications such as human–computer interaction, healthcare, computer-assisted learning, serious games, and security. Among the various channels of the human body that communicate emotions, non-verbal information, such as facial expressions, plays an important role in the analysis of human affective behaviors [1].

The Facial Action Coding System (FACS) proposed by Ekman and Friesen [1] is a method of describing facial movement based on the atomic activity of individual or groups of muscles called action units (AUs) (see Table 1). The FACS defines a total of 44 AUs, from which it is possible to represent nearly all human facial expressions by some combination thereof; this includes the facial expressions for the six basic emotions — anger, disgust, fear, happiness, sadness, and surprise. For example, the prototypical facial expression displaying "happiness" is configured in terms of AU6 (Check Raiser) and AU12 (Lip Corner Puller), according to the emotion prediction table featured in [1].

Automatic emotion recognition based on facial cues has been extensively exploited in the past [2]–[8], and related existing techniques can be roughly classified into either single-phase recognition — where emotions are directly recognized from face images — or two-phase recognition — where AUs are first detected and underlying emotions are then inferred from the detected AUs. The former often leads to a complex recognition model that consists of a larger number of variables, thus requiring longer training time and more training data. Detecting AUs prior to the occurrence of a facial emotion makes emotion recognition more interpretable by providing an explicit visual evidence for the recognized emotion.

Table 1. Visual examples of selected FACS AUs [1].

AU1	AU2	AU4	AU5	AU6	AU7
Inner Brow Raiser	Outer Brow Raiser	Brow Lowered	Upper Lid Raiser	Cheek Raiser	Lid Tightener
AU9	AU12	AU14	AU15	AU17	AU20
Nose Wrinkler	Lip Corner Puller	Dimpler	Lip Corner Depressor	Chin Raiser	Lip Stretcher
AU23	AU24	AU25	AU26	AU27	
Lip Tightener	Lip Pressor	Lips Part	Jaw Drop	Mouth Stretch	

A major issue with two-phase emotion recognition is that of the error accumulation problem, which, in fact, is inherent in any multiphase scheme. That is, the performance of emotion inference from automatically detected AUs is hindered by incorrectly identified or missed AUs. To attain recognition rates similar to that of a single-phase approach, a two-phase approach requires both robust AU detection and mapping of AUs to emotions to reduce any accumulated errors.

Automatic AU detection has been extensively explored and many good techniques have been developed [3]–[7]. However, none of them has achieved satisfactory performance for all considered AUs. According to a recent comparison of four state-of-the-art AU detection methods [4], those whose AU detectors used geometric facial features outperformed those who used texture features, on AUs whose activations are well characterized by morphological changes, such as AU1 and AU2. However, some of them fail in detecting AUs where there are distinct changes in skin texture, such as AU7 and AU15.

Motivated by our key observation that a single AU detector does not perform equally well for all AUs, we propose a new two-phase facial emotion recognition framework, where the presence of AUs is determined by group decisions of multiple independent AU detectors and a target emotion is inferred from the combined AU detection results. The proposed framework consists of three major components — multiple AU detection, AU detection fusion, and AU-to-emotion mapping. The emphasis of this research is on the second of these components, where we introduce two types of decision fusion methods called label-output fusion and probability-output fusion, which combines the individual decisions of the multiple AU detectors and presents the combined decision in a categorical and dimensional value, respectively. Given the automatically detected and fused AUs, emotion recognition can be performed by mapping the combined AU detection decision to six basic emotions through either a rule-based, longest-common-subsequence-based, or model-based AU-to-emotion mapping model.

To demonstrate the effectiveness of our proposed framework, the recognition performance of every possible combination of eight AU detection fusion methods and three mapping models (13 combinations in total) is evaluated and compared with that of previous two-phase emotion recognition, which employs a single AU detector over two real-world face image databases.

The rest of this paper is organized as follows. Section II reviews previous work on automatic methods for emotion-specified facial expression recognition. The proposed two-phase emotion recognition framework is presented in detail in Section III. Section IV provides experimental evaluations, and Section V concludes the paper.

II. Related Work

This section describes related work on emotion recognition based on facial cues. Detailed reviews related to this field of work can be found in [8].

1. Single-Phase Facial Emotion Recognition

In single-phase emotion recognition, emotions are directly recognized from input face images. Previous efforts have emphasized types of facial features and classifiers. Cohen and others [9] employed a tree-augmented naïve Bayes classifier to learn the dependencies among different motion features and hidden Markov models (HMMs) to recognize the emotions from these correlated motion features. Barlett and others [10] empirically demonstrated that the best results for classifying facial expressions into basic emotions are achieved by using multiclass support vector machines (SVMs) as classifiers and feature selection by AdaBoost.

Recent single-phase methods consider the relationship between FACS's AUs and emotions to improve the performance of emotion recognition. Chang and others [11] employed emotion-related facial AUs as partially observed hidden state variables in their graphical model based on hidden conditional random fields, and demonstrated that knowing the AUs provides useful evidence for distinguishing emotions. Zhang and Ji [12] exploited the dependencies between AUs and basic emotions, and established a Bayesian network (BN) model consisting of three layers — classification layer, AU layer, and sensory data layer — to classify input images into six basic emotions.

Direct emotion recognition from facial features attains better

recognition rates than a two-phase approach. However, this often leads to a complex recognition model that consists of a larger number of variables and requires longer training time as well as more training data.

2. Two-Phase Facial Emotion Recognition

Two-phase emotion recognition consists of two parts — AU detection and AU-to-emotion mapping. Previous work on automatic detection of facial AUs has focused mainly on the types of facial features and classifiers similar to those found in one-phase emotion recognition. Donato and others [4] empirically compared various representations of face images and demonstrated that Gabor wavelet features and independent component analysis are useful for classifying facial actions. Bartlett and others [10] applied SVMs to the Gabor wavelet coefficients of a face image to detect AUs. Valsta and Pantic [6] combined SVMs and HMMs to create AU classifiers that can incorporate the temporal dynamics of AU activation, and demonstrated that AU detectors employing such classifiers outperform an SVM-only approach, for many AUs. Li and others [3] proposed a data-free prior model for facial AU detection that generalizes to new databases. Although many good techniques have been developed for automatic detection of AUs, none of them has achieved satisfactory performance for all considered AUs.

The mapping of AUs to emotions has also been explored in the past. A few deterministic rules that map facial AUs to emotions have been developed by exploiting the linguistic description of emotions in terms of AUs provided by domain experts [13]–[14]. Valstar and Pantic [15] have formulated a set of mapping rules based on emotional FACS and also used artificial neutral networks (ANNs) to map AUs to six basic emotions. Alternatively, Velusamy and others [16] derived most discriminant AUs for each emotion and inferred an underlying emotion by comparing detected AUs with the selected discriminant AUs of each emotion using the longest common subsequence (LCS) distance.

While most rule-based mapping methods use strict matching, an LCS-based mapping allows partial matching, to make an AU-to-emotion mapping robust to false positives and misses among automatically detected AUs.

III. Proposed Method

In this section, we present our two-phase facial emotion recognition framework, which decides the presence of AUs by group decisions of multiple AU detectors and infers a target basic emotion from a set of combined AU detection results. The proposed framework consists of three major components
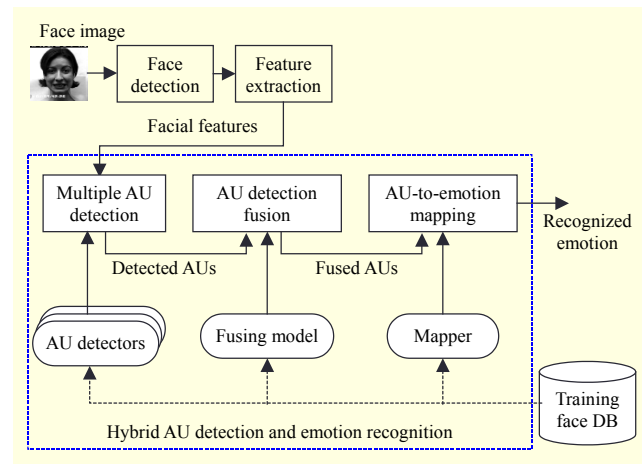
Fig. 1. Flow of our emotion recognition framework.

— multiple AU detection, AU detection fusion, and AU-to-emotion mapping. The first two components compose the AU detection phase and the last constitutes the second phase of the two-phase emotion recognition scheme.

Given a set of training face images fully labelled with the expected AUs and emotions, our framework first trains models with the training data for the three components, named AU detector, fusing model, and mapper, respectively. Once the models are obtained from the training data, the proposed framework takes the facial features extracted from the detected face region in an unknown input face image; detects the presence of target AUs using the trained multiple AU detectors; combines the individual decisions of the multiple AU detectors by the fusing model to decide a final decision on the AU presence; and infers the emotions from the fused AUs using the AU-to-emotion mapper. Figure 1 shows the flow of our two-phase emotion recognition framework. A detailed description of each part is given in the following sections.

1. Multiple AU Detection

Our two-phase framework employs multiple AU detectors to detect the presence, or absence, of target AUs with group decisions. Based on our detailed review on the linguistic description of emotions in terms of FACS AUs [13]–[14] and available face databases coded with AUs and emotions [14], [17], we chose 17 AUs that are found to be relevant to the recognition of six basic emotions (see the selected AUs in Table 1). For each of the 17 selected AUs, multiple AU detectors are individually trained with different detection methods or different portions of training data. Such AU detectors take various types of facial features as the description of an input face image and recognize the presence (or absence) of target AUs.

Classifiers such as ANN, SVM, boosting, HMM, and dynamic Bayesian networks (DBNs) as well as facial features such as grayscale pixels, edges, and appearance descriptors can be employed to model an individual AU detector. However, it is assumed that the output of such a detector is presented in a binary form representing the complete presence (or absence) of a target AU.

2. AU Detection Fusion

AU detection fusion can be considered as the process by which individual decisions of multiple AU detectors are combined to produce a final decision on the presence of target AUs for the subsequent AU-to-emotion mapping phase.

Let d_k be the AU detector that detects the presence or absence of a certain AU, where $k = 1, \dots, K$; here, K is the total number of AU detectors. Given an input face image, x, $d_k(x) = \hat{y}_k$ means AU detector d_k assigns the input x to value $\hat{y}_k$, where $\hat{y}_k = 1$ if the presence of the target AU is detected from the input x; otherwise, $\hat{y}_k = 0$. Then, the AU detection fusion can be formulated as follows: given outputs of multiple AU detectors for a target AU $d_1(x) = \hat{y}_1, d_2(x) = \hat{y}_2, \dots,$ $d_K(x) = \hat{y}_K$, the goal is to determine $D(x) = \hat{y}$, where D is the fusing function that combines the outputs of the multiple AU detectors and assigns x to a value $\hat{y} \in [0,1] = \{\hat{y} \in \mathbb{R} \mid 0 \le \hat{y} \le 1\}$. The output of the fusing function can be interpreted as the class membership probability that reflects the uncertainty with which the given face image x can be assigned to the target AU class. The closer to one the output value is, the more likely the target AU is presented in the input face image.

According to the output types, the fusing function is further divided into either *probability-output* fusion, where the fusing function assigns any real number between zero and one, including both or *label-output* fusion, where the outputs are limited to only two endpoints (that is, $\hat{y} \in \{0,1\}$) representing the complete absence or presence of the target AU, respectively.

A. Label-Output Fusion

Since the outputs of individual AU detectors are binary numbers, the label-output fusion method can obtain a decision via a voting-based scheme. We propose five label-output fusion methods by adopting conventional voting schemes such as majority, unanimous, and weighted-majority voting.

First, a majority vote–based label-output fusion method determines a final decision by selecting the decision that more than half of the individual AU detectors agree on. Thus, the resulting fusing function for this method, D_m, can be formulated as follows:

$$D_\mathrm{m}(x) = \hat{y} = \begin{cases} 1 & \sum_{k=1}^{K} \hat{y}_k > K/2, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Alternatively, a unanimous vote–based label-output fusion method determines the presence of a target AU only when all individual AU detectors agree on the presence of the target AU. Therefore, the resulting unanimous vote–based fusing function, D_u, can be defined as follows:

$$D_\mathrm{u}(x) = \hat{y} = \begin{cases} 1 & \sum_{k=1}^{K} \hat{y}_k = K, \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Both the majority and unanimous vote–based label-output fusion methods are straightforward in that they do not need any training. Also, as can be seen in (1) and (2), they treat the outputs of individual AU detectors with equal weight assuming no *a-priori* knowledge on the behaviors of individual AU detectors. However, when *a-priori* information about the quality of individual AU detectors is available, better fusion methods can be explored.

Our weighted majority vote–based fusion method exploits the performance of individual AU detectors for known face images to derive the weights of multiple AU detectors and then a final decision is made by taking the decision that has higher weighted votes. The weighted majority vote–based fusion function D_w is thus formulated as follows:

$$D_\mathrm{w}(x) = \hat{y} = \begin{cases} 1 & \sum_{k=1;\hat{y}_k=1}^{K} w_k > \sum_{k=1;\hat{y}_k=0}^{K} w_k, \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

where w_k is the weight on AU detector $d_k(x) = \hat{y}_k$, where $k = 1, \dots, K$. The weights are iteratively determined by exploiting the performance of individual AU detectors over those training face images that are fully labelled with the expected presence of the target AU. Let $S = \{(x^i, y^i) \mid 1 \le i \le N\}$ be the training face image set, where x^i is the face image, y^i is the expected AU label of the image representing the presence or absence of the target AU, and N is the total number of training images. Initially, the weight w_k is set to 1 for all K AU detectors. For each pair (x^i, y^i), the decisions of individual AU detectors are obtained and the combined decision $\hat{y}^i$ is determined by the weighted vote in (3). Then, the weights of the AU detectors that incorrectly assign a value different from the given expected value y^i are decreased to $\beta \cdot w_k$, where $\beta \in [0,1)$. As a result, the individual AU detectors that correctly detect the presence or absence of the target AU receive higher weightings. A detailed description on weighted majority voting can be found in [18].

It is well known that the best performance of weighted majority vote–based fusion is bounded by some small constant fraction of the best performance among the individual AU

Training image set	Results of three AU detectors	BKS look-up table		
		$c_{\hat{y}_1,\hat{y}_2,\hat{y}_3}$	$y=1$	$y=0$
$(x^1, y^1=1)$	$(d_1^1=1, d_2^1=0, d_3^1=1)$	$c_{0,0,0}$		
$(x^2, y^2=1)$	$(d_1^2=1, d_2^2=0, d_3^2=1)$	$c_{0,0,1}$		
$(x^3, y^3=1)$	$(d_1^3=0, d_2^3=0, d_3^3=1)$	$c_{0,1,0}$		
$(x^4, y^4=1)$	$(d_1^4=1, d_2^4=0, d_3^4=1)$	$c_{0,1,1}$		
$(x^5, y^5=1)$	$(d_1^5=1, d_2^5=0, d_3^5=1)$	$c_{1,0,0}$		
$(x^6, y^6=0)$	$(d_1^6=1, d_2^6=0, d_3^6=1)$	$c_{1,0,1}$	3	1
…	…	$c_{1,1,0}$		
$(x^N, y^N=0)$	$(d_1^N=1, d_2^N=0, d_3^N=0)$	$c_{1,1,1}$		

Fig. 2. Example of BKS look-up table built with results of three AU detectors on training face images set.

detectors. This can be a problem when the performance of the best AU detector is not satisfactory. Therefore, we employ a randomized weighted majority algorithm [18] to mitigate this dependence issue of the straightforward weighted majority voting algorithm.

Our randomized weighted majority vote–based fusion method computes the fraction of the multiple AU detectors detecting the presence or absence of the target AU using their weights and determines the final decision by randomly predicting according to that fraction. The associated fusing function, D_r, can be formally described as follows:

$$D_r(x) = \hat{y} = \begin{cases} 1 & \text{with probability} \sum_{k=1; \hat{y}_k=1}^{K} w_k / W, \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

where $W = \sum_{k=1; \hat{y}_k=1}^{K} w_k + \sum_{k=1; \hat{y}_k=0}^{K} w_k$. It is proven that the number of mistakes the randomized weighted majority vote–based fusion is going to make is halved due to the introduced randomization to the deterministic fusion in (5) [18].

The last label-output fusion method adopts the behavior knowledge space (BKS) method and organizes the given knowledge on the empirical performance of multiple AU detectors on the training data. The BKS method constructs a look-up table consisting of cells, each of which corresponds to every possible combination of multiple AU detector outputs. Since only a binary classifier is considered for an individual AU detector, our BKS look-up table contains 2^K cells. Each cell contains two numbers; that is, the total number of training face images assigned to the corresponding output combination and known to have the target AU or not, respectively. Intuitively, the BKS look-up table maintains the training data counts of every possible combination of individual decisions of multiple AU detectors and the combined final decision.

Figure 2 illustrates an example of such a BKS look-up table, constructed with the detection results of three AU detectors on a training face image set. Since three AU detectors are used in the example, the BKS look-up table is composed of eight cells representing every possible output combination of three AU detectors (see the first column of the look-up table in Fig. 2). Then, each cell is filled by two numbers representing the total number of training samples detected as the corresponding output combination and originally labeled with the expected AU value $y=1$ and $y=0$, respectively.

Once the look-up table is constructed, our BKS-based label-output fusing function D_k determines the final decision given an unknown face image x by the following decision rule:

$$D_k(x) = \hat{y} = \begin{cases} 1 & c_{\hat{y}_1,\hat{y}_2,...,\hat{y}_K}(y=1) > c_{\hat{y}_1,\hat{y}_2,...,\hat{y}_K}(y=0), \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

where $c_{\hat{y}_1,\hat{y}_2,...,\hat{y}_K}(y=1)$ and $c_{\hat{y}_1,\hat{y}_2,...,\hat{y}_K}(y=0)$ are the two counts stored in the cell $c_{\hat{y}_1,\hat{y}_2,...,\hat{y}_K}$, representing the number of training images labelled with the expected AU value $y=1$ and $y=0$, respectively. For example, if an unknown image x is detected as $d_1(x) = \hat{y}_1 = 1, d_2(x) = \hat{y}_2 = 0, d_3(x) = \hat{y}_3 = 1$ by the three AU detectors in Fig. 2, then it is finally assigned to the AU value $\hat{y} = 1$ since $c_{\hat{y}_1,\hat{y}_2,\hat{y}_3}(y=1) = 3 > c_{\hat{y}_1,\hat{y}_2,\hat{y}_3}(y=0) = 1$ according to the BKS look-up table in Fig. 2 and the decision rule in (5). Further details on the original BKS method can be found in [19].

B. Probability-Output Fusion

The probability-output fusion method produces a combined decision that ranges in degree between 0 and 1. Since the outputs of the AU detection fusion are fed into the subsequent AU-to-emotion mapping, obtaining a dimensional value can be more suitable in terms of information loss. To this end, three probability-output fusion methods are proposed; the first two adapt the randomized weighted majority vote–based label-output fusing function in (4) and the BKS-based decision rule in (5), respectively, and the last one adopts BNs to derive a probability estimate for the combined decision.

First, our weight-based probability-output fusion method employs the probability obtained in the randomized weighted majority vote–based label-output fusion method defined in (4). Let w_k be the weight on AU detector $d_k(x) = \hat{y}_k$ for $k = 1, \dots, K$ and $\sum_{k=1; \hat{y}_k=1}^{K} w_k$ be the total weight of AU detectors that detect the presence of the target AU (that is, $\hat{y}_k = 1$). Then, the weight-based probability-output fusing function D_{wp} estimates the probability with which the target AU is present in an input face image x as follows:

$$D_{wp}(x) = \hat{y} = \sum_{k=1; \hat{y}_k=1}^{K} w_k / W, \quad (6)$$

where $W = \sum_{k=1; \hat{y}_k=1}^{K} w_k + \sum_{k=1; \hat{y}_k=0}^{K} w_k$.

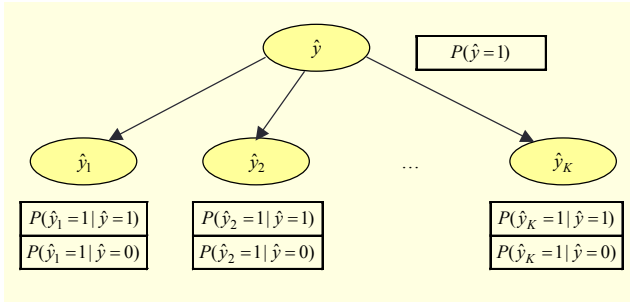The second probability-output fusion method is derived from

Fig. 3. Structure and parameters of BN used by D_{bn}.

the data stored in the BKS look-up table. Given the individual decisions of multiple AU detectors $d_1(x) = \hat{y}_1, d_2(x) = \hat{y}_2, \ldots, d_K(x) = \hat{y}_K$, the BKS-based probability-output fusing function D_{kp} decides the combined decision with the two numbers stored in the cell $c_{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_K}$ of the BKS look-up table as follows:

$$D_{kp}(x) = \hat{y} = c_{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_K}(y=1)/C, \quad (7)$$

where $C = \sum_{i \in \{0,1\}} c_{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_K}(y=i)$.

Our last probability-output fusion method employs BNs to represent the probabilistic relationship between the individual decisions of multiple AU detectors and the combined decision on the presence of the target AU. A BN is a directed acyclic graph that represents a joint probability distribution among a set of random variables [20]. Figure 3 shows the structure of the constructed BN and the parameters that need to be specified. The BN consists of a set of random variables representing the individual decisions of multiple AU detectors and the final decision of the fusing function and the conditional dependences among the variables. The BN-based probability-output fusing function D_{bn} decides the probability of the presence of the target AU by calculating the conditional probability as follows:

$$
\begin{aligned}
D_{bn}(x) = \hat{y} &= P\left(\hat{y}=1 \mid \hat{y}_1, \hat{y}_2, \ldots, \hat{y}_K\right) \\
&= \frac{P\left(\hat{y}=1, \hat{y}_1, \hat{y}_2, \ldots, \hat{y}_K\right)}{P\left(\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_K\right)} \\
&= \frac{P(\hat{y}=1)P(\hat{y}_1 \mid \hat{y}=1)P(\hat{y}_2 \mid \hat{y}=1)\cdots P(\hat{y}_K \mid \hat{y}=1)}{\sum_{\hat{y} \in \{0,1\}} P(\hat{y})P(\hat{y}_1 \mid \hat{y})P(\hat{y}_2 \mid \hat{y})\cdots P(\hat{y}_K \mid \hat{y})}.
\end{aligned}
$$
$$(8)$$

The conditional probability distribution (CPD) for each node given its parents and the prior probability of parent node in (8) is the parameters of our BN. These parameters are estimated from the decisions of multiple AU detectors on the training face images and their original AU values representing the presence or absence of a target AU by maximizing the

likelihood of the training data. Further details on BN learning and inference can be found in [20].

3. AU-to-Emotion Mapping

AU-to-emotion mapping aims to infer a target emotion from detected AUs. Detected AUs are represented by either categorical or dimensional values depending on the AU fusion method. Therefore, the mapping method needs to deal with detected AUs represented by both categorical and dimensional values. We present three AU-to-emotion mapping methods to recognize an emotion displayed in an input face image.

A. Rule-Based Mapping

Several rules that classify facial actions into the basic emotion categories have been acquired in a straightforward manner from the linguistic descriptions of the prototypic facial expressions given by Ekman and Friesen [1]. A set of deterministic AU-to-emotion mapping rules is formulated from the emotion descriptions in terms of AUs present in [13] and [14], which is presented in Table 2.

Although some rules deal with the degree of AU presence, most of them only consider whether the target AUs are present or not to map to emotions. Therefore, a rule-based mapping is suitable for inferring emotions from the combined AU detection obtained by the label-output fusion method in our two-phase facial emotion recognition framework.

B. LCS-Based Mapping

A heuristic rule-based mapping method can suffer from the noisy outputs of the preceding automatic AU detection and fusion. A more robust mapping that allows partial matching using the well-known LCS distance is recently proposed by Velusamy and others [16] and is employed for our AU-to-emotion mapping.

The LCS-based mapping method consists of two parts. First,

Table 2. AU-to-emotion mapping rules.

Emotion	Description in terms of AUs
Anger	AU23 ∧ AU24 ∧ ¬AU20
Disgust	(AU9 ∨ AU10) ∧ ¬AU20
Fear	(AU1 ∧ AU2 ∧ AU4) ∨ AU20
Joy	(AU10 ∨ ¬AU9) ∧ ¬AU20
Sadness	(((AU1 ∧ AU4) ∨ AU6) ∧ AU15)) ∧ (¬AU9 ∨ ¬AU10) ∧ ¬AU20
Surprise	(AU1 ∧ AU2) ∨ AU5 ∧ ¬AU4 ∧ (¬AU9 ∨ ¬AU10) ∧ ¬AU20

a set of *discriminant* AUs are determined per emotion using a concept called discriminant power [16]. The discriminant power of an AU for an emotion is defined as follows:

$$H_{ij} = P\left(\mathrm{AU}_j | \omega_i\right) - P\left(\mathrm{AU}_j | \overline{\omega}_i\right), \qquad (9)$$

where $(\mathrm{AU}_j | \omega_i)$ is the probability of the action unit AU_j's presence given that the emotion ω_i has occurred, and $P(\mathrm{AU}_j | \overline{\omega}_i)$ is the probability of the action unit AU_j's presence given that the emotion ω_i has not occurred. For each emotion, the top five highly discriminant AUs that have the largest positive discriminant power are selected as discriminant AUs of that emotion.

Once the discriminant AUs are determined, the emotion to be recognized, in a new image, is determined by comparing the detected AUs (from the new image) and the predetermined discriminant AUs of every emotion by the LCS similarity measure and by choosing the emotion with the LCS as the answer. For example, suppose that the discriminant AUs of anger and happiness are determined as {AU2, AU4, AU7, AU17, AU23} and {AU1, AU5, AU6, AU12, AU26}, respectively, according to the discriminant power defined in (9). If the detected AUs are (AU1, AU6, AU7) for an input image, then it is now mapped to the emotion happiness, although some AUs, such as AU5, AU12, and AU26, are missing in the detected AUs and AU7 is not part of the discriminant AUs of happiness. Similar to the rule-based mapping, the LCS-based mapping takes a binary input and can thus only be paired with label-output fusion methods.

C. Model-Based Mapping

A model-based mapping method is proposed to deal with the dimensional outputs of our probability-output fusion methods. SVM is employed as the mathematical model that takes as an input the list of detected AUs presented in ranges between 0 and 1 and classifies them into emotions. Six binary SVM classifiers are trained to model each of the six basic emotions and the outputs of the trained six SVM classifiers are transformed to an emotion probability using the softmax function to determine one single recognized emotion.

Let $\mathbf{O} = \left(o_1, o_2, \ldots, o_6\right)^{\mathrm{T}}$ be the output vector of the six trained SVM classifiers given a list of combined decisions on the presence of all considered AUs, where o_i is the distance of the given input from the decision hyperplane of the ith SVM classifier. Then, the recognized emotion $\hat{\omega}$ is determined as follows:

$$\hat{\omega} = \operatorname*{argmax}_{i \in \{1,2,\ldots,6\}} e^{o_i} / \sum_{j=1}^{6} e^{o_j}. \qquad (10)$$

IV. Experiments

The performance of the proposed two-phase facial emotion recognition framework consisting of multiple AU detection, AU detection fusion, and AU-to-emotion mapping is evaluated over two widely used face databases labelled with true AUs and emotions.

1. Database

The Extended Cohn-Kanade (CK+) database [13] contains 593 image sequences of posed and non-posed spontaneous expressions with frontal face. The final frame in each image sequence contains the emotion-specified facial expression at the apex state and is coded with 39 FACS AUs and 6 emotions. This dataset is widely used as a benchmark database to evaluate facial AU detection and emotion-specified facial expression recognition.

The ISL Facial Expression Database (ISL) [17] consists of 42 image sequences obtained from 10 subjects. The image sequences contain several posed facial expressions from neutral, through a series of onset, apex, and offset phases and back again to a neutral state with nearly frontal. Only the training portion of the database composed of 28 image sequences of 6 subjects is used in our experiments because they are coded with 16 FACS AUs, whereas the testing portion is coded with only 4 AUs. Furthermore, those face images where the target AU is present in low intensity are excluded, which results in 2,083 face image frames.

2. Results of Multiple AU Detection and Fusion

Three AU detectors are used for multiple AU detection. Two of them are trained individually from the CK+ and ISL databases and named CKD and ISLD, respectively. Each database is randomly partition into five folds, where four folds are used for the training and the remaining fold is left out for the validation. Of the 593 CK+ face images, 531 are used to train the CKD, and 1,666 out of the 2,083 face images are used for training the ISLD.

From the training face images, frontal human faces with a head rotation of no more than ±10° are first detected and the detected faces are then rescaled to the size of 64 × 64 pixels. A set of Gabor filters at eight orientations and seven spatial frequencies (2:16 pixels per cycle at 1/2 octave steps) are applied on each normalized face to extract a feature vector of 64 × 64 × 56 dimensions. Given the facial feature vectors, AdaSVM [4] is then employed to detect a target AU present in the image features while simultaneously reducing the dimension of the feature vector. The AdaSVM employs Adaboost as the feature selection method and SVM as the

Table 3. AU detection results of individual AU detectors in terms of detection accuracy rate (unit: %).

DB	Detector	AU1	AU2	AU4	AU5	AU6	AU7	AU9	AU12	AU14	AU15	AU17	AU20	AU23	AU24	AU25	AU26	AU27
CK+	CKD	**90.16**	**91.80**	**75.41**	**85.25**	73.77	70.49	**96.72**	**95.08**	**96.72**	85.25	80.33	**88.52**	83.61	**85.25**	**85.25**	93.44	**95.08**
	ISLD	72.13	75.41	62.30	70.49	**75.41**	62.30	80.33	73.77	N/A	65.57	67.21	N/A	**83.61**	**85.25**	68.85	**95.08**	77.05
	LAUD	68.85	73.77	65.57	70.49	67.21	**75.41**	80.33	93.44	**96.72**	**90.16**	N/A	**88.52**	N/A	N/A	72.13	N/A	75.41
ISL	CKD	51.56	85.13	60.67	61.39	84.41	66.19	94.72	84.17	N/A	72.18	60.43	N/A	71.22	72.18	51.56	**81.06**	**93.76**
	ISLD	**100**	**99.52**	**99.52**	**98.32**	**98.56**	**99.52**	**100**	**100**	N/A	**97.60**	**99.28**	N/A	**97.84**	**99.04**	**99.52**	79.38	83.33
	LAUD	58.03	71.22	58.03	68.82	84.89	59.23	89.69	81.06	N/A	76.26	N/A	N/A	N/A	N/A	39.09	N/A	82.73

weak classifier.

Although the CK+ database is coded with 39 AUs, some AUs appear too infrequently. As stated in Section III-1, based on our detailed review on the linguistic description of emotions in terms of FACS AUs [13]–[14] and the occurring frequency of AUs in CK+ and ISL, we chose 17 AUs that are found to be relevant to the recognition of six basic emotions as in Table 1. Then, a single AdaSVM classifier is trained with the training portion of the CK+ database to detect the presence or absence of each of the 17 AUs and the resulting 17 AdaSVM classifiers compose the CKD. For the ISL database coded with 16 AUs, 15 AUs except AU45 (that is, eye closed) that are overlapped with the chosen 17 AUs are finally selected. The ISLD detector consists of the 15 AdaSVM classifiers trained with the training portion of the ISL database.

For the third AU detector, we employ a local binary pattern (LBP)–based detector named LAUD [7]. While the CKD and ISLD detectors use the well-known Gabor filter features, the LAUD detector exploits LBP features as the face description. The LBP features have been successfully applied to face recognition and have been recently extended to facial expression recognition [9]. The LAUD detector also adopts SVM as classifiers for AU detection and its implementation comes with 14 SVMs trained from the MMI Facial Expression Database [21]. Similarly, the AU45 detector of LAUD is excluded in our experiments, since AU45 (that is, blink) is irrelevant to emotion description. As a result, the LAUD detector used in the experiments consists of 13 LBP-based SVM classifiers.

Table 3 shows the AU detection results of individual AU detectors on the CK+ and ISL databases in terms of detection accuracy rates. Since four folds of the two databases are used to train the CKD and ISLD detectors, the performance comparison of three AU detectors are conducted on the remaining one fold of data. The best AU detection performance among three AU detectors is marked bold in the table. As expected, all three AU detectors do not perform equally well for all the target AUs. For example, the CKD detector trained with the CK+ database performs better for most of the AUs of the CK+ database than the other two detectors. However, IDSL and LAUD perform better on some AUs (that is, AU6, AU23, AU26, AU7, and AU15) although they are obtained with totally different databases. On the ISL database, the ISLD detector detects better most of the AUs except a few AUs such as AU26 and AU27 where CKD performs better. These experimental results confirm our observation that a single AU detector does not perform equally well for all AUs.

The individual decisions obtained by the three AU detectors are then combined with the proposed eight AU-detection fusion methods to finally determine the combined decision on the presence of a target AU. Table 4 shows the performance of the proposed label-output fusion methods on the CK+ and ISL databases. The detection accuracy rate is again used to evaluate the binary outputs of the label-output fusion methods. As expected, the fusion methods that exploit the *a-priori* information about the quality of individual AU detectors (D_w, D_r, and D_k) outperform the ones that treat the individual AU detectors with equal weights (D_m and D_u). The performance of the weighted majority vote–based fusion is bounded by the best performance of the three individual AU detectors. Only a marginal improvement is obtained by the randomized weighted majority vote–based fusion on AU6 and AU23 in the CK+ database.

To evaluate the probability outputs of the three probability-output fusion methods, the log-loss function defined as

$$-\frac{1}{N} \sum_{i=1}^{N} \left[y^i \log(\hat{y}^i) + (1 - y^i) \log(1 - \hat{y}^i) \right], \text{ where } y^i \in \{0,1\}$$

is the true AU label and $\hat{y}^i = P(y^i = 1)$ is the probability estimate for N number of face images, is employed as an error metric. Note that the probability estimate $\hat{y}^i$ is bounded from the extremes (that is, 0 and 1) by a small value to prevent an infinite error. Table 5 shows the evaluation results on the three probability-output fusion methods in terms of the log loss. Overall the BKS-based (D_{kp}) and the BN-based (D_{bn}) fusion methods attain fewer detection errors than the weight-based fusion method (D_{wp}) over most of the AUs, which implies that

Table 4. Combined AU detection results of five label-output fusion methods in terms of detection accuracy rate (unit: %).

DB	Fusion	AU1	AU2	AU4	AU5	AU6	AU7	AU9	AU12	AU14	AU15	AU17	AU20	AU23	AU24	AU25	AU26	AU27
CK+	D_m	81.97	86.89	68.85	80.33	70.49	70.49	81.97	**96.72**	**96.72**	86.89	75.41	86.89	83.61	**85.25**	83.61	90.16	81.97
	D_u	60.66	72.13	65.57	65.57	67.21	**73.77**	80.33	86.89	**96.72**	**90.16**	72.13	**90.16**	83.61	**85.25**	62.30	**98.36**	75.41
	D_w	**90.16**	**91.80**	**75.41**	**85.25**	73.77	70.49	**96.72**	95.08	**96.72**	85.25	80.33	88.52	83.61	**85.25**	**85.25**	93.44	**95.08**
	D_r	**90.16**	**91.80**	**75.41**	81.97	**75.41**	70.49	**96.72**	95.08	**96.72**	85.25	80.33	88.52	**85.25**	**85.25**	**85.25**	93.44	**95.08**
	D_k	**90.16**	**91.80**	**75.41**	**85.25**	73.77	70.49	**96.72**	95.08	**96.72**	85.25	80.33	88.52	83.61	**85.25**	**85.25**	93.44	**95.08**
ISL	D_m	89.21	92.57	87.05	84.65	84.89	69.54	94.72	87.29	N/A	80.10	78.18	N/A	81.53	78.42	61.15	64..75	93.56
	D_u	71.70	75.54	73.14	71.94	84.89	58.99	89.69	85.37	N/A	79.14	81.53	N/A	87.53	92.81	92.81	95.68	85.61
	D_w	**100**	**100**	**99.52**	**98.32**	**98.56**	**99.52**	**100**	**100**	N/A	**97.60**	**99.28**	N/A	**97.84**	**99.04**	**99.52**	81.06	**93.76**
	D_r	**100**	**100**	**99.52**	**98.32**	**98.56**	**99.52**	**100**	**100**	N/A	**97.60**	**99.28**	N/A	**97.84**	**99.04**	**99.52**	80.58	**93.76**
	D_k	**100**	**100**	**99.52**	**98.32**	**98.56**	**99.52**	**100**	**100**	N/A	**97.60**	**99.28**	N/A	**97.84**	**99.04**	**99.52**	97.84	**93.76**

Table 5. Combined AU detection results of three probability-output fusion methods in terms of log loss.

DB	Detector	AU1	AU2	AU4	AU5	AU6	AU7	AU9	AU12	AU14	AU15	AU17	AU20	AU23	AU24	AU25	AU26	AU27
CK+	D_{wp}	0.77	0.64	2.35	1.38	2.04	2.37	0.25	0.19	0.38	0.85	2.10	0.76	1.08	1.41	1.36	0.35	0.44
	D_{kp}	**0.36**	**0.37**	**1.02**	**0.74**	1.23	1.35	**0.17**	**0.17**	**0.21**	0.62	0.96	**0.55**	0.77	0.83	**0.64**	0.30	**0.27**
	D_{bn}	0.74	0.45	1.13	0.81	**0.69**	**1.06**	0.18	0.27	0.25	**0.60**	**0.76**	0.71	**0.54**	**0.64**	0.76	**0.09**	0.46
ISL	D_{wp}	**0.0001**	**0.0001**	0.056	0.193	0.143	0.056	0.0005	**0.0001**	N/A	0.277	0.082	N/A	0.249	0.110	0.055	1.093	0.150
	D_{kp}	0.004	0.005	**0.022**	**0.086**	0.051	**0.029**	0.0017	0.009	N/A	**0.078**	**0.033**	N/A	**0.061**	0.038	0.028	**0.096**	**0.050**
	D_{bn}	**0.0001**	0.0002	0.030	0.100	**0.050**	0.034	**0.0001**	0.006	N/A	**0.078**	**0.033**	N/A	0.067	**0.035**	**0.027**	0.121	0.052

the former is more certain about the true presence of AUs.

3. Results of Emotion Recognition

In this section, we compare the emotion recognition ability of various combinations of proposed AU detection fusion and AU-to-emotion recognition methods. Emotion recognition is conducted over 309 face images from the CK+ database, which are fully labelled both for the AUs and the six basic emotions. Among the selected 309 face images, 248 images used for the training of the CKD detectors in the preceding AU detection fusion are again used to learn AU-to-emotion mappings and the rest are used for testing.

The LCS-based and model-based AU-to-emotion mappings require training, unlike the rule-based mapping. In the LCS-based mapping, a set of five discriminant AUs is first obtained for each of the six basic emotions from the training portion of CK+ database as in [17]. These are {AU23, AU24, AU17, AU4, AU7} for anger, {AU9, AU17, AU7, AU4, AU6} for disgust, {AU20, AU1, AU4, AU25, AU5} for fear, {AU12, AU6, AU25, AU26, AU14} for happiness, {AU15, AU17, AU1, AU4, AU7} for sadness, and {AU2, AU27, AU5, AU1, AU25} for surprise. In the model-based mapping, a binary

SVM classifier with a radial basis function kernel is trained for each emotion and the outputs of six trained SVM classifiers are combined to finally determine the underlying emotions as in (10). The SVM parameters σ and C are determined by 3-fold cross-validation on the training data.

Figure 4 shows the correct recognition rates obtained by our two-phase emotion recognition framework. The combination of BKS-based probability-output fusion (D_{kp}) and model-based AU-to-emotion mapping achieves the best emotion recognition rate (86.89%) among 13 different AU detection and AU-to-emotion mapping combinations in our framework. This best performance is only 1.64% lower than the best performance obtained by the LCS-based emotion mapping over the human-labelled AUs (88.52%); however, it is slightly better than the correct recognition rate (85.25%) obtained by the conventional rule-based AU-to-emotion mapping on the manually labelled AUs (see the leftmost group of bars labelled with "True" in Fig. 4).

The emotion recognition using the combined decisions of the straightforward AU detection fusion methods that consider individual AU detectors with equal weights (that is, D_m and D_u) attains the worst accuracy rates, which is even worse than
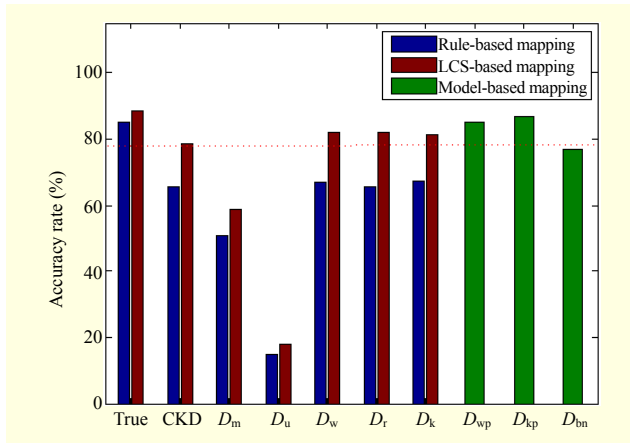
Fig. 4. Accuracy rates of emotion recognition.

Table 6. Confusion matrix of D_{kp} + model-based mapping.

	Anger	Disgust	Fear	Joy	Sadness	Surprise
Anger	**66.7**	0.0	0.0	0.0	22.2	11.1
Disgust	0.0	**100**	0.0	0.0	0.0	0.0
Fear	20.0	0.0	**60.0**	0.0	0.0	20.0
Joy	0.0	0.0	0.0	**100**	0.0	0.0
Sadness	0.0	0.0	40.0	0.0	**60.0**	0.0
Surprise	0.0	0.0	0.0	5.9	0.0	**94.1**

those of previous two-phase emotion recognition framework using a single AU detector (see the second to the leftmost group of bars labelled with "CKD" in Fig. 4). However, the label-output fusion methods that exploits the behaviors of multiple AU detectors (that is, D_w, D_r, and D_k) and the robust LCS-based AU-to-emotion mapping combinations as well as the probability-output fusion and the model-based AU-to-emotion mapping pairs show improved recognition accuracy rates as shown in Fig. 4.

In summary, the combination of probability-output fusion and model-based AU-to-emotion mapping method yield better recognition rates than the combination of label-output fusion or rule-based mapping methods. Also, the LCS-based AU-to-emotion mapping that allows partial template matching outperforms the rule-base mapping that allows only strict matching for the binary outputs of fusion methods.

Table 6 shows the confusion matrix of the best emotion recognition result achieved by the combination of BKS-based probability-output fusion (D_{kp}) and model-based AU-to-emotion mapping. The most recognized emotions are disgust and joy (that is, happiness). Fear is misclassified to anger (20%) and surprise (20%) with similar visual facial expressions and sadness is confused with fear (40%). From a further

comparison against other state-of-the-art two-phase and one-phase approaches, our two-phase method using multiple AU detectors outperformed the previous two-phase approach using a single AU detector and attained comparable emotion recognition rates over three emotions (that is, disgust, joy, and surprise) as compared with the state-of-the-art one-phase approaches. Unless temporal features are used, the performance difference between our method and the most up-to-date one-phase approaches is not significant.

V. Conclusion

In this paper, we proposed a new two-phase facial emotion recognition framework consisting of multiple AU detection, AU detection fusion, and AU-to-emotion mapping. In our framework, the presence of AUs is detected by group decisions of multiple AU detectors and a target emotion is inferred from the combined AU detection decisions. The proposed framework is evaluated over two real-world face databases. The experimental results demonstrate the improved performance over the previous two-phase framework using a single AU detector in terms of both AU detection accuracy and correct emotion recognition rate.

References

[1] P. Ekman and W. Friesen, "*The Facial Action Coding System: A Technique for the Measurement of Facial Movement,*" Palo Alto, CA, USA: Consulting Psychologists Press, 1978.

[2] W. Yun et al., "Hybrid Facial Representations for Emotion Recognition," *ETRI J.*, vol. 35, no. 6, Dec. 2013, pp. 1021–1028.

[3] Y. Li et al., "Data-Free Prior Model for Facial Action Unit Recognition," *IEEE Trans. Affective Comput.*, vol. 4, no. 2, Apr. 2013, pp. 127–141.

[4] G. Donato et al., "Classifying Facial Actions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 10, 1999, pp. 974–989.

[5] Y. Zhu et al., "Dynamic Cascades with Bidirectional Bootstrapping for Action Unit Detection in Spontaneous Facial Behavior," *IEEE Trans. Affective Comput.*, vol. 2, no. 2, Apr.–June 2011, pp. 79–91.

[6] M.F. Valstar and M. Pantic, "Combined Support Vector Machines and Hidden Markov Models for Modeling Facial Action Temporal Dynamics," *IEEE Int. Conf. Human-Comput. Interaction*, Rio de Janeiro, Brazil, Oct. 20, 2007, pp. 118–127.

[7] B. Jiang, M.F. Valstar, and M. Pantic, "Action Unit Detection Using Sparse Appearance Descriptors in Space-Time Video Volumes," *IEEE Int. Conf. Automat. Face Gesture Recogn.*, Santa Barbara, CA, USA, Mar. 21–25, 2011, pp. 314–321.

[8] Z. Zeng et al., "A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions," *IEEE Trans. Pattern Anal.*

Mach. Intell., vol. 31, no. 1, Jan. 2009, pp. 39–58.

[9] I. Cohen et al., "Facial Expression Recognition from Video Sequences: Temporal and Static Modeling," *Comput. Vis. Image Understanding*, vol. 91, no. 1–2, July 2003, pp. 160–187.

[10] M.S. Bartlett et al., "Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behavior," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn.*, San Diego, CA, USA, vol. 2, June 20–25, 2005, pp. 568–573.

[11] K.-Y. Chang, T.-L. Liu, and S.-H. Lai, "Learning Partially-Observed Hidden Conditional Random Fields for Facial Expression Recognition," *IEEE Conf. Comput. Soc. Vis. Pattern Recogn.*, Miami, FL, USA, June 20–25, 2009, pp. 533–540.

[12] Y. Zhang and Q. Ji, "Active and Dynamic Information Fusion for Facial Expression Understanding from Image Sequences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, May 2005, pp. 699–714.

[13] P. Lucey et al., "The Extended Cohn-Kanade Dataset (CK+): A Complete Dataset for Action Unit and Emotion-Specified Expression," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn. Workshop*, San Francisco, CA, USA, June 13–18, 2010, pp. 94–101.

[14] M. Pantic and L.J.M. Rothkrantz, "An Expert System for Multiple Emotional Classification of Facial Expressions," *IEEE Int. Conf. Tools Artif. Intell.*, Chicago, IL, USA, 1999, pp. 113–120.

[15] M.F. Valstar and M. Pantic, "Biologically vs. Logic Inspired Encoding of Facial Actions and Emotions in Video," *IEEE Int. Conf. Multimedia Expo*, Toronto, Canada, 2006, pp. 325–328.

[16] S. Velusamy et al., "A Method to Infer Emotions from Facial Action Units," *IEEE Int. Conf. Acoust.*, *Speech Signal Process.*, Prague, Czech Republic, May 22–27, 2011, pp. 2028–2031.

[17] Q. Ji, *ISL Facial Expression Databases*, Intelligent Systems Lab, Rensselaer Polytechnic Institute. Accessed Feb. 26, 2014. http://www.ecse.rpi.edu/~cvrl/database/database.html

[18] N. Littlestone and M.K. Warmuth, "The Weighted Majority Algorithm," *Inf. Comput.*, vol. 108, no. 2, Feb. 1994, pp. 212–261.

[19] Y.S. Huang and C.Y. Suen, "The Behavior-Knowledge Space Method for Combination of Multiple Classifiers," *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn.*, New York, USA, June 15–17, 1993, pp. 347–352.

[20] D. Heckerman, "A Tutorial on Learning with Bayesian Networks," in *Learning in Graphical Models*, Cambridge, MA, USA: MIT Press, 1999, pp. 301–354.

[21] M.F. Valstar and M. Pantic, "Induced Disgust, Happiness and Surprise: An Addition to the MMI Facial Expression Database," *Int. Conf. Language Resources Evaluation*, Istanbul, Turkey, May 21–27, 2010, pp. 65–70.

Hyunjin Yoon received her BS and MS degrees in computer science and engineering from Ewha Womans University, Seoul, Rep. of Korea, in 1996 and 1998, respectively, and her PhD degree in computer science from the University of Southern California, Los Angeles, USA, in 2009. Since 2010, she has been with ETRI, where she is now a senior researcher. Her main research interests include multidimensional data analysis and machine learning algorithms with applications in human–computer interactions.

Sangwook Park received his BS and MS degrees from Kyungpook National University, Daegu, Rep. of Korea, in 2001 and 2004, respectively. In March 2004, he joined ETRI. He developed IPv6 Routers and oversaw their operation, administration, and maintenance from 2004 to 2006. He is currently working in multisensory representation technology as a senior researcher of the Real and Emotional Sense Platform Research Section and has been engaged in projects involving Giga Korea Software Platform applications and multimedia services. His research interests include IP mobility and routing protocols; multimedia streaming technology; and real-sensory services framework technology.

Yongkwi Lee received his MS degree in medical engineering from Yonsei University, Seoul, Rep. of Korea, in 2009 and is currently working toward his PhD degree in medical engineering at Chungnam National University, Daejeon, Rep. of Korea. In 2009, he worked for Korea Food & Drug Administration, Osong, Rep. of Korea. Since 2009, he has been with ETRI, where he is currently a senior researcher. His research interests include human–computer interfaces, wearable-device designs, emotion-related physiological signal acquisition, mobile health, wireless body area networks, emotion recognition, emotional-ICT, realistic devices, and smart factory.

Mikyong Han received her MS degree in computing engineering from the School of Electronics and Information, Kyung Hee University, Seoul, Rep. of Korea, in 1993. She joined ETRI in 1993 and is currently a principal research member. From March 2012 to February 2013, she was a visiting professor at the School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, USA. Her major research interests include multi-media service platforms, emotional service platforms, and immersive media service platforms.

Jong-Hyun Jang received his PhD degree in computer science and engineering from Hankuk University of Foreign Studies, Yongin, Rep. of Korea, in 2004. Since 1994, he has been with ETRI, where he is currently a principle member of the research staff as well as a team leader of the Real and Emotional Sense Platform Research Section. He has worked on several projects for the development of programming environments and PCS since 1994. His research interests are real-time middleware for telecommunication systems, home networking systems, and real-sense media services.