Human Facial Expression Recognition using Stepwise Linear Discriminant Analysis and Hidden Conditional Random Fields†

Muhammad Hameed Siddiqi, Rahman Ali, Adil Mehmood Khan, Young-Tack Park, *Member, IEEE,* and Sungyoung Lee*, *Member, IEEE,*

***Abstract*—This paper introduces an accurate and robust facial expression recognition (FER) system. For feature extraction, the proposed FER system employs Stepwise Linear Discriminant Analysis (SWLDA). SWLDA focuses on selecting the localized features from the expression frames using the partial F-test values, thereby reducing the within class variance and increasing the low between variance among different expression classes. For recognition, the Hidden Conditional Random Fields (HCRF) model is utilized. HCRF is capable of approximating a complex distribution using a mixture of Gaussian density functions. To achieve optimum results, the system employs a hierarchical recognition strategy. Under these settings, expressions are divided into three categories based on parts of the face that contribute most towards an expression. During recognition, at the first level, SWLDA and HCRF are employed to recognize the expression category; whereas, at the second level, the label for the expression within the recognized category is determined using a separate set of SWLDA and HCRF, trained just for that category. In order to validate the system, four publicly available datasets were used, and a total of four experiments were performed. The weighted average recognition rate for the proposed FER approach was 96.37% across the four different datasets, which is a significant improvement in contrast to the existing FER methods.**

***Index Terms*—Facial Expressions, Stepwise Linear Discriminant Analysis, Hidden Markov Models, Hidden Conditional Random Fields.**

I. Introduction

Expressions play a vital role in our daily communications, and recent years have witnessed a great amount of work being done to develop accurate and reliable facial expressions recognition (FER) systems. Such systems can be employed in many applications, such as in daily communications, personality and child development [3], neuroscience and psychology [4], access control and surveillance [5], and human behavior studies in telemedicine and e-health environments [6].

FER systems can be categorized into two types: posed expression recognition systems [7], [8] and spontaneous expression recognition systems [9], [10]. Former case deals with recognizing artificial expressions: expressions produced by people when they are asked to do so [5]. On the other hand, the latter case deals with the expressions that people give out spontaneously, and these are the ones that can be observed on a day-to-day basis, such as during conversations or while watching movies [5]. The focus of this study is posed FER systems.

A typical FER system employs either frame-based classification or sequence-based classification. In frame-based classification methods, only the current frame is utilized with or without a reference image (neutral face image) in order to recognize the expressions; whereas, in sequence-based classification methods, the temporal information of the sequences are utilized in order to recognize the expressions in one or more frames [11]. In sequence-based methods, the geometrical displacement of facial feature points between the current frame and the initial frame are calculated [12]; whereas, frame-based methods do not have this property. The temporal information of expression in sequences of frames is important for facial expression analysis [13].

In this research study, we propose the use of Stepwise Linear Discriminant Analysis (SWLDA) coupled with Hidden Conditional Random Fields (HCRF) for a sequence-based FER system named SH-FER. The block diagram of the SH-FER is shown in Fig. 1. Though SWLDA has been used in many different areas before [14], it is for the first time that it is being utilized as a feature extraction technique in an FER system. The purpose of using SWLDA as a feature extraction technique is to extract the localized features from faces that the previous feature extraction techniques were limited in analyzing. As for the HCRF, the existing HCRF models are limited by their independence suppositions [15], which may reduce classification accuracy. In this work, we have tried to overcome this limitation by approximating the complex distributions by using a mixture of full covariance Gaussian density function.

Another important aspect of this work is that our system is based on the theory that different expressions can be grouped into three categories based on the part of the face that most contributes to the expression [16], [17], [18]. This classification is shown in Table I.

Those expressions in which the lips have a major contribution are labeled as lip-based expressions. In lip-eyes-based expressions, both the lips and eyes equally contribute to the expressions. In lip-eyes-forehead expressions, the lips, eyes, and eyebrows or forehead have equal roles. In our FER system

†This work is the extension of our previous works [1], [2].

M.H. Siddiqi, R. Ali, and S. Lee are with the Department of Computer Engineering, Kyung Hee University, Suwon, Rep. of Korea (emails: {siddiqi,rahmanali,sylee}@oslab.khu.ac.kr)

A.M. Khan is with the Department of Computer Science, Innopolis University, Kazan, Russia (email: amtareen@gmail.com)

Y.-T. Park is with the School of Information Technology, Soongsil University, Seoul, Rep. of Korea (email: park@ssu.ac.kr)

*Corresponding authors' email: sylee@oslab.khu.ac.kr

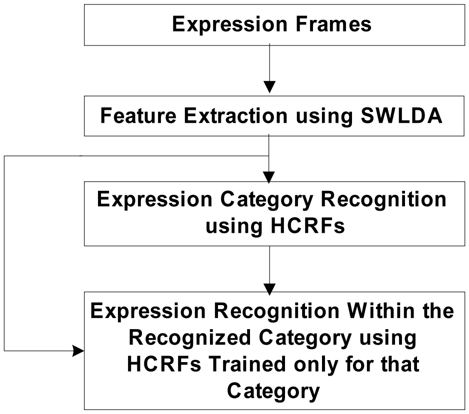

Fig. 1. Block diagram of the SH-FER.

TABLE I
THE CLASSIFIED CATEGORIES AND FACIAL EXPRESSIONS USED IN THIS STUDY.

Category	Facial Expressions
Lips-Based	Happy
	Sad
Lips-Eyes-Based	Surprise
	Disgust
Lips-Eyes-Forehead-Based	Anger
	Fear

(SH-FER), as shown in Fig. 1, an expression is classified into one of these three categories at the first level; then at the second level, SWLDA and HCRF (trained for the recognized category) are employed to label this expression within the recognized category. The SH-FER yielded a weighted average recognition rate of 96.37% when tested on four publicly available standard datasets of facial expressions, which is a significant improvement in accuracy.

The rest of the paper is organized as follows. Section II reviews related work regarding feature extraction and classification modules and their limitations in the FER domain. Section III presents an overview of the SH-FER system. The experimental setup of the SH-FER is represented in Section IV. The experimental results and discussion are given in Section V. Finally, the paper concludes with future directions in Section VI.

II. RELATED WORK

A. *Feature Extraction*

Feature extraction deals with extracting distinguishable features from each facial expression shape and quantizing them into discrete symbols [19]. According to the face descriptors, there are two types of features: global features and local features. For global features, features are extracted from the entire face, whereas for local features, parts of the face, such as eyes, mouth, nose and forehead are used.

Global feature extraction methods are known as holistic methods. These include Nearest Features Line-based Subspace Analysis [20], Eigenfaces and Eigenvector [21], [22] and [23], Fisherfaces [24], global features [25], Independent Component Analysis (ICA) [26], [27], Principal Component Analysis (PCA) [28], [29], [30], frequency-based methods [31], Gabor wavelet [32]. One problem that can be associated with the use of these methods is the fact that they are very sensitive to variations in pose, illumination, occlusion, aging, and rotation changes of the face [33], [34]. Furthermore, these techniques are poor at handling data where classes do not follow the Gaussian distribution. Also, these techniques do not work well in case of a small sample size [34]. Furthermore, complexity-wise, most of these techniques are much expensive because of considering the entire face, as this requires more memory [35]. Lastly, these methods work well mostly in a controlled environment [36].

On the other hand, local feature extraction methods compute local descriptors from parts of the face and then integrate this information into one descriptor. These include Local Feature Analysis (LFA) [37], Gabor features [38], Non-negative Matrix Factorization (NMF) and Local non-negative Matrix Factorization (LNMF) [39], and Local Binary Pattern (LBP) [12], [40]. Among these methods, LBP is the most commonly employed feature extraction technique. However, LBP does not provide the directional information of the facial frame [1].

Some recent studies have tried to solve the limitations of LBP. These methods include Local Transitional Pattern (LTP) [41], Local Directional Pattern (LDP) [42], Local Directional Pattern Variance (LDPv) [43]. Most of these methods exploited other information instead of employing intensity to overcome the problems due to noise and illumination change [44]. However, performance of these methods still degrade in non-monotonic illumination change, noise variation, change in pose, and expression conditions [44]. Another commonly used local feature extraction method for expression recognition is Local Fisher Discriminant Analysis (LFDA) [45]. But, LFDA fails to determine the essential assorted structure when face image space is highly nonlinear [46]. Furthermore, authors of [47] employed pixel and color segmentation for feature extraction to detect facial expressions. However, the performance of this approach also degrades with variation in illumination.

B. *Classification*

As for the classification module, a large number of methods have been employed for accurate expression classification. In [48], authors exploited artificial neural networks (ANNs) in order to classify different facial expressions and achieved a 73% recognition rate. However, ANN is a black box and has incomplete capability to explicitly categorize possible fundamental relationships [49]. Besides, ANNs may take long time to train and may trap in a bad local minima. Moreover, authors of [50] and [12] employed support vector machines (SVMs) for their FER system. But, in SVMs, the observation probability is calculated using indirect techniques; in other words, there is no direct estimation of the probability [51]. Furthermore, SVMs simply disregard temporal dependencies among video frames, and thus each frame is expected to be statistically independent from the rest. Similarly, authors of [52] and [53] utilized Gaussian mixture models (GMMs) to recognize different types of facial expressions. But facial features could be very sensitive to noise; therefore, fast variations in facial frames cannot be modeled by GMMs and might cause misclassification [11].

Most of the aforementioned classifiers were employed for frame-based classification. On the other hand, the most commonly used sequence-based classification method is the Hidden Markov Models (HMMs) [54], [55]. HMMs have their own advantage in handling sequential data when frame-level features are used, whereas vector-based classifiers, such as GMMs, ANNs, and SVMs, fail to learn the sequence of the feature vectors.

Nevertheless, conventional HMMs are based on Markovian property, which presumes that the current state depends only on the previous state. Because of this assumption, labels of two contiguous states must hypothetically occur consecutively in the observed sequence. Unfortunately, this presumption is not always true in reality. Some other limitations of HMMs include their generative nature and the independence assumption between states and observations [56]. A non-generative model such as maximum entropy Markov model (MEMM) was developed in order to resolve the limitations of HMM, and it produced better results compared to HMM [57]. However, MEMM has a commonly known drawback called the "*label bias problem*".

Conditional Random Fields (CRF) [56] and HCRF [58], the generalizations of MEMM, were then proposed to take the full advantage of MEMM and to solve the "*label bias problem*" [56]. HCRF extends the capability of CRF with hidden states making it able to learn hidden structure of the sequential data. Both of them use global normalization instead of per-state normalization. Thus, they allow weighted scores, making the parameter space larger than those of MEMM and HMM. The following discussion provides the underlying theory of HCRF, and analyzes the limitations in their existing implementations.

We consider a task of mapping from inputs X to labels $Y \in \Gamma$, for instance, Γ = {happy, anger, sad, surprise, disgust, fear} in an FER problem. Each input X is a sequence of T frames, $X = x_1, x_2, ..., x_T$. The training set contains N pairs $(X_i, Y_i), i = 1, 2, ..., N$. In a Q-state HCRF, the conditional probability of a class label Y given input X and set of parameters of the model Λ is computed as

$$p\left(Y|X;\Lambda\right) = \frac{\sum\limits_{\overline{S}} \exp\left\{\Lambda \cdot f\left(Y, \overline{S}, X\right)\right\}}{z\left(X, \Lambda\right)}, \tag{1}$$

where

$$z\left(X, \Lambda\right) = \sum_{Y'\bar{S}} \exp\left\{\Lambda \cdot f\left(Y', \overline{S}, X\right)\right\}, \tag{2}$$

is the normalization factor to guarantee the sum-to-one rule of the conditional probability, where, Y' is the predicted label for the sequence, and $\overline{S} = \{s_1, s_2, \ldots, s_T\}$ is a sequence of hidden states. Each $s_i, i = 1, 2, ..., T$, can have an integer value from 1 to Q, the number of states, Λ is the parameter vector and $f\left(Y, \overline{S}, X\right)$ is known as the feature vector that consists of the following sufficient statistics used by the model.

$$\overset{Pr}{\underset{y'}{f}}\left(Y, \overline{S}, X\right) = \delta\left(y = y'\right), \quad \forall y' \in Y, \tag{3}$$

$$\overset{Tr}{\underset{ss'}{f}}\left(Y, \overline{S}, X\right) = \sum_{t=1}^{T} \delta(s_{t-1} = s)\, \delta(s_t = s'), \quad \forall \{ss'\} \in \overline{S}, \tag{4}$$

$$\overset{Occ}{\underset{s}{f}}\left(Y, \overline{S}, X\right) = \sum_{t=1}^{T} \delta(s_t = s), \quad \forall s \in \overline{S}, \tag{5}$$

$$f_s^{M_1}\left(Y, \overline{S}, X\right) = \sum_{t=1}^{T} \delta(s_t = s) x_t, \quad \forall s \in \overline{S}, \tag{6}$$

$$f_s^{M_2}\left(Y, \overline{S}, X\right) = \sum_{t=1}^{T} \delta(s_t = s) x_t^2, \quad \forall s \in \overline{S}, \tag{7}$$

where $\delta\left(s = s'\right)$ is equal to one when $s = s'$, otherwise equal to zero. Thus, $\overset{Pr}{\underset{y'}{f}}\left(Y, \overline{S}, X\right)$ in (3) tracks the number of times the predicted labels are equal to the original labels. Similarly, $\overset{Tr}{\underset{ss'}{f}}\left(Y, \overline{S}, X\right)$ in (4) determines the number of times the transition ss' occurs in $\overline{S}$, and this process is repeated for the entire state sequence. Likewise, $\overset{Occ}{\underset{s}{f}}\left(Y, \overline{S}, X\right)$ in (5) counts the occurrence of the state s. The first and second moments $f_s^{M_1}$ and $f_s^{M_2}$ in (6) and (7) respectively are the sum and sum of the squares of observations that align with the state s. It is to be noted that the term feature vector does not refer to the input features, but refers to the vector of sufficient statistics used by the model. Latter is referred to as the observation vector. The choice of the feature vector determines the dependencies of the HCRF model.

It can be seen from the above equations that with some specific set of parameters (Λ), HCRF's dependencies are similar to those of HMM. For example with above feature vector, the diagonal-covariance Gaussian distribution can be defined as

$$\overset{Pr}{\underset{y'}{\Lambda}} = \log(u_{y'}), \quad \forall y' \in Y, \tag{8}$$

$$\overset{Tr}{\underset{ss'}{\Lambda}} = \log(A_{ss'}), \quad \forall \{ss'\} \in \overline{S}, \tag{9}$$

$$\overset{Occ}{\underset{s}{\Lambda}} = -\frac{1}{2}\left(\log\left(2\pi\sigma_s^2\right) + \frac{\mu_s^2}{\sigma_s^2}\right), \tag{10}$$

$$\Lambda_s^{M_1} = \frac{\mu_s}{\sigma_s^2}, \tag{11}$$

$$\Lambda_s^{M_2} = -\frac{1}{2\sigma_s^2}, \tag{12}$$

where u in (8) is the prior distribution of Gaussian-HMM, and A in (9) is a transition matrix, then the numerator of the condition probability can be written as

$$\sum_{\overline{S}} \exp\left\{\Lambda \cdot f\left(Y, \overline{S}, X\right)\right\} = \sum_{\overline{S}} u(s_1) \prod_{t=1}^{T} A\left(s_{t-1}, s_t\right) N\left(x_t^2, \mu_{S_t}, \sigma_{S_t}\right), \tag{13}$$

where N denotes the Gaussian distribution. The conditional probability of X given Y is computed with a Gaussian-HMM by (13) that has a prior distribution u, and a transition matrix A.

A more generalized version of the HCRF model has been proposed by [59] in order to handle more complex distributions using a linear mixture of Gaussian density functions, and is given as

$$p(Y|X;\Lambda) = \frac{\sum_{\overline{S}} \sum_{m=1}^{M} \exp\left\{\Lambda \cdot f\left(Y, \overline{S}, m, X\right)\right\}}{z(X,\Lambda)}, \quad (14)$$

where M is the number of components in the Gaussian mixture.

Although, there are some existing works that employed the above HCRF model and showed good results [60], [61]. They did not address and overcome the limitations of the model. As we can see in the above equation, that the model can only utilize diagonal-covariance Gaussian distribution. In other words, the variables (columns of $x_i, i = 1, 2, ..., N$) are assumed to be pair-wise independent. Hereafter, we call this model *diagonal covariance Gaussian mixture hidden conditional random fields* (DCGM-HCRF). In addition, equations (10), (11), and (12) imply that with a particular set of values, the observation density at each state will converge to Gaussian form. Unfortunately, there is algorithm that could guarantee this convergence. Therefore, these assumptions may result in a decrease of accuracy.

In order to inherit the advantages of HCRF model and completely tackle the limitations of the existing work, we propose the use of HCRF algorithm that is able to explicitly utilize mixture of *full covariance Gaussian mixture hidden conditional random fields* (FCGM-HCRF).

III. Materials and Methods

A. Stepwise Linear Discriminant Analysis (SWLDA)

Dimension reduction by extracting discriminating features is based on the idea of maximizing the total scatter of the data while minimizing the variance within classes. It can be seen in Fig. 2 that the feature values (gray scale values) for the six classes are highly merged, which can result in a high misclassification rate. Please note that the actual number of features (gray scale values) could be more than three, however, for the sake of visualization, the first three features were picked in order to create Fig. 2.

The problem shown in Fig. 2 is due to similarities among the expressions that result in high within-class variance and low between-class variance. Therefore, a method is required that not only provides the dimension reduction, but also increases the low between-class variance to increase class separation before the features are fed to the classifier.

In order to solve this problem, several methods have been proposed in the machine learning literature, such as kernel discriminant analysis (KDA) [62], generalized discriminant analysis (GDA) [63], and linear discriminant analysis (LDA) [64]. Among these, LDA has been widely employed in FER systems. However, LDA is a linear technique that is limited in flexibility when applied to more complex datasets. For more details on LDA, please refer to a previous study [65]. Accordingly, this work employs a robust feature extraction technique called Stepwise Linear Discriminant Analysis (SWLDA). SWLDA is easy to explain, has good predictive ability, and computationally, it is less expensive than other existing methods [14]. Some limitations of the existing works, such as illumination change, do not affect the performance of the SWLDA. SWLDA only extracts a small set of features by employing forward and backward regression models. In forward regression, the most correlated features are selected based on partial F-test values, whereas in backward regression, the least significant features are removed from the regression model. In both cases, F-test values are calculated on the basis of defined class labels. The advantage of this method is that it is very efficient for seeking localized features. The actual number of the extracted features used is 200. For more details on SWLDA, please refer to a previous study [14].

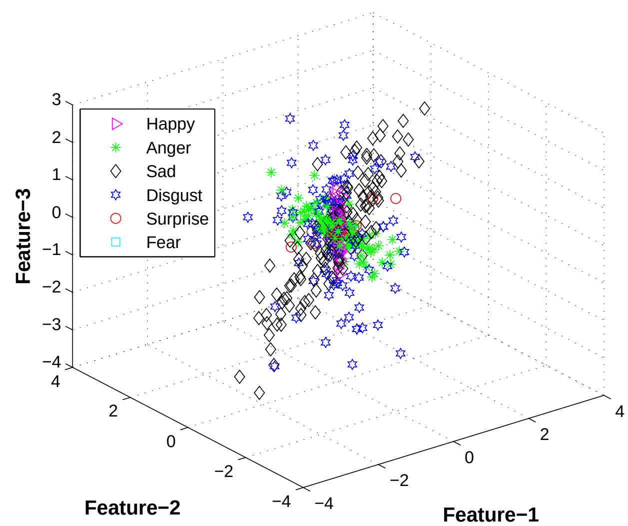

Fig. 2. 3D-feature (gray scale values) plot for six types of facial expressions.

B. Hidden Conditional Random Fields (HCRF)

As mentioned before that the existing HCRF utilizes diagonal covariance Gaussian distributions in the feature function and does not guarantee the convergence of its parameters to some specific values at which the conditional probability is modeled as a mixture of normal density functions. Because of this property, the existing HCRF losses a lot of information. This is one of the main disadvantages of the existing HCRF model.

In order to solve this limitation, we explicitly involve full covariance Gaussian distributions in the feature functions at the observation level. For the prior and transition probabilities, we used the same equations of [59] as described in (3) and

(4). Mathematically, our contribution can be explained as

$$f_s^{Ob}\left(Y,\overline{S},X\right) = \sum_{t=1}^{T} \log\left(\sum_{m=1}^{M}\Gamma_{s,m}^{Obs} N\left(x_t^2,\mu_{s,m},\Sigma_{s,m}\right)\right)\left(\delta\left(s_t=s\right)\right), \tag{15}$$

The (15) presents the observation of the input at each state. Where M is the number of density functions, Γ is used in order to consider the contextual information of the whole observation, $\Gamma_{s,m}^{Obs}$ is the mixing weight of the m^{th} component with mean $\mu_{s,m}$ and covariance matrix $\sum_{s,m}$. $N\left(x_t^2,\mu_{s,m},\Sigma_{s,m}\right)$ in (15) can be computed as

$$N\left(x_t^2,\mu_{s,m},\Sigma_{s,m}\right) = \frac{1}{(2\pi)^{D/2}|\Sigma_{s,m}|^{1/2}}\exp\left(-\frac{1}{2}\left(x_t^2-\mu_{s,m}\right)'\Sigma_{s,m}^{-1}(x_t^2-\mu_{s,m})\right), \tag{16}$$

where D is the dimension of the observation, and $\sum_{s,m}$ is the full covariance matrix.

As we can see in (15), by changing Γ, μ and $\sum$ we can create any mixture of the normal densities. So, the corresponding observation weight (Λ_s^{Obs}) is not necessary to be updated during the training phase. Therefore,

$$\Lambda_s^{Ob} = 1, \quad \forall s \in \overline{S}, \tag{17}$$

As a result, the conditional probability that is used to model the system can be rewritten as

$$p\left(Y|X;\Lambda,\Gamma,\mu,\Sigma\right) = \frac{\sum\limits_{\overline{S}}\exp\left(P\left(\overline{S}\right)+T\left(\overline{S}\right)+O\left(\overline{S}\right)\right)}{z\left(X,\Lambda,\Gamma,\mu,\Sigma\right)}, \tag{18}$$

where

$$P\left(\overline{S}\right) = \sum_{s\in\overline{S}}\Lambda_{y'}^{Pr} f_{y'}^{Pr}\left(Y,\overline{S},X\right), \tag{19}$$

$$T\left(\overline{S}\right) = \sum_{\{ss'\}\in\overline{S}}\Lambda_{ss'}^{Tr} f_{ss'}^{Tr}\left(Y,\overline{S},X\right), \tag{20}$$

$$O\left(\overline{S}\right) = \sum_{s\in\overline{S}} f_s^{Ob}\left(Y,\overline{S},X\right), \tag{21}$$

By putting the values of $P\left(\overline{S}\right)$, $T\left(\overline{S}\right)$, and $O\left(\overline{S}\right)$ from (19), (20), and (21) respectively in (18), the updated conditional probability can be rewritten in (22).

As mentioned before, our contribution is at the observation level; therefore, by putting the value of $f_s^{Ob}\left(Y,\overline{S},X\right)$ from (15), the updated conditional probability for the system can be rewritten in (23). The simple form of the conditional probability is defined in (24).

$$p\left(Y|X;\Lambda,\Gamma,\mu,\Sigma\right) = \frac{Score\left(Y|X;\Lambda,\Gamma,\mu,\Sigma\right)}{z\left(X;\Lambda,\Gamma,\mu,\Sigma\right)}, \tag{24}$$

The procedure of the proposed HCRF follows exactly the procedure of the [59]. Based on equations (23) and (24), we can further update the conditional probability using the well-known forward and backward algorithms (as the algorithms used in HMM), which are defined in equations (25) and (26) respectively.

Therefore, the $Score\left(Y|X;\Lambda,\Gamma,\mu,\Sigma\right)$ of (25) is equal to the forward algorithm (α) and backward algorithm (β) as in (27).

$$Score\left(Y|X;\Lambda,\Gamma,\mu,\Sigma\right) = \sum_{s\in\overline{S}}\alpha_T(s) = \sum_{s\in\overline{S}}\beta_1(s). \tag{27}$$

In the training phase, our goal was to find the parameters ($\Lambda,\Gamma,\mu,$ and $\sum$) to maximize the conditional probability of the training data. In SH-FER, we utilize (Limited-memory Broyden-Fletcher-Goldfarb-Shanno) L-BGFS method to search the optimal point. However, instead of repeating the forward and backward algorithms to compute the gradients as others did [59], we run the forward and backward algorithms only when calculating the conditional probability, then we reuse the results to compute the gradients.

IV. Experimental Setup

The SH-FER has been tested and validated on four publicly available standard datasets of facial expressions [66], [67], [68], and [69]. In total, four experiments were performed. All the experiments were performed in Matlab using an Intel® Pentium® Dual-Core^{TM} (2.5 GHz) with a RAM capacity of 3 GB. Detailed description of the four datasets used in this study are as follows:

- *Extended Cohn-Kanade Dataset (CK+)*:
  This facial expressions dataset contains 593 video sequences on seven facial expressions recorded from 123 subjects (university students). The age range of the subjects was from 18 to 30 years and most of them were female. Out of 593 video sequences, 309 were used in this work, corresponding to the six expressions mentioned in Table I. The original size of each facial frame in some of the images is 640×480, and 640×490 pixel in others, with 8-bit precision for grayscale values.
- *Japanese Female Facial Expressions (JAFFE) Dataset*:
  The expressions in this dataset were posed by 10 different subjects (Japanese female). Most of the expression frames were taken from the frontal view of the camera with tied hair in order to expose all the sensitive regions of the face. In the whole dataset, there is a total of 213 facial frames, which consist of seven expressions including neutral. Therefore, we selected only 195 expression frames, corresponding to the six facial expressions. The original size of each facial frame is 256×256 pixel.
- *Extended Yale B Face (B+) Dataset*:
  This dataset contains a total of 16128 facial frames taken under a single light source performed by 28 distinct subjects for 576 viewing conditions (9 poses × 64 illumination conditions). The original size of each facial frame is 320×243 pixel.
- *MMI Dataset*:
  The MMI dataset of facial expressions is a fully web-searchable collection of visual and audio-visual recordings of subjects displaying a facial expression. This dataset contains a total of 238 video sequences performed by 28 subjects (male and female). The original size of each facial frame is 720×576 pixel.

$$p\left(Y|X;\Lambda,\Gamma,\mu,\Sigma\right)=\frac{\sum\limits_{\overline{S}}\exp\left(\sum\limits_{s\in\overline{S}}\Lambda_{y'}^{Pr}f_{y'}^{Pr}\left(Y,\overline{S},X\right)+\sum\limits_{\{ss'\}\in\overline{S}}\Lambda_{ss'}^{Tr}f_{ss'}^{Tr}\left(Y,\overline{S},X\right)+\sum\limits_{s\in\overline{S}}f_{s}^{Ob}\left(Y,\overline{S},X\right)\right)}{z\left(X,\Lambda,\Gamma,\mu,\Sigma\right)}, \tag{22}$$

$$p\left(Y|X;\Lambda,\Gamma,\mu,\Sigma\right)=\frac{\sum\limits_{\overline{S}=s_1,s_2,..,s_T}\exp\left(\Lambda_{y'}^{Pr}+\sum\limits_{t=1}^{T}\left(\Lambda_{s_{t-1},s_t}^{Tr}\right)+\log\left(\sum\limits_{m=1}^{M}\Gamma_{s_t,m}^{Obs}N\left(x_t^2,\mu_{s_t,m},\Sigma_{s_t,m}\right)\right)\right)}{z\left(X,\Lambda,\Gamma,\mu,\Sigma\right)}, \tag{23}$$

$$\alpha_\tau=\sum_{\overline{S}=s_1,s_2,..,\{s_\tau=s\}}\exp\left(\Lambda_{y'}^{Pr}+\sum_{t=1}^{\tau}\left(\Lambda_{s_{t-1},s_t}^{Tr}\right)+\log\left(\sum_{m=1}^{M}\Gamma_{s_t,m}^{Obs}N.\left(x_t^2,\mu_{s_t,m},\Sigma_{s_t,m}\right)\right)\right),$$

$$\alpha_\tau=\sum_{s'\in\overline{S}}\alpha_{\tau-1}\left(s'\right)\exp\left(\Lambda_{s's}^{Tr}+\log\left(\sum_{m=1}^{M}\Gamma_{s,m}^{Obs}N\left(x_\tau,\mu_{s,m},\Sigma_{s,m}\right)\right)\right), \tag{25}$$

$$\beta_\tau\left(s\right)=\sum_{\overline{S}=\{s_\tau=s\},s_{\tau+1},..,s_T}\exp\left(\Lambda_{y'}^{Pr}+\sum_{t=\tau}^{T}\left(\Lambda_{s_{t-1},s_t}^{Tr}\right)+\log\left(\sum_{m=1}^{M}\Gamma_{s_t,m}^{Obs}N\left(x_t^2,\mu_{s_t,m},\Sigma_{s_t,m}\right)\right)\right),$$

$$\beta_\tau\left(s\right)=\sum_{s'}\beta_{\tau+1}\left(s'\right)\exp\left(\Lambda_{ss'}^{Tr}+\log\left(\sum_{m=1}^{M}\Gamma_{s,m}^{Obs}N\left(x_\tau,\mu_{s,m},\Sigma_{s,m}\right)\right)\right), \tag{26}$$

The four experiments are explained as follows:

- In the first experiment, SH-FER was validated using a 10-fold cross-validation rule for each dataset. In other words, each dataset was divided into 10 subsets. Out of these 10 subsets, one subset was used as the validation data, whereas the remaining nine subsets were used as the training data. For each dataset, this process was repeated 10 times, with data from each subset used exactly once as the validation data.
- In the second experiment, $n-$fold cross-validation rule based on dataset was performed (in our case n =4). It means that from the four datasets, data from the three datasets were retained as the validation data for testing the system, and the data from the remaining dataset was used as the training data. This process was repeated four times, with data from each dataset used exactly once as the training data.
- In the third experiment, a set of three sub-experiments were performed in order to show the effectiveness of sub-components of SH-FER, i.e., SWLDA and HCRF. For this purpose, from the first experiment, the best case (dataset) was selected based on the recognition rate. Next, three sub-experiments were performed using the 10-fold validation rule. In the first case, ICA (a well-known local feature extraction technique) was utilized with HCRF instead of SWLDA. In the second case, ICA was coupled with LDA (a well-known discriminant analysis approach) before feeding the features to HCRF. Finally, in the third case, the existing HCRF [59] was used with SWLDA instead of using proposed HCRF.
- Lastly, in the fourth experiment, the performance of SH-FER was compared with some well-known existing FER systems, including [44], [70], [71], [72], [73], [12], [74], [75]. We borrowed the implementations for some of the methods, whereas for other methods, the published results in their respective papers were used.

V. Experimental Results and Discussion

A. First Experiment

The classification results at the first level of SH-FER (expression category classification) are shown in Table II(A) (using CK+), Table II(B) (using JAFFE), Table II(C) (using Extended Yale), and in Table II(D) (using MMI). The Feature plots for these cases are shown in Fig. 3, 4, 5, and 6 respectively. Once again, please note that the actual number of features used is 200. However, for the sake of visualization, we just picked the first three features in order to create all the feature plots. These plots show that at the first level, SH-FER provided a clear separation among the three categories for each dataset.
The overall classification results for the second level classification (expression classification within each category) using CK+, JAFFE, Extended Yale B, and MMI datasets are shown in Table III(A), Table III(B), Table III(C), and Table III(D), respectively. Similarly, the feature plots for these cases are shown in Fig. 7, Fig. 8, Fig. 9, and Fig. 10, respectively. These results indicate that the SH-FER consistently achieved a high recognition rate when applied to these datasets separately.

B. Second Experiment

For the second experiment, the overall results are shown in Table IV. It is clear from Tables IV(A) and IV(B) that SH-FER achieved a high recognition rate when it was trained using the CK+ and Extended Yale B face datasets. However, the system achieved low accuracy when it was trained on the JAFFE and

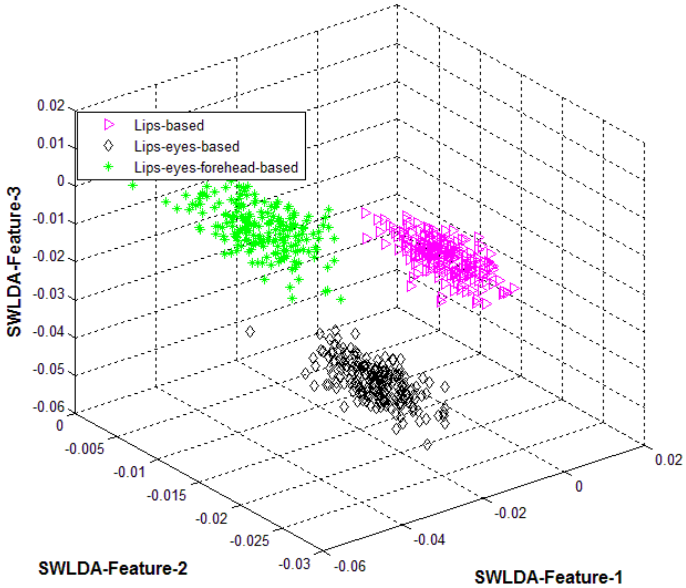

Fig. 3. 3D feature plots for the three expression-categories after applying SWLDA at the first level of SH-FER on CK+ dataset.

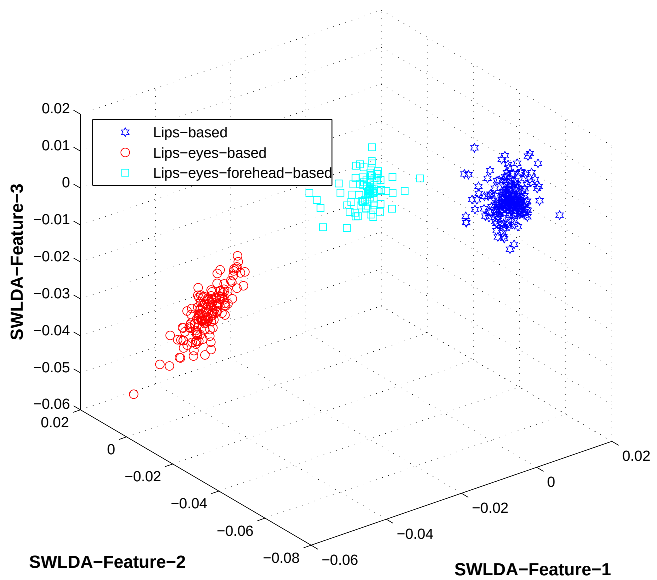

Fig. 4. 3D feature plots for the three expression-categories after applying SWLDA at the first level of SH-FER on JAFFE dataset.

MMI datasets (shown in Table IV(C) and (D)). This might be because the datasets have different facial features; for instance, some of the subjects in the Extended Yale B face dataset have worn glasses, whereas subjects in the CK+ and JAFFE datasets did not wear glasses. Furthermore, eye features in the JAFFE dataset are very different from those in CK+, Extended Yale B and MMI datasets. Similarly, some of the subjects in MMI dataset are at 30^o to 45^o angle to the camera, which might have resulted in low accuracy. Nevertheless, the results are very encouraging and this suggests that the SH-FER is robust, i.e., the system not only achieved a high recognition rate on

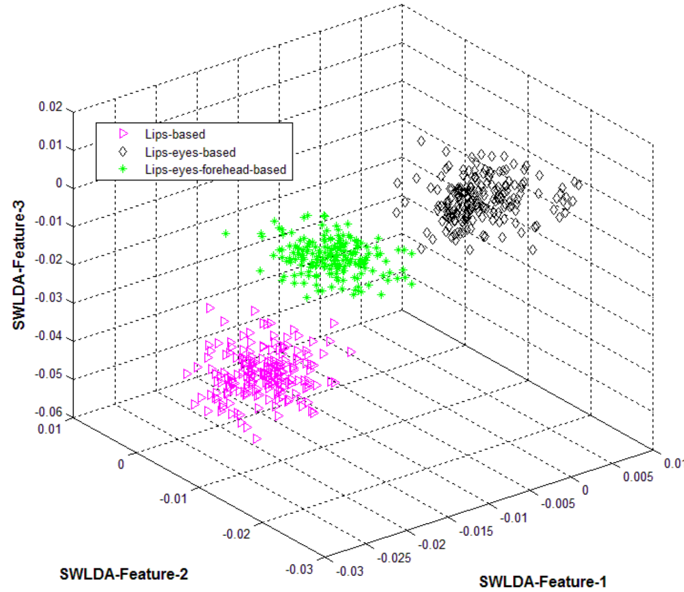

Fig. 5. 3D feature plots for the three expression-categories after applying SWLDA at the first level of SH-FER on Extended Yale B face dataset.

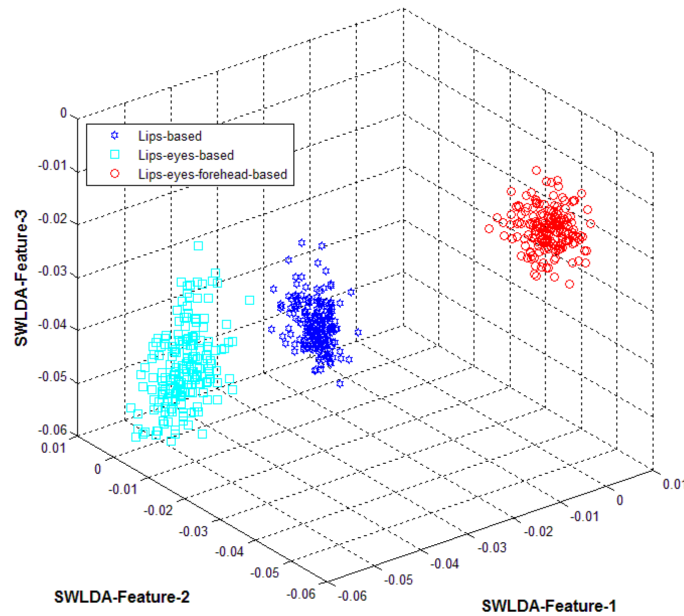

Fig. 6. 3D feature plots for the three expression-categories after applying SWLDA at the first level of SH-FER on MMI dataset.

one dataset, it also provided good recognition rates when used across multiple datasets.

C. Third Experiment

The overall results for the three cases are shown in Table V. These results indicate that both the SWLDA and the HCRF played vital roles in the high accuracy of SH-FER system. It is apparent from Table V(A) and (B) that when SWLDA was replaced with ICA, the system was unable to achieve adequate recognition accuracy. The reason behind the better performance of SWLDA is apparent in Fig. 7, 8, 9, 10. When compared to Fig. 2, one can notice a clear separation among

TABLE II
CLASSIFICATION RESULTS OF SH-FER FOR EXPRESSION CATEGORY CLASSIFICATION ON: (A) CK+ DATASET, (B) JAFFE DATASET, (C) EXTENDED YALE B FACE DATASET (B+), AND (D) MMI DATASET OF FACIAL EXPRESSIONS. (UNIT: %).

Expressions Category	Classification Rate
Lips-based	99
Lips-eyes-based	99
Lips-eyes-forehead-based	98
Average	**98.66**

(A)

Expressions Category	Classification Rate
Lips-based	99
Lips-eyes-based	97
Lips-eyes-forehead-based	99
Average	**98.33**

(B)

Expressions Category	Classification Rate
Lips-based	97
Lips-eyes-based	98
Lips-eyes-forehead-based	99
Average	**98.00**

(C)

Expressions Category	Classification Rate
Lips-based	98
Lips-eyes-based	99
Lips-eyes-forehead-based	98
Average	**98.33**

(D)

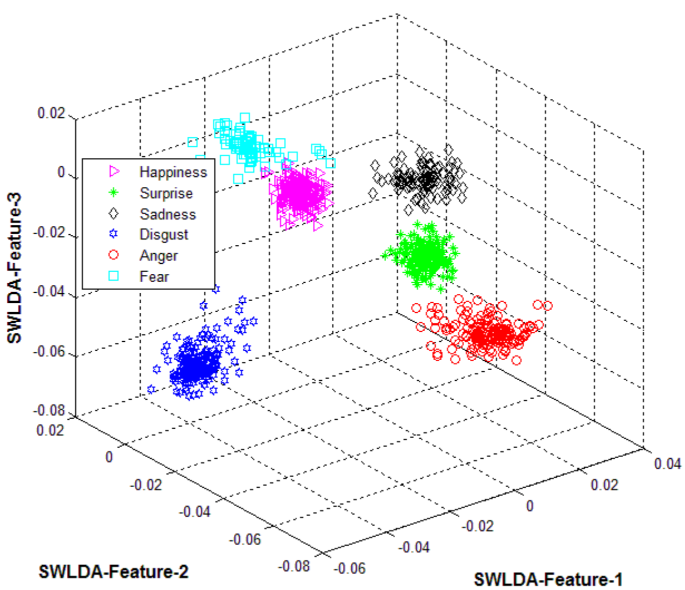

Fig. 7. 3D feature plots for the six expressions after applying SWLDA at the second level of SH-FER on CK+ dataset.

all the expression classes in these figures. Thus, SWLDA not only provides dimension reduction, it also increases the low between-class variance to increase the class separation before the features are fed to the classifier. The low within class variance and high between class variance are achieved because of the forward and backward regression models in the SWLDA.

Likewise, it is also obvious from Table V(C) that when HCRF was replaced with existing HCRF [59], the system was unable to achieve good recognition rate. Thus the proposed HCRF model successfully addresses the limitations of HMM and existing HCRFs, which has widely been used for sequential FER.

TABLE III
CLASSIFICATION RESULTS OF SH-FER FOR EXPRESSION CLASSIFICATION AT THE SECOND LEVEL USING: (A) CK+ DATASET (B) JAFFE DATASET, (C) EXTENDED YALE B FACE DATASET, AND (D) MMI DATASET OF FACIAL EXPRESSIONS. (UNIT: %).

	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	**97**	1	0	1	1	0
Sad	2	**96**	1	0	1	0
Anger	0	1	**97**	1	0	1
Disgust	0	1	0	**98**	1	0
Surprise	0	1	0	1	**98**	0
Fear	1	0	3	0	2	**95**
Average			**96.83**			

(A)

	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	**96**	1	0	1	1	0
Sad	1	**97**	1	1	0	0
Anger	0	1	**98**	0	0	1
Disgust	1	0	2	**95**	2	0
Surprise	1	1	0	1	**97**	0
Fear	0	0	2	2	1	**95**
Average			**96.33**			

(B)

	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	**97**	2	0	1	0	0
Sad	1	**96**	0	2	1	0
Anger	0	1	**95**	2	1	1
Disgust	0	1	0	**98**	1	0
Surprise	0	2	0	1	**95**	2
Fear	0	1	0	2	1	**95**
Average			**96.00**			

(C)

	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	**95**	2	0	1	1	1
Sad	1	**96**	1	0	2	0
Anger	0	1	**97**	0	0	2
Disgust	1	1	0	**97**	1	0
Surprise	0	2	0	1	**97**	0
Fear	0	2	0	1	1	**96**
Average			**96.33**			

(D)

D. Fourth Experiment

The recognition rates for the eight methods, chosen for this experiment, along with the SH-FER are summarized in Table VI. It can be seen that the SH-FER outperformed the existing methods. Thus, the proposed system shows significant potential in its ability to accurately and robustly recognize human facial expressions using video data.

VI. CONCLUSION

Over the past two decades, FER systems have received a great deal of attention from the research community due to their application in many areas of pattern recognition and computer vision. However, recognizing human facial expressions accurately is still a major concern. This lack of accuracy can

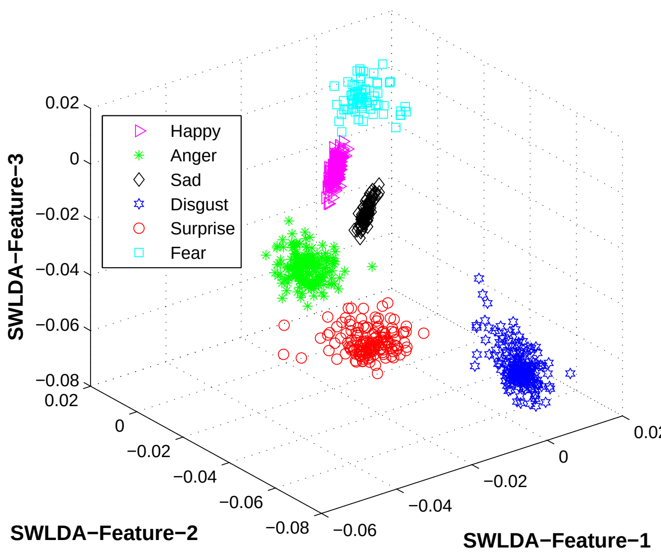

Fig. 8. 3D feature plots for the six expressions after applying SWLDA at the second level of SH-FER on JAFFE dataset.

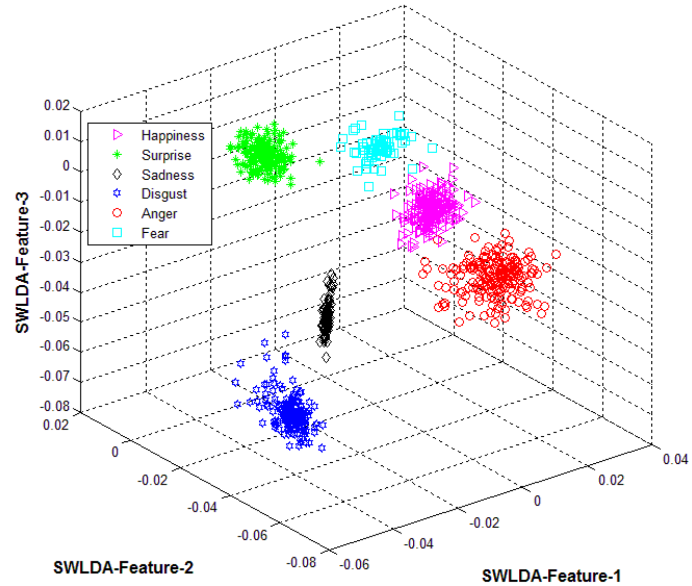

Fig. 10. 3D feature plots for the six expressions after applying SWLDA at the second level of SH-FER on MMI dataset.

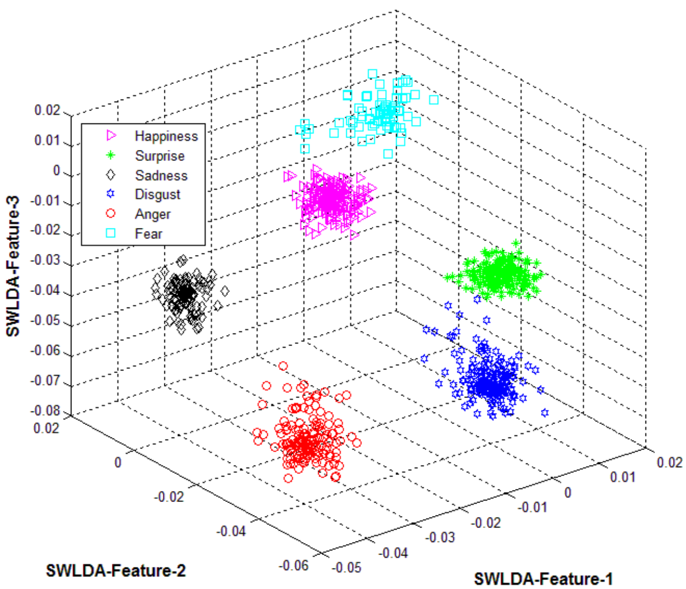

Fig. 9. 3D feature plots for the six expressions after applying SWLDA at the second level of SH-FER on Extended Yale B face dataset.

be attributed to various causes, such as the failure to extract prominent features, and the high similarity among different facial expressions that results due to the presence of low between-class variance in the feature space.

Accordingly, the purpose of this study was to propose an accurate and robust FER system, called SH-FER, which is capable of exhibiting high recognition rate. The SH-FER uses SWLDA and HCRF as its feature extraction and classification techniques, respectively. SWLDA helps the system in extracting the most significant features thereby reducing the high within class variance and increasing the low between class variance. HCRF then uses these features to accurately classify the human facial expressions. This model is capable of approximating the complex distributions using a mixture of full covariance Gaussian density functions.

The proposed SH-FER system has been validated using four publicly available datasets. Each dataset consisted of six facial expressions, i.e., happy, sad, surprise, disgust, anger, and fear, performed by different people, and each expression was composed of several sequence of expression frames. All of these experiments were performed in the laboratory using offline validation. Though the system was very successful in recognizing each of the six expressions in all of these experiments with a very high accuracy, its performance in real environment is yet to be investigated. The system performance could degrade in real-life tests, especially when used with various face angles and clutter (unnecessary objects in a test image). To resolve these issues, a real-time and robust segmentation technique would be required. Moreover, in a real environment, the facial frames may have different angles (different side views). Therefore, further research is required to maintain and improve the same recognition rate with different facial angles and clutter.

Since SH-FER employs two-level recognition with SWLDA and HCRFs at each level, this might lead to complexity issues. As described before, in the proposed HCRF model we explicitly involve full covariance Gaussian distribution instead of diagonal distribution and we believe that the proposed HCRF could take more time to train as compared to the existing HCRF. However, the proposed HCRF model also showed significant improvement over existing work in terms of recognition accuracy.

One solution could be to use a light weight classifier, like K-nearest neighbor (K-NN), at the first level; however, k-NN has some limitations. For example, it is very sensitive to noise and to the presence of inappropriate parameters as well. Therefore, further research is required in order to investigate ways to maintain the high recognition rate of the SH-FER

TABLE IV
CLASSIFICATION RESULTS OF SH-FER FOR THE SECOND EXPERIMENT. (A) TRAINING ON EXTENDED YALE B FACE DATASET AND TESTING ON CK+, JAFFE, AND MMI DATASETS, (B) TRAINING ON CK+ DATASET AND TESTING ON JAFFE, EXTENDED YALE B, AND MMI DATASETS, (C) TRAINING ON JAFFE DATASET AND TESTING ON CK+, EXTENDED YALE B, AND MMI DATASETS, (D) TRAINING ON MMI DATASET AND TESTING ON CK+, EXTENDED YALE B, AND JAFFE DATASETS (UNIT: %).

	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	**87**	3	2	2	3	3
Sad	3	**89**	2	1	3	2
Anger	0	2	**90**	2	4	2
Disgust	0	4	3	**89**	1	3
Surprise	3	2	5	4	**80**	6
Fear	1	2	5	3	2	**87**
Average			**87.0**			

(A)

	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	**89**	4	2	3	2	0
Sad	3	**91**	3	2	1	0
Anger	2	3	**90**	0	2	3
Disgust	0	2	4	**91**	2	1
Surprise	1	0	5	3	**88**	3
Fear	2	0	4	3	4	**87**
Average			**89.3**			

(B)

	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	**79**	7	4	3	2	5
Sad	6	**80**	4	4	3	3
Anger	2	3	**83**	5	3	4
Disgust	0	2	4	**90**	1	3
Surprise	1	5	2	4	**85**	3
Fear	2	6	3	3	4	**82**
Average			**83.1**			

(C)

	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	**86**	4	3	3	1	3
Sad	3	**82**	5	3	4	3
Anger	3	1	**84**	2	6	4
Disgust	3	0	4	**88**	3	2
Surprise	4	2	3	0	**89**	2
Fear	2	3	10	4	6	**75**
Average			**84.0**			

(D)

while improving its efficiency at the same time.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2013-067321)).
This work was supported by the Industrial Strategic Technology Development Program (10035348, Development of a Cognitive Planning and Learning Model for Mobile Platforms) funded by the Ministry of Knowledge Economy (MKE, Korea).

TABLE V
CLASSIFICATION RESULTS OF SH-FER FOR THE THIRD EXPERIMENT USING EXTENDED COHN-KANADE (CK+) DATASET OF FACIAL EXPRESSIONS. (A) ICA WITH HCRF, INSTEAD OF USING SWLDA, (B) ICA+LDA+HCRF, (C) SWLDA WITH EXISTING HCRF [59], INSTEAD OF USING THE PROPOSED HCRF MODULE (UNIT: %).

	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	**87**	3	2	4	2	2
Sad	3	**89**	2	2	3	1
Anger	2	1	**90**	2	3	2
Disgust	2	2	1	**89**	4	2
Surprise	2	0	4	3	**87**	2
Fear	2	3	4	2	4	**85**
Average			**87.83**			

(A)

	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	**88**	4	3	2	2	1
Sad	3	**93**	1	2	0	1
Anger	3	2	**90**	1	1	3
Disgust	2	3	2	**89**	3	1
Surprise	2	1	2	3	**90**	2
Fear	0	3	2	3	4	**88**
Average			**89.66**			

(B)

	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	**88**	2	0	3	4	3
Sad	0	**91**	1	2	4	2
Anger	2	1	**92**	2	0	3
Disgust	0	2	3	**93**	0	2
Surprise	3	1	3	2	**90**	1
Fear	2	1	3	2	3	**89**
Average			**90.50**			

(C)

TABLE VI
COMPARISON OF SH-FER WITH EXISTING FER SYSTEMS USING THE DATASETS MENTIONED IN THEIR RESPECTIVE PAPERS: (A) CK+, (B) EXTENDED YALE B, (C) MMI, AND (D) JAFFE DATASETS (UNIT: %).

Existing Works	**[44]**	**[70]**	**SH-FER**
Recognition Rates	89	90	**96**

(A)

Existing Works	**[71]**	**[72]**	**SH-FER**
Recognition Rates	90	87	**96**

(B)

Existing Works	**[73]**	**[12]**	**SH-FER**
Recognition Rates	73	86	**96**

(C)

Existing Works	**[74]**	**[75]**	**SH-FER**
Recognition Rates	84	92	**96**

(D)

REFERENCES

[1] M. H. Siddiqi, F. Farooq, and S. Lee, "A robust feature extraction method for human facial expressions recognition systems," in *Proceedings of the 27th Conference on Image and Vision Computing New Zealand*, pp. 464–468, ACM, 2012.

[2] M. H. Siddiqi, A. M. Khan, T. C. Chung, and S. Lee, "A precise recognition model for human facial expressions recognition system," in *Proceedings of the 26th IEEE Canadian Conference on Electrical and Computer Engineering*, IEEE, 2013.

[3] M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Measuring facial expressions by computer image analysis," *Psychophysiology*, vol. 36, no. 2, pp. 253–263, 1999.

[4] A. Mehrabian, "Communication without words," *Psychological today*, vol. 2, pp. 53–55, 1968.

[5] V. Bettadapura, "Face expression recognition and analysis: the state of the art," *arXiv preprint arXiv:1203.6722*, 2012.

[6] C. Lisetti and C. LeRouge, "Affective computing in tele-home health:

design science possibilities in recognition of adoption and diffusion issues," in *In Proceedings 37th IEEE Hawaii International Conference on System Sciences, Hawaii, USA*, 2004.

[7] X. Wu and J. Zhao, "Curvelet feature extraction for face recognition and facial expression recognition," in *Natural Computation (ICNC), 2010 Sixth International Conference on*, vol. 3, pp. 1212–1216, IEEE, 2010.

[8] S. Moore and R. Bowden, "The effects of pose on facial expression recognition," in *Proceedings of the British Machine Vision Conference*, pp. 1–11, 2009.

[9] Z. Zhu and Q. Ji, "Robust real-time face pose and facial expression recovery," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1, pp. 681–688, IEEE, 2006.

[10] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: machine learning and application to spontaneous behavior," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, pp. 568–573, IEEE, 2005.

[11] Y.-L. Tian, T. Kanade, and J. F. Cohn, "Facial expression analysis," in *Handbook of face recognition*, pp. 247–275, Springer, 2005.

[12] C. Shan, S. Gong, and P. W. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, vol. 27, no. 6, pp. 803–816, 2009.

[13] C. Shan and R. Braspenning, "Recognizing facial expressions automatically from video," in *Handbook of ambient intelligence and smart environments*, pp. 479–509, Springer, 2010.

[14] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced p300 speller performance," *Journal of neuroscience methods*, vol. 167, no. 1, pp. 15–21, 2008.

[15] M. Ager, Z. Cvetkovic, and P. Sollich, "Phoneme classification in high-dimensional linear feature domains," *arXiv preprint arXiv:1312.6849*, pp. 1–12, 2013.

[16] M. Nusseck, D. W. Cunningham, C. Wallraven, and H. H. Bülthoff, "The contribution of different facial regions to the recognition of conversational expressions," *Journal of vision*, vol. 8, no. 8, 2008.

[17] K. Kaulard, D. W. Cunningham, H. H. Bülthoff, and C. Wallraven, "The mpi facial expression databasea validated database of emotional and conversational facial expressions," *PloS one*, vol. 7, no. 3, p. e32321, 2012.

[18] K. L. Schmidt and J. F. Cohn, "Human facial expressions as adaptations: Evolutionary questions in facial expression research," *American journal of physical anthropology*, vol. 116, no. S33, pp. 3–24, 2001.

[19] X. Wang and K. K. Paliwal, "Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition," *Pattern recognition*, vol. 36, no. 10, pp. 2429–2439, 2003.

[20] Y. Pang, Y. Yuan, and X. Li, "Iterative subspace analysis based on feature line distance," *Image Processing, IEEE Transactions on*, vol. 18, no. 4, pp. 903–907, 2009.

[21] G. Aguilar-Torres, K. Toscano-Medina, G. Sanchez-Perez, M. Nakano-Miyatake, and H. Perez-Meana, "Eigenface-gabor algorithm for feature extraction in face recognition," *International Journal of Computers*, vol. 3, no. 1, pp. 20–30, 2009.

[22] S. R. V. Kittusamy and V. Chakrapani, "Facial expressions recognition using eigenspaces," *Journal of Computer Science*, vol. 8, no. 10, pp. 1674–1679, 2012.

[23] J. Kalita and K. Das, "Recognition of facial expression using eigenvector based distributed features and euclidean distance based decision making technique," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 4, no. 2, pp. 196–202, 2013.

[24] Z. Abidin and A. Harjoko, "A neural network based facial expression recognition using fisherface," *International Journal of Computer Applications*, vol. 59, no. 3, 2012.

[25] V. J. Mistry and M. M. Goyani, "A literature survey on facial expression recognition using global features," *International Journal of Engineering and Advanced Technology*, vol. 2, no. 4, pp. 653–657, 2013.

[26] F. Long, T. Wu, J. R. Movellan, M. S. Bartlett, and G. Littlewort, "Learning spatiotemporal features by using independent component analysis with application to facial expression recognition," *Neurocomputing*, vol. 2, no. 1, pp. 126–132, 2012.

[27] K. Karande, S. Talbar, and S. Inamdar, "Face recognition using oriented laplacian of gaussian (olog) and independent component analysis (ica)," in *Digital Information and Communication Technology and it's Applications (DICTAP), 2012 Second International Conference on*, pp. 99–103, IEEE, 2012.

[28] M. Kaur, R. Vashisht, and N. Neeru, "Recognition of facial expressions with principal component analysis and singular value decomposition," *International Journal of Computer Applications*, vol. 9, no. 12, pp. 36–40, 2010.

[29] A. Halder, A. Jati, G. Singh, A. Konar, A. Chakraborty, and R. Janarthanan, "Facial action point based emotion recognition by principal component analysis," in *Proceedings of the International Conference on Soft Computing for Problem Solving (SocProS 2011) December 20-22, 2011*, pp. 721–733, Springer, 2012.

[30] M. Z. Uddin, T.-S. Kim, and B. C. Song, "An optical flow feature-based robust facial expression recognition with hmm from video," *International Journal of Innovative Computing, Information and Control*, vol. 9, no. 4, pp. 1409–1421, 2013.

[31] L. Ma, "Facial expression recognition using 2-d dct of binarized edge images and constructive feedforward neural networks," in *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*, pp. 4083–4088, IEEE, 2008.

[32] M. Kumbhar, A. Jadhav, and M. Patil, "Facial expression recognition based on image feature," *International Journal of Computer and Communication Engineering*, vol. 1, no. 2, pp. 117–119, 2012.

[33] S. Joseph, "Ecological and biochemical studies on cyanobacteria of cochin estuary and their application as source of antioxidants," 2012.

[34] S. Chitra and D. G. Balakrishnan, "A survey of face recognition on feature extraction process of dimensionality reduction techniques," *Journal of Theoretical and Applied Information Technology*, vol. 36, no. 1, pp. 92–100, 2012.

[35] S. C. Christabel, M. Annalakshmi, M. D. P. Winston, *et al.*, "Facial feature extraction based on local color and texture for face recognition using neural network," *International Journal of Science and Engineering Applications*, vol. 2, no. 4, pp. 78–82, 2013.

[36] L. Zhang and D. Tjondronegoro, "Feature extraction and representation for face recognition," *Face Recognition, Milos Oraves (Ed.),*, pp. 1–20, 2010.

[37] S. Z. Li, X. W. Hou, H. J. Zhang, and Q. S. Cheng, "Learning spatially localized, parts-based representation," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, pp. I–207, IEEE, 2001.

[38] W. Gu, C. Xiang, Y. Venkatesh, D. Huang, and H. Lin, "Facial expression recognition using radial encoding of local gabor features and classifier synthesis," *Pattern Recognition*, vol. 45, no. 1, pp. 80–91, 2012.

[39] I. Buciu and I. Pitas, "Application of non-negative and local non negative matrix factorization to facial expression recognition," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 1, pp. 288–291, IEEE, 2004.

[40] S. Zhang, X. Zhao, and B. Lei, "Facial expression recognition based on local binary patterns and local fisher discriminant analysis," *WSEAS Transactions on Signal Processing*, vol. 8, no. 1, pp. 21–31, 2012.

[41] T. Ahsan, T. Jabid, U. Chong, *et al.*, "Facial expression recognition using local transitional pattern on gabor filtered facial images," *IETE Technical Review*, vol. 30, no. 1, pp. 47–52, 2013.

[42] T. Jabid, M. H. Kabir, and O. Chae, "Robust facial expression recognition based on local directional pattern," *ETRI journal*, vol. 32, no. 5, pp. 784–794, 2010.

[43] M. H. Kabir, T. Jabid, and O. Chae, "A local directional pattern variance (ldpv) based face descriptor for human facial expression recognition," in *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, pp. 526–532, IEEE, 2010.

[44] A. Ramirez Rivera, J. Rojas Castillo, and O. Chae, "Local directional number pattern for face analysis: Face and expression recognition," *Image Processing, IEEE Transactions on*, vol. 22, no. 5, pp. 1740–1752, 2013.

[45] Y. Rahulamathavan, R. Phan, J. Chambers, and D. Parish, "Facial expression recognition in the encrypted domain based on local fisher discriminant analysis," vol. 4, no. 1, pp. 83–92, 2013.

[46] Z. Wang and X. Sun, "Manifold adaptive kernel local fisher discriminant analysis for face recognition," *Journal of Multimedia*, vol. 7, no. 6, pp. 387–393, 2012.

[47] C. S Patil and A. J Patil, "A review paper on facial detection technique using pixel and color segmentation," *International Journal of Computer Applications*, vol. 62, no. 1, pp. 21–24, 2013.

[48] M. Gargesha and P. Kuchi, "Facial expression recognition using artificial neural networks," *Artif. Neural Comput. Syst*, pp. 1–6, 2002.

[49] J. V. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *Journal of clinical epidemiology*, vol. 49, no. 11, pp. 1225–1231, 1996.

[50] H.-H. Tsai, Y.-S. Lai, and Y.-C. Zhang, "Using svm to design facial expression recognition for shape and texture features," in *Machine Learning and Cybernetics (ICMLC), 2010 International Conference on*, vol. 5, pp. 2697–2704, IEEE, 2010.

[51] "Support vector machines (svms)." http://scikit-learn.org/stable/modules/svm.html, 2011. (Last visited in Sunday 15 December 2013).

[52] W.-f. Liu, J.-l. Lu, Z.-f. Wang, and H.-j. Song, "An expression space model for facial expression analysis," in *Image and Signal Processing, 2008. CISP'08. Congress on*, vol. 2, pp. 680–684, IEEE, 2008.

[53] M. Schels and F. Schwenker, "A multiple classifier system approach for facial expressions in image sequences utilizing gmm supervectors," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, pp. 4251–4254, IEEE, 2010.

[54] T. Bouwmans, F. El Baf, *et al.*, "Modeling of dynamic backgrounds by type-2 fuzzy gaussians mixture models," *MASAUM Journal of of Basic and Applied Sciences*, vol. 1, no. 2, pp. 265–276, 2009.

[55] M. Z. Uddin, J. Lee, and T.-S. Kim, "An enhanced independent component-based human facial expression recognition from video," *Consumer Electronics, IEEE Transactions on*, vol. 55, no. 4, pp. 2216–2224, 2009.

[56] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," pp. 282–289, 2001.

[57] S. Kumar and M. Hebert, "Discriminative fields for modeling spatial dependencies in natural images," NIPS, 2003.

[58] A. Quattoni, S. Wang, L.-P. Morency, M. Collins, and T. Darrell, "Hidden conditional random fields," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 10, pp. 1848–1852, 2007.

[59] A. Gunawardana, M. Mahajan, A. Acero, and J. C. Platt, "Hidden conditional random fields for phone classification," in *Proc. Interspeech*, vol. 2, pp. 1117–1120, Citeseer, 2005.

[60] M. Mahajan, A. Gunawardana, and A. Acero, "Training algorithms for hidden conditional random fields," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 1, pp. I–273–I–276, IEEE, 2006.

[61] S. Reiter, B. Schuller, and G. Rigoll, "Hidden conditional random fields for meeting segmentation," in *Multimedia and Expo, 2007 IEEE International Conference on*, pp. 639–642, IEEE, 2007.

[62] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. Mullers, "Fisher discriminant analysis with kernels," in *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop.*, pp. 41–48, IEEE, 1999.

[63] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural computation*, vol. 12, no. 10, pp. 2385–2404, 2000.

[64] S. Mika, *Kernel fisher discriminants*. PhD thesis, Universitätsbibliothek, 2002.

[65] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 711–720, 1997.

[66] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pp. 94–101, IEEE, 2010.

[67] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pp. 200–205, IEEE, 1998.

[68] K.-C. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 5, pp. 684–698, 2005.

[69] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pp. 5–pp, IEEE, 2005.

[70] D. Ghimire and J. Lee, "Geometric feature-based facial expression recognition in image sequences using multi-class adaboost and support vector machines," *Sensors*, vol. 13, no. 6, pp. 7714–7734, 2013.

[71] S. Du and R. K. Ward, "Adaptive region-based image enhancement method for robust face recognition under variable illumination conditions," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 20, no. 9, pp. 1165–1175, 2010.

[72] P. Forczmański, G. Kukharev, and N. Shchegoleva, "Simple and robust facial portraits recognition under variable lighting conditions based on two-dimensional orthogonal transformations," in *Image Analysis and Processing–ICIAP 2013*, pp. 602–611, Springer, 2013.

[73] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang, and D. N. Metaxas, "Learning active facial patches for expression analysis," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 2562–2569, IEEE, 2012.

[74] X. Zhao and S. Zhang, "Facial expression recognition using local binary patterns and discriminant kernel locally linear embedding," *EURASIP journal on Advances in signal processing*, vol. 2012, no. 1, pp. 1–9, 2012.

[75] Y. Rahulamathavan, R. Phan, S. Veluru, K. Cumanan, and M. Rajarajan, "Privacy-preserving multi-class support vector machine for outsourcing the data classification in cloud," *Dependable and Secure Computing, IEEE Transactions on*, 2013.

Muhammad Hameed Siddiqi is PhD student in Ubiquitous Computing (UC) Lab, Department of Computer Engineering, Kyung Hee University, South Korea. He did his Master of Engineering from the Department of Computer Engineering, Kyung Hee University, South Korea in 2012, and Bachelor of Computer Science (Hons) from Islamia College University of Peshawar, N-W.F.P, Pakistan in 2007. He was a Graduate Assistant at Universiti Teknologi PETRONAS, Malaysia from 2008 to 2009. His research interest focuses on Image Processing, Pattern Recognition, Machine Intelligence, Activity Recognition, and Facial Expression Recognition.

Rahman Ali is a PhD student in Ubiquitous Computing Laboratory (UCLab), Department of Computer Engineering, Kyung Hee University, South Korea. He got his M.Phil Degree in Computer Science from Department of Computer Science, University of Peshawar, Pakistan in 2009. He secured his M.Sc Degree (in Computer Science) from Hazara University, Mansehra, Pakistan in 2005 and Bachelor Degree (in Computer Science) from Govt. Jahanzeb College Saidu Sharif, Swat, Pakistan back in 2002. He has been a lecturer in Computer Science at University of Peshawar since 2009. He also served Institute of Information Technology, University of Science and Technology, Bannu as a lecturer in computer science and the Laboratoire d'Informatique de l'Universit du Maine, France as a research assistant. His current research includes Machine Learning, Knowledge Acquisition and Reasoning.

Adil Mehmood Khan received his Ph.D. degree from the Department of Computer Engineering of Kyung Hee University, Republic of Korea in 2011. He is now working as a faculty member with the Department of Computer Science, Innopolis University, Kazan, Russia. His research interest includes pattern recognition, signal processing, ubiquitous computing, and machine learning.

Young-Tack Park received his B.S. fromSeoul National University, Korea in 1978. He got M.S. and Ph.D. degrees in Computer Science from KAIST Korea and University of Illinois at Urbana-Champaign, USA in 1980 and 1992, respectively. Since 1993, he has been a professor at School of Computing, Soongsil University, Korea.

Sungyoung Lee received his B.S. from Korea University, Seoul, Korea. He got his M.S. and Ph.D. degrees in Computer Science from Illinois Institute of Technology (IIT), Chicago, USA in 1987 and 1991 respectively. He has been a professor in the department of Computer Engineering, Kyung Hee University, Korea since 1993. He is a founding director of the Ubiquitous Computing Laboratory, and has been affiliated with a director of Neo Medical ubiquitous-Life Care Information Technology Research Center, Kyung Hee University since 2006. Before joining Kyung Hee University, he was an assistant professor in the Department of Computer Science, Governors State University, llinois, USA from 1992 to 1993. His current research focuses on Ubiquitous Computing and Applications, Wireless Ad-hoc and Sensor Networks, Context-aware Middleware, Sensor Operating Systems, Real-Time Systems and Embedded Systems, Activity and Emotion Recognition. He is a member of ACM and IEEE.