IETE Technical Review
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/titr20

Depth Camera-Based Facial Expression Recognition System Using Multilayer Scheme

Muhammad Hameed Siddiqi[a], Rahman Ali[a], Abdul Sattar[b], Adil Mehmood Khan[c] & Sungyoung Lee[a]

[a] Department of Computer Engineering, Kyung Hee University, Suwon, 446-701, Republic of Korea

[b] Department of Biomedical Engineering, Kyung Hee University, Suwon, 446-701, Republic of Korea

[c] Division of Information and Computer Engineering, Ajou University, Suwon, 443-749, Republic of Korea

Published online: 14 Aug 2014.

To cite this article: Muhammad Hameed Siddiqi, Rahman Ali, Abdul Sattar, Adil Mehmood Khan & Sungyoung Lee (2014) Depth Camera-Based Facial Expression Recognition System Using Multilayer Scheme, IETE Technical Review, 31:4, 277-286, DOI: 10.1080/02564602.2014.944588

To link to this article: http://dx.doi.org/10.1080/02564602.2014.944588

Depth Camera-Based Facial Expression Recognition System Using Multilayer Scheme

Muhammad Hameed Siddiqi[1], Rahman Ali[1], Abdul Sattar[2], Adil Mehmood Khan[3] and Sungyoung Lee[1]

[1]Department of Computer Engineering, Kyung Hee University, Suwon, 446-701, Republic of Korea, [2]Department of Biomedical Engineering, Kyung Hee University, Suwon, 446-701, Republic of Korea, [3]Division of Information and Computer Engineering, Ajou University, Suwon, 443-749, Republic of Korea

ABSTRACT

The analysis of a facial expression in telemedicine and healthcare plays a significant role in providing sufficient information about patients such as stroke and cardiac in monitoring their expressions for better management of their diseases. Facial expression recognition (FER) improves the level of interaction between human-to-human communications. The human face has a major contribution for such types of communications, which consists of lips, eyes and forehead that are considered the most informative features for FER. There are some parameters that make FER a challenging task that includes high similarity among different expressions that makes it difficult to distinguish these expressions with a high accuracy. Moreover, most of the previous works used existing available standard datasets and all the datasets were pose-based datasets, and they have some privacy issues because of utilizing video (RGB) cameras. Accordingly, this work presents a multilayer scheme for FER to handle these issues. In the proposed FER system, we utilized a depth camera in order to solve the privacy concerns, and the accuracy of this camera is not affected by any kind of environmental parameters. Similarly, the depth camera automatically detects and extracts the faces based on the distance between the camera and subject. For global and local feature extraction, principal component analysis (PCA) and independent component analysis (ICA) were used. A hierarchical classifier was used, where the expression category was recognized at the first level, followed by the actual expression recognition at the second level. For the entire experiments, an *n*-fold cross-validation scheme (based on subjects) was employed. The proposed FER system achieved a significant improvement in accuracy (98.0%) against the existing methods.

Keywords:
Facial expressions, Depth camera, Principal component analysis, Independent component analysis, Linear discriminant analysis, Hidden Markov model.

1. INTRODUCTION

Cure provisioning to patients by remotely analysing their facial expressions can be promising breakthrough development in telemedicine and healthcare domains. This will assist clinicians in their decision-making process by monitoring patients' facial expressions remotely, specifically in disease such as stroke. A concept of telestroke was introduced in [1] that explains a case study for monitoring acute stroke patient using telemedicine. Stroke patients' state can be defined using facial expression identification techniques and facial exercises can be recommended. Sad and happy facial expressions can be identified for stroke patients and therefore guidelines can be remotely provided for better care of patient. Another example is that psychiatrist can use telemedicine technology for treating a patient with post-traumatic stress disorder (PTSD) by monitoring his/her facial expressions remotely [2]. Study has even been carried out to monitor heart-failure patients using telemedicine [3], a prime candidate for a facial expression analysis. These developments will facilitate clinicians and physicians for efficiently monitoring and management of patient disease. In summary, facial expression recognition (FER) is an observable indication of person's sentimental state, mental activity and behaviour [4].

Telemedicine and healthcare applications that employ video technologies raise privacy concerns since it can lead to situations where subjects may not know that their private information is being shared and thus become exposed to a threat [5]. Unlike RGB-cameras, depth-cameras only capture the depth information and do not reveal the identity of the subject or other sensitive information, which makes them a superior choice over RGB-cameras [6]. Therefore, we choose the depth-camera over RGB-cameras for the proposed FER system. To the best of our knowledge, there are no any sufficient works which have been done to study the expression recognition with a depth camera.

Over the past decade, human FER has emerged as an important research area. Human FER systems can be classified into two categories: pose-based expression recognition systems [7−9] and spontaneous expression recognition systems [10−12]. Pose-based expressions are the artificial expressions produced by people when they are asked to do so [13]. Similarly, spontaneous expressions are those that people give out spontaneously, and they can be observed on a day-to-day basis, such as during conversations or while watching movies [13]. The focus of this work is pose-based.

Commonly, there are three basic modules in a typical FER system: preprocessing, feature extraction and recognition. Preprocessing module diminishes illumination and other light effects to increase the recognition accuracy. Feature extraction module deals with getting the distinguishable features of each expression and quantizing it as a discrete symbol. While in recognition module, a classifier such as hidden Markov model (HMM), or Gaussian mixture model (GMM) or support vector machine (SVM) is first trained with training data and then used to generate labels for the expressions in the incoming video data.

Several factors make FER a challenging research problem. These include varying light conditions in training and test images; and high similarity among different expressions that makes it difficult to distinguish these expressions with a high accuracy. Uddin et al. [14] have proposed a complete approach for FER systems that provided high classification accuracy for the depth database of facial expressions. In their work, they employed independent component analysis (ICA) for feature extraction. Once extracted, features were subject to generalized discriminant analysis (GDA) to find the most relevant features. The result after applying GDA was fed to an HMM. The recognition rate of their technique was 97.08%. They applied specific training and testing strategy; however, if we change their training and training scheme, then their work failed in exhibiting the same accuracy. Low accuracy in these new experiments could be attributed to the following two reasons. First, in some of the datasets (like our dataset) of facial expressions, the subjects have worn glasses and some subjects have beard that make it difficult to extract useful features from some parts of the face, such as the eyes and lips. Second, most of the expressions share high similarity, and thus their features overlap significantly in the feature space. Uddin et al. [14] applied GDA to the extracted feature space to improve the class separation among different classes with the assumption that the variance is distributed uniformly among all the classes. However, this is not the case; for example, expressions like happiness and sadness are very similar to each other but can easily be distinguished from anger and fear (another pair with high similarity).

Accordingly, this work implements a multilayer scheme-based FER that is capable of performing accurate FER across depth dataset. In the proposed FER system, first, we utilized an Intel® Creative depth camera that automatically removes background from the expression frame (based on the distance between the camera and subject). Second, principal component analysis (PCA) and ICA were used for feature extraction. Finally, a hierarchical classifier was used, where the expression category was recognized at the first level, followed by the actual expression recognition at the second level. The proposed FER system has been validated using our own created depth dataset of facial expressions, and, therefore, succeeded in providing high recognition accuracy across depth dataset. To the best of our knowledge, very limited works have been done on FER using depth data.

The rest of the paper is organized as follows. Section 2 discusses some related work about FER. The proposed FER system is described in detail in Section 3. Then the experimental setup for the proposed FER system is described in Section 4. Section 5 presents the experimental results along with some discussion on the results and talks about the factors that could degrade the performance of the system if tested in real-life scenarios. Finally, the paper is concluded with some future directions in Section 6.

2. RELATED WORK

Automatic FER has become an important research area for many applications from last two decades. As described earlier typical FER system has three basic modules such as preprocessing, feature extraction and recognition. For each module, lots of work has been done in order to improve the accuracy of such FER systems. However, most of them suffered from low accuracy using depth dataset. Preprocessing module removes the environmental parameters and improves the quality of the expression frames; however, these parameters do not affect the accuracy of the depth camera. Moreover, the depth camera has the capability to automatically detect and extract the faces from the expression frames based on distance between the camera and subject.

There, a huge amount of works have been done for feature extraction module to improve the accuracy of FER systems. However, most of them have their own limitations. These methods include nearest feature line-based subspace analysis [15], eigenfaces and eigenvectors [16], [17], fisherfaces [18] and global features [19].

However, all these holistic methods do not know what exact facial features are the most important for FER systems. Moreover, these methods ignore higher order correlation value and might not work if the data sources are dependent [20].

Moreover, some local feature-based methods have been proposed to compute the local descriptors from some parts of the face and then integrate this information into one descriptor. These methods include local feature analysis (LFA) [21], Gabor features [22], and local binary pattern (LBP) [23]. Among these methods, LBP achieved better performance. However, LBP does not provide the directional information of the facial frame [24]. Some recent works have proposed solutions to the limitations of LBP. These methods include local transitional pattern (LTP) [25], local directional pattern (LDP) [26] and local directional pattern variance (LDPv) [27]. Most of these methods exploited other information instead of employing intensity to overcome the problems due to noise and illumination change [28]. However, the performance of these methods still degrades in non-monotonic illumination change, noise variation, change in pose and expression conditions [29].

As for the recognition module, several classifiers have been investigated. The authors of [29] employed artificial neural networks (ANNs) to recognize different types of facial. However, an ANN is a black box and has incomplete capability to explicitly categorize possible fundamental relationships [30]. Moreover, the FER systems proposed in [31,32] used support vector machines (SVMs). However, in SVMs, the probability is calculated using indirect techniques; in other words, there is no direct estimation of the probability, these are calculated by employing five-fold cross-validation due to which SVM suffers from the lack of classification [33]. Similarly, in [34,35], GMMs were employed to recognize different types of facial expressions. As stated earlier, the features could be very sensitive to noise; therefore, fast variations in the facial frames cannot be modelled by GMMs and produce problems for sensitive detection [36]. HMMs are mostly used to handle sequential data when frame-level features are used. In such cases, other vector-based classifiers, such as GMMs, ANNs and SVMs, have difficulty in learning the dependencies in a given sequence of frames. Due to this capability, some well-known FER systems, including [37–39], utilized HMM as a classifier. In conclusion, a large number of feature extraction techniques and classifiers have been employed for video-based FER systems. Among them, PCA and ICA have been the most widely used feature extraction techniques, and HMMs have been the most commonly used classifier.

Accordingly, this work presents a multilayer scheme in order to solve the limitations of the existing works. In this work, at the first level, a set of PCA and ICA were first applied to extract the features from all the classes and then HMM was utilized to recognize the three expression categories. Once the category of the given expression has been determined, the label for the expression within the recognized category is recognized at the second level. For this purpose, PCA and ICA were applied separately as a feature extraction technique to each category and the result was used to train three HMMs, one HMM per category.

3. PROPOSED FACIAL EXPRESSION RECOGNITION SYSTEM

3.1 Pre-processing

Preprocessing module is used to diminish the environmental effects due to which the quality of the expression frames is improved. This module also improves the efficiency of the FER systems. In order to enhance the expression frames in this module, lots of existing methods such as histogram equalization, median filter and homomorphic filter can be applied. However, this module does not affect the accuracy of the depth cameras.

3.2 Feature Extraction

Feature extraction deals with getting the distinguishable features of each expression and quantizing it as a discrete symbol. There, lots of techniques have been proposed and validated for feature extraction for FER systems. Among them, PCA and ICA are the most commonly used methods, and their performance has already been validated in [39]. Therefore, we decided to use PCA and ICA for feature extraction to extract both the global and local features, respectively.

3.2.1 Principal Component Analysis (PCA)

PCA is applied to extract the global features. PCA is a second-order approach that offers an easy way of reducing a complex set of data by assigning it onto a space with a small dimension while protecting as much of the unpredictability as possible. PCA fabricates the best linear least-squares decomposition of a training set. It is the most common feature extraction technique that has been employed in FER systems. This method has the benefit of being linear and makes no hypothesis concerning the data distribution. The role of PCA is to estimate the original data with lower dimensional features, which represents the data economically. It also focuses on the global features of the grey-scale faces. In this case, there is a strong correlation among observed variables. For this work, the

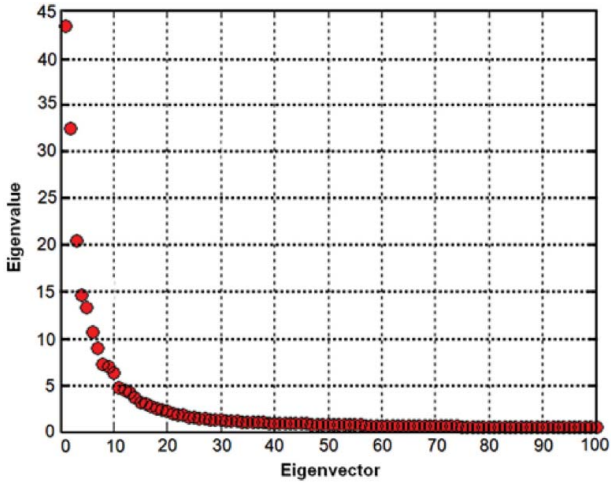

Figure 1: Top 100 eigenvalues along with their corresponding eigenvectors.

main purpose for using PCA was to express the large one-dimensional (1D) vector of pixels constructed from the 2D image into the compact principal components of the feature space. This is called eigenspace projection. The primary job of PCA is to compute the eigenvectors of the covariance data matrix M and then, by the combination of a few top eigenvectors, the approximation is done. Top 100 eigenvectors along with their eigenvalues are shown in Figure 1, where a total of 600 facial image vectors were considered for PCA. For further reading on PCA, please refer to [40].

However, PCA is an unsupervised technique that locates PCs at the optimally diminished dimension of the input. For FER, it only focuses on the global information and extracts the global features from the whole face image, which results in low accuracy. Furthermore, PCA yields uncorrelated components. If the data have a Gaussian distribution, the uncorrelated components are independent. However, if the data are merged non-Gaussian components, then PCA fails to extract components having non-Gaussian distributions [41]. Therefore, for accurate classification, we need local features as well, which are extracted by utilizing ICA.

3.2.2 *Independent Component Analysis (ICA)*

ICA is a technique used to seek independent components from multivariate statistical data. ICA assumes that the underlying sources are linearly mixed and statistically independent. General implementations of ICA can be found in the literature [42,43].

If we assume that the sources are denoted by

$$S(t) = [s_1(t), s_2(t), \ldots, s_m(t)]^T \tag{1}$$

and the multichannel observations are denoted by

$$X(t) = [x_1(t), x_2(t), \ldots, x_m(t)]^T, \tag{2}$$

then the linear mixture can be represented by

$$x_j = a_{j1}s_1 + a_{j2}s_2 + \cdots + a_{jn}s_n \quad \text{for all } j \tag{3}$$

or can be written as

$$X = As, \tag{4}$$

where the matrix A of size $n \times m$ represents linear memory-less mixing channels. The statistical model presented in Eq. (4) is called the independent component analysis or ICA model. The ICA model is a generative model, i.e. it describes how the observed data are generated by a process of mixing the components S_k. The independent components are latent variables, i.e. they cannot be directly observed. Moreover, the mixing matrix is assumed to be unknown. We observe the random vector x, and we must estimate both A and s by using it.

ICA starts with the very simple assumption that the components S_k are statistically *independent*. It will be seen below that we must also assume that the independent components have *non-Gaussian* distributions. However, in the basic model, we do *not* assume that these distributions are known, but note that if they are known, the problem is considerably simplified. For simplicity, we assume that the unknown mixing matrix is square, but this assumption can sometimes be relaxed. Then, after estimating the matrix A, we can compute its inverse, W, and obtain the independent component simply by

$$S = Wx, \tag{5}$$

where $W = [w_1, w_2, \ldots, w_n]$ is the matrix of size of $n \times m$. In general, ICA assumes that the number of channels is equal to the number of independent sources, i.e. $n = m$. Thus, n channels of data are decomposed into n ICs.

We note that, in the ICA model, the time index t is dropped, as seen in Eqs. (4) and (5). We assume that each mixture x_j and each independent component S_k are random variables, instead of a proper time signal.

3.3 Expression Category Recognition

The proposed FER system is based on the theory that different expressions can be grouped into three categories based on the muscles movements: first category (lips-based movement), e.g. happy and sad; second

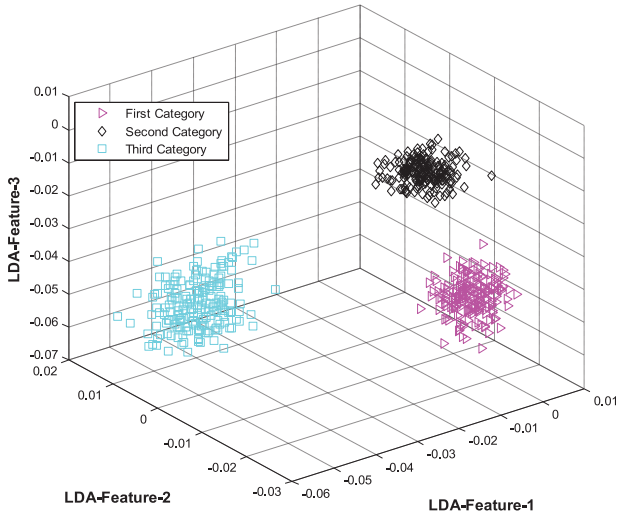

Figure 2: 3D plot of the proposed FER system at the first level of classification, i.e. category classification.

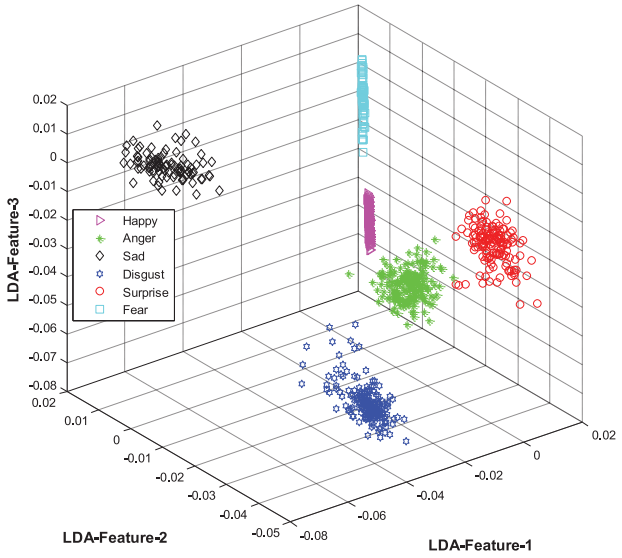

Figure 3: 3D plot of the proposed FER system at the second level of classification, i.e. expressions classification.

category (lips−eyes-based movement), e.g. surprise and disgust; and third category (lips−eyes−forehead-based movement), e.g. anger and fear. In each category, the corresponding parts of face have much contribution in expressions making. Therefore, in the proposed FER system, an expression is classified into one of these three categories at the first level, then, at the second level, classifier (trained for the recognized category) is employed to give a label to this expression within this category.

At the first level, linear discriminant analysis (LDA) was first applied to the extracted features from all the classes and an HMM was trained to recognize the three expression categories: first category, second category and third category expressions. The LDA-features for these three categories are shown in Figure 2.

A clear separation could be seen among the categories, and this is why the proposed FER system achieved 100% recognition accuracy at the first level.

3.4 Expressions Recognition

Once the category of the given expression has been determined, then at the second level, the label of the expression has been determined within that category. For this purpose, LDA was applied separately to the feature space of each category and the result was used to train three HMMs, one HMM per category. Collectively, the overall results for all the expression classes are shown in Figure 3.

It can be seen from Figure 3 that the proposed FER system achieved better expressions classification results.

4. EXPERIMENTAL SETUP

We used our own created depth dataset of facial expressions in order to assess the performance of the proposed FER system. The dataset displays frontal view of the face and each expression is composed of several sequences of expression frames. During each experiment, we reduced the size of each input image (expression frame) to 60 × 60, where the images were first converted to a zero-mean vector of size 1 × 3600 for feature extraction. We utilized all the six expressions for the whole experiments that were performed in Matlab using an Intel® Pentium® Dual-Core™ (2.5 GHz) with a RAM capacity of 3 GB.

In this dataset, there were 25 subjects (university students) that performed six expressions such as happy, anger, sad, disgust, surprise and fear. The age ranges of the subjects were from 25 to 35 years old and most of them were male. All the expressions were in frontal view, meaning that the depth camera was normal to the subjects. Each subject performed six expressions and each expression contained of 15 expression frames. For the recognition purpose, all the expression frames were utilized from each expression sequence, which resulted in a total of 2250 expression images. For a thorough validation, four experiments were performed in the following way.

1. In the first experiment of the proposed FER system, an n-fold cross-validation rule (based on subjects) was utilized, meaning that out of n subjects, data from a single subject were taken as the validation data for testing the proposed FER system, whereas

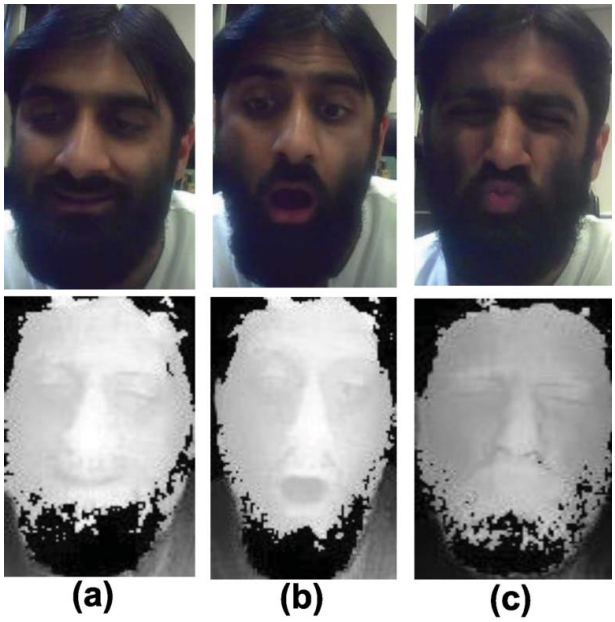

Figure 4: Sample RGB (top row) and depth (down row) images of (a) happy, (b) surprise and (c) disgust expressions captured by depth camera.

the data for the remaining $n-1$ subjects were used as the training data. This process was repeated n times, with data from each subject used exactly once as validation data. The value of n varied according to the dataset used. The benefit of this rule is that each expression was used for both training and testing.

2. In the second experiment of the proposed FER system, the performance of the sub-component, i.e. a multilayer scheme (hierarchical), was analysed.
3. In the third experiment, the performance of the proposed FER system was compared with previous existing methods.
4. In the last experiment, the performance of different approaches with different combination was analysed against the proposed FER system.

The depth sample images are shown in Figure 4.

5. RESULTS AND DISCUSSION

In the first experiment of the proposed FER system, an n-fold cross-validation rule (based on subjects) was utilized, meaning that out of n subjects, data from a single subject were taken as the validation data for testing the proposed FER system, whereas the data for the remaining $n-1$ subjects were used as the training data. This process was repeated n times, with data from each subject used exactly once as validation data. The value of n varied according to the dataset used. The benefit of this rule is that each expression was used for both training and testing. The overall results of the proposed FER system at the first and second levels of classifications are shown in Tables 1 and 2, respectively.

Table 1: Confusion matrix of the proposed FER system at the first level of classification on depth database of facial expressions (unit: %)

Expressions category	Classification rate
First category	97
Second category	97
Third category	99
Average	97.67

Table 2: Confusion matrix of the proposed FER system at the second level of classification on depth database of facial expressions (unit: %)

Expressions	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	98	0	1	0	1	0
Sad	1	97	0	2	0	0
Anger	1	1	97	0	1	0
Disgust	1	0	0	99	0	0
Surprise	0	1	0	0	98	1
Fear	1	0	0	0	0	99
Average				98.00		

It can be seen from Tables 1 and 2 that the proposed FER system achieved better recognition results on both levels of classification when applied on depth dataset.

In the second experiment of the proposed FER, the effectiveness of the proposed multilayer scheme was analysed. This experiment was performed under the absence of a multilayer scheme (hierarchical LDA and HMMs), meaning that the experiment was performed by using single HMM. The overall results are shown in Table 3.

It can be seen from Table 3 that the proposed FER system does not achieve a high recognition rate, meaning that the multilayer scheme has much contribution and

Table 3: Confusion matrix of the proposed FER system on depth database of facial expressions while removing the multilayer (hierarchical recognition) scheme (unit: %)

Expressions	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	87	2	3	4	3	1
Sad	1	90	2	2	3	2
Anger	4	3	84	4	3	2
Disgust	1	2	3	88	3	3
Surprise	4	4	2	1	88	1
Fear	1	4	3	1	0	91
Average				88.00		

Table 4: Comparison results of the proposed FER system against some existing state-of-the-art methods on depth dataset for six facial expressions under the same settings as described in Section 4 (unit: %)

Existing work	[14]	[18]	[25]	[28]	[39]	Proposed FER system
Average accuracy rate	88	79	80	90	84	98

is the main module for high recognition accuracy of the proposed system. These results support the theory that the problem of high similarity among the features of different expressions is a local problem. In other words, the features exist in the form of groups in the overall feature space. The expressions within one group are very similar, whereas they are easily distinguishable from those in the other groups; therefore, to overcome this problem in an effective manner, these groups (or expression categories) should be separated first and then techniques like LDA should be applied to each category separately.

In the third experiment, the performance of the proposed FER system has been compared against recent existing methods. These methods have been implemented by us under the same guidelines which were provided in their respective manuscripts. An n-fold cross-validation scheme was utilized, i.e. out of n subjects, data from a single subject were retained as the validation data for testing the proposed system, whereas the data for the remaining $n-1$ subjects were used as the training data. This process was repeated n times, where the value of n varied according to the dataset used. The weighted average recognition rates for all state-of-the-art methods are indicated in Table 4

It can be seen from Table 4 that the proposed FER system has achieved best recognition result than of the existing state-of-the-art methods. All the experiments for the proposed FER system and for the existing methods have been performed in laboratory (offline validation) by using our own depth dataset, and it can be seen from the experiments that proposed FER system achieved a high recognition rate on depth dataset.

Table 5: Confusion matrix of ICA and single-HMM using depth dataset of facial expressions (unit: %)

Expressions	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	49.9	11	9.9	12	9	8.2
Sad	8.9	51.3	13	8.9	6.9	11
Anger	14	10.7	50	9.8	5.7	9.8
Disgust	5.1	8.9	9	60	11	6
Surprise	9.2	7.2	11	8.6	56	8
Fear	10.1	8.9	9.4	5.6	13	53
Average				53.36		

Table 6: Confusion matrix of ICA+PCA and single-HMM using depth dataset of facial expressions (unit: %)

Expressions	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	69.1	9.1	6.2	5.5	5.8	4.3
Sad	10	65.5	4.8	7.9	5.8	6
Anger	3.3	5.1	75	3.4	5.2	8
Disgust	4.3	7.5	6.5	66.7	8.9	6.1
Surprise	5.7	3.2	6.8	9	71	4.3
Fear	3.6	6.1	8.1	4.4	4.8	73
Average				70.05		

Table 7: Confusion matrix of ICA+LDA and single-HMM using depth dataset of facial expressions (unit: %)

Expressions	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	75.7	7.2	4.7	0	5.2	7.2
Sad	6.9	73	5.2	5.9	4.3	4.7
Anger	4.6	5.2	71.2	4.8	8	6.2
Disgust	6.7	0	8.4	69.9	9	6
Surprise	3	3.2	3.8	6	80	4
Fear	0	0	8.9	6.6	7.8	76.7
Average				74.41		

Table 8: Confusion matrix of PCA+ICA+LDA and single-HMM using depth dataset of facial expressions (unit: %)

Expressions	Happy	Sad	Anger	Disgust	Surprise	Fear
Happy	86.1	6	2.5	3.4	2	0
Sad	5	82	4	3	3	3
Anger	1	2	86	3	2	6
Disgust	0	0	0	90.1	9.9	0
Surprise	2	2	2	6	84	4
Fear	2	3	15	2	3	75
Average				83.87		

In the last experiment, a set of experiments were performed using different combinations of various previously used feature extraction and classification approaches on depth dataset. The overall results for these tests are shown in Tables 5–8, respectively

6. CONCLUSION

Communication through facial expressions plays a significant role in telemedicine, and social interactions. Recently, automatic FER using depth data has received a lot of attention. Several FER systems have been proposed; however, recognizing human facial expressions accurately is still a major concern for most of these systems on depth data. The human face has a major contribution for such types of communications, which consists of lips, eyes and forehead that are considered the most informative features for expressions recognition. There are some parameters that make FER a

challenging task that includes high similarity among different expressions that makes it difficult to distinguish these expressions with a high accuracy and there are also some privacy issues because of utilizing video (RGB) cameras. Unlike the previous systems, the proposed FER system utilized a depth camera, which does not affect the lighting parameters, and for global and local feature extraction, we used PCA and ICA. Finally, we used a multilayer scheme to overcome the problem of high similarity among different expressions. This work is based on the theory that different expression can be grouped into three categories based on the muscles movements: first category (lips-based movement), second category (lips−eyes-based movement) and third category (lips−eyes−forehead-based movement). Therefore, in the proposed FER system, an expression is classified into one of these three categories and LDA coupled with HMM has been used to find one of these categories at the first level. Then, at the second level, the label for an expression within the recognized category is recognized using a separate set of LDA and HMM, trained just for that category. The proposed FER system has been validated and tested on our own created depth dataset. There are 25 subjects that performed six basic expressions and each expression video has 15 expression frames. For the recognition purpose, all the expression frames were utilized from each expression sequence, which resulted in a total of 2250 expression images. For the entire experiments of the proposed FER system, an n-fold cross-validation rule (based on subjects) was utilized. The benefit of this rule is that each expression was used for both training and testing. This is one of the limitations of the previous works, i.e. if we swap their training and testing strategy, then they cannot achieve better results. The whole experiments were performed in laboratory (offline validation). Therefore, further research is needed to employ this work either in real environments of healthcare or in smart home environments.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) [grant number 2013−067321]. This research was also supported by the MSIP (Ministry of Science, ICT & Future Planning), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2014-(H0301-14-1003)).

REFERENCES

1. B. M. Demaerschalk, M. L. Madeline, T. E. Kiernan, B. J. Bobrow, D. A. Corday, K. E. Wellik, and M. I. Aguilar, et al., "Stroke telemedicine," *Mayo Clin. Proc.*, Vol. 84, no. 1, pp. 53−64, Jan. 2009.
2. Telemedicine: Extending Specialist Care to Rural Areas, Newsletter Article, CISCO. Available: http://www.cisco.com/web/strategy/docs/gov/fedbiz081810healthpresence.pdf
3. I. H. Kraai, M. L. Luttik, R. M. de Jong, T. Jaarsma, and H. L. Hillege, "Heart failure patients monitored with telemedicine: Patient satisfaction, a review of the literature," *J. Cardiac Failure*, Vol. 17, no. 8, pp. 684−90, Aug. 2011.
4. J. M. Bogdan, W. Quan, and L.-K. Shark, "Facial expression recognition," in *Biometrics − Unique and Diverse Applications in Nature, Science, and Technology*, Midori Albert, Ed. 2011, pp. 57−88, ISBN: 978-953-307-187-9, InTech.
5. R. Rusyaizila, Z. Nasriah, and S. Putra, "Privacy issues in pervasive healthcare monitoring system: A review," *International Science Index,* Vol. 4, no. 12, pp. 640−646. Dec. 2010, pp. 741−7.
6. M. H. Siddiqi, A. M. Khan, and S.-W. Lee, "Active contours level set based still human body segmentation from depth images for video-based activity recognition," *Trans. Internet Inf. Syst.*, Vol. 7, no. 11, pp. 2829−52, Nov. 2013.
7. 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. 253−63, Mar. 1999.
8. X. Wu, and J. Zhao, "Curvelet feature extraction for face recognition and facial expression recognition," in *2010 Sixth International Conference on Natural Computation (ICNC)*, Vol. 3, Yantai: IEEE, 10−12 Aug. 2010, pp. 1212−6.
9. S. Moore, and R. Bowden, "The effects of pose on facial expression recognition," in *Proceedings of the British Machine Vision Conference*, London, 7−10 Sep. 2009, pp. 1−11.
10. Z. Zhu, and Q. Ji, "Robust real-time face pose and facial expression recovery," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, New York: IEEE, 17−22 Jun. 2006, pp. 681−8.
11. 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 2005 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, Vol. 2, San Diego: IEEE, 20−26 Jun. 2005, pp. 568−73.
12. Y.-L. Tian, T. Kanade, and J. F. Cohn, "Facial expression analysis," in S.Z. Li and A.K. Jain (Ed.), *Handbook of Face Recognition*, New York: Springer, 2005, pp. 247−75.
13. A. Mehrabian, "Communication without words," *Psychol. Today*, Vol. 2, pp. 53−5, Sep. 1968.
14. M. Z. Uddin, "A depth video-based facial expression recognition system," *IETE Tech. Rev.*, Vol. 29, no. 2, pp. 169−78, Mar. 2012.
15. Y. Pang, Y. Yuan, and X. Li, "Iterative subspace analysis based on feature line distance," *IEEE Trans. Image Process.*, Vol. 18, no. 4, pp. 903−7, Mar. 2009.
16. S. Ragavan, V. Kittusamy, and V. Chakrapani, "Facial expressions recognition using eigenspaces," *J. Comput. Sci.*, Vol. 8, no. 10, pp. 1674−9, Aug. 2012.
17. J. Kalita, and K. Das, "Recognition of facial expression using eigenvector based distributed features and euclidean distance based decision making technique," *In. J. Adv. Comput. Sci. Appl.*, Vol. 4, no. 2, pp. 196−202, Feb. 2013.
18. Z. Abidin, and A. Harjoko, "A neural network based facial expression recognition using fisherface," *Int. J Comput. Appl.*, Vol. 59, no. 3, pp. 30−4, Dec. 2012.
19. V. J. Mistry, and M. M Goyani, "A literature survey on facial expression recognition using global features," *Int. J. Eng. Adv. Technol.*, Vol. 2, no. 4, pp. 653−7, Apr. 2013.
20. S. Chitra, and D. G. Balakrishnan, "A survey of face recognition on feature extraction process of dimensionality reduction techniques," *J. Theor. Appl. Inf. Technol.*, Vol. 36, no. 1, pp. 92−100, Feb. 2012.

21. S. Z. Li, X. W. Hou, H. J. Zhang, and Q. S. Cheng, "Learning spatially localized, parts-based representation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* Kauai, HI, Vol. 1, pp. 207–12, 8–14 Dec. 2001.
22. W. Gu, C. Xiang, Y. V. Venkatesh, D. Huang, and H. Lin, "Facial expression recognition using radial encoding of local gabor features and classifier synthesis," *Pattern Recognit.*, Vol. 45, no. 1, pp. 80–91, Jan. 2012.
23. S. Zhang, X. Zhao, and B. Lei, "Facial expression recognition based on local binary patterns and local fisher discriminant analysis," *WSEAS Trans. Signal Process.*, Vol. 8, no. 1, pp. 21–31, Jan. 2012.
24. M. H. Siddiqi, F. Farooq, and S. Lee, "A robust feature extraction method for human facial expressions recognition systems," in *27th Conference on Image and Vision Computing New Zealand*, New Zealand, 26–28 Nov. 2012, pp. 464–8.
25. T. Ahsan, T. Jabid, and U. P, Chong, et al., "Facial expression recognition using local transitional pattern on gabor filtered facial images," *IETE Tech. Rev.*, Vol. 30, no. 1, pp. 47–52, Jan. 2013.
26. L. D. Introna, and D. Wood, "Picturing algorithmic surveillance: The politics of facial recognition systems," *Surveillance Soc.*, Vol. 2, no. 2/3, pp. 177–198, 2004.
27. M. H. Kabir, T. Jabid, and O. Chae, "A local directional pattern variance (ldpv) based face descriptor for human facial expression recognition," in *7th International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Boston, 29 August – 1 September, 2010, pp. 526–32.
28. A. R. Rivera, J. R. Castillo, and O. Chae, "Local directional number pattern for face analysis: Face and expression recognition," *IEEE Trans. Image Process.*, Vol. 22, no. 5, pp. 1740–52, May. 2013.
29. D. Filko, and M. Goran, "Emotion recognition system by a neural network based facial expression analysis," *Autom.: J. Control Meas. Electron. Comput. Commun.*, Vol. 54, no. 2, pp. 263–72, 2013.
30. J. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes." *J. Clin. Epidemiol.*, Vol. 49, no. 11, pp. 1225–31. Nov. 1996.
31. G. Kharat, and S. Dudul, "Human emotion recognition system using optimally designed SVM with different facial feature extraction techniques," *WSEAS Trans. Comput.*, Vol. 7, no. 6, pp. 650–9, Jun. 2008.
32. C. Shan, S. Gong, and P. McOwan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image Vis. Comput.*, Vol. 27, no. 6, pp. 803–16, May. 2009.
33. Support Vector Machines. Available: http://scikit-learn.org/stable/modules/svm.html (accessed on Thursday 1 May 2014).
34. W. Liu, J. Lu, Z Wang, and H. Song, "An expression space model for facial expression analysis," in *Congress on Image and Signal Processing*, Sanya, Hainan, Vol. 2, 27–30 May. 2008, pp. 680–4.
35. M. Schels, and F. A. Schwenker, "Multiple classifier system approach for facial expressions in image sequences utilizing GMM supervectors," In *20th International Conference on Pattern Recognition*, Istanbul, 23–26 Aug. 2010, pp. 4251–4.
36. T. Bouwmans, and F. El. Baf, "Modeling of dynamic backgrounds by type-2 fuzzy Gaussians mixture models," *MASAUM J. Basic Appl. Sci.*, Vol. 1, no. 2, pp. 265–76, Feb. 2010.
37. M. Yeasin, B. Bullot, and R. Sharma, "Recognition of facial expressions and measurement of levels of interest from video," *IEEE Trans. Multimedia*, Vol. 8, no. 3, pp. 500–8, Jun. 2006.
38. P. S. Aleksic, and A. K. Katsaggelos, "Automatic facial expression recognition using facial animation parameters and multistream HMMs," *IEEE Trans. Inf. Forensics Secur.*, Vol. 1, no. 1, pp. 3–11, Mar. 2006.
39. M. Z. Uddin, J. J. Lee, and T.-S. Kim, "An enhanced independent component-based human facial expression recognition from video," *IEEE Trans. Consum. Electron.*, Vol. 55, no. 4, pp. 2216–24, Nov. 2009.
40. I. T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer, 2002.
41. I. Buciu, C. Kotropoulos, and I. Pitas, "Comparison of ICA approaches for facial expression recognition," Signal *I*mage *V*ideo *P*rocess., Vol. 3, no. 4, pp. 345–61, Dec. 2009.
42. A. Hyvarinen, and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Comput.*, Vol. 9, no. 7, pp. 1483–92, Oct. 1997.
43. A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: John Wiley & Sons, 2001.

Authors

Muhammad Hameed Siddiqi is a PhD student in the Ubiquitous Computing (UC) Lab, Department of Computer Engineering, Kyung Hee University, South Korea. He completed his Master's degree in Engineering at the Department of Computer Engineering, Kyung Hee University, South Korea in 2012, and received his Bachelor's (Hons) degree in Computer Science 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 interests are image processing, pattern recognition, machine intelligence and activity and expression recognition.

E-mail: siddiqi@oslab.khu.ac.kr

Rahman Ali is a PhD student in the Ubiquitous Computing Laboratory (UCLab), Department of Computer Engineering, Kyung Hee University, South Korea. He got his MPhil degree in Computer Science from the Department of Computer Science, University of Peshawar, Pakistan in 2009. He secured his MSc degree (in Computer Science) from Hazara University, Mansehra, Pakistan in 2005 and his Bachelor's 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'Université du Maine, France as a research assistant. His current research interests are machine learning, knowledge acquisition, and reasoning.

E-mail: rahmanali@oslab.khu.ac.kr

Abdul Sattar received the BS degree in telecommunication engineering from Mehran University of Engineering and Technology, Jamshoro Sindh Pakistan, in 2008. Currently, he is working toward the MS leading PhD degree at the U-Health Lab Biomedical Department, Kyung Hee University, Suwon, South Korea. His research interests include embedded systems, u-healthcare systems, smart health and Internet of Things in health.

E-mail: sattar.abdul@.khu.ac.kr

Adil Mehmood Khan received his PhD 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 Division of Information and Computer Engineering, Ajou University, Republic of Korea. His research interests include pattern recognition, signal processing, ubiquitous computing, and machine learning.

E-mail: amtareen@ajou.ac.kr

Sungyoung Lee received his BS from Korea University, Seoul, Korea. He got his MS and PhD 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, Ilinois, 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.

E-mail: sylee@oslab.khu.ac.kr

DOI: 10.1080/02564602.2014.944588;