J Electr Eng Technol.2017; 12(4): 1657-1662
http://doi.org/10.5370/JEET.2017.12.4.1657

ISSN(Print) 1975-0102
ISSN(Online) 2093-7423

Facial Expression Recognition using 1D Transform Features and Hidden Markov Model

Ahmad Jalal*, Shaharyar Kamal† and Daijin Kim**

Abstract – Facial expression recognition systems using video devices have emerged as an important component of natural human-machine interfaces which contribute to various practical applications such as security systems, behavioral science and clinical practices. In this work, we present a new method to analyze, represent and recognize human facial expressions using a sequence of facial images. Under our proposed facial expression recognition framework, the overall procedure includes: accurate face detection to remove background and noise effects from the raw image sequences and align each image using vertex mask generation. Furthermore, these features are reduced by principal component analysis. Finally, these augmented features are trained and tested using Hidden Markov Model (HMM). The experimental evaluation demonstrated the proposed approach over two public datasets such as Cohn-Kanade and AT&T datasets of facial expression videos that achieved expression recognition results as 96.75% and 96.92%. Besides, the recognition results show the superiority of the proposed approach over the state of the art methods.

Keywords: Facial expression, 1D transform, Hidden Markov model

1. Introduction

Facial expression recognition (FER) is considered to be one of the fundamental topics of computer vision and attracted by a lot of researchers in the past few decades. It has been widely used for many real-world applications [1-4] such as automatic action detection, identity authentication, face mask generation, human-computer interaction and surveillance systems [5-7]. Human beings use facial expression to express their emotions and also provide important cues during the communication such as showing how much they are taking interest during communication, waiting for their turn to speak in social interaction and understanding the information by giving continuous feedback. Usually facial expressions contribute 55 percent of the effect while communicating and hence it makes major part of the human daily life communication. The basic purpose of FER system is to recognize the face expression [8] correctly among several different expressions in the database.

Recently, several researches developed many effective facial features representation and extraction techniques from the original images to successfully recognize the facial expressions. In [9], non-verbal multi-model has been proposed, which examine the facial expression of human based on pixel differences. To classify data, they used Gaussian process classification. However, all these methods are very complex to implement, as well as computationally expensive. Thus, we need more robust and less computational processing system for FER. While, other intensity-based cameras [10-15] play vital role in many real-world applications and are promoted towards FER.

In this method, we propose a novel feature extraction method for FER using sequence of facial images. Initially, we remove noisy effects in facial sequential images and develop more robust, efficient and compact features as 1D transform features. However, these features are further computed by principal component analysis (PCA) for global features properties and reduced dimensions. Our proposed feature extraction method has less computational cost and it is effective to recognize the different facial expressions. Moreover, our experimental results proved that the proposed method outperforms state of the art methods.

The rest of the paper is structured as follows. In section 2 we explain the basic architecture of our FER system by considering feature extraction techniques, data clustering and training/testing using HMM. Experimental settings and results are shown in section 3. Finally, we conclude the paper in section 4.

2. System Methodology

The proposed FER system consists of incoming face images of various resolution and background. The preprocessing step is used to improve the quality of image

† Corresponding Author: Dept. of Electronics and Radio Engineering, Kyung Hee University, Korea. (shka@khu.ac.kr)
* Dept. of Computer Science, Air University, Islamabad, Pakistan. (ahmadjalal@mail.au.edu.pk)
** Dept. of Computer Science and Engineering, POSTECH, Korea. (dkim@postech.ac.kr)
Received: October 16, 2015; Accepted: April 5, 2017

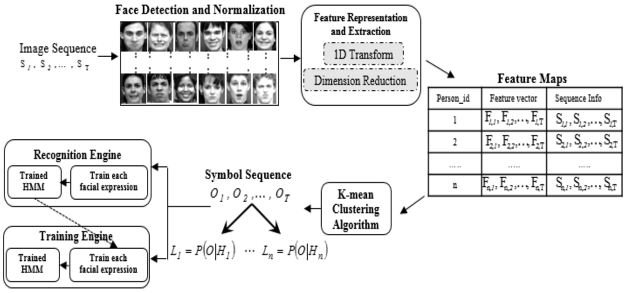

Fig. 1. Overall flow architecture of proposed FER system

by applying background subtraction, face detection, noise removal, face alignment and restricting the image size using region of interest (ROI). These detected face images are used for feature extraction based on 1D transform technique. Then, these features are further processed for dimension reduction via PCA followed by discriminating method for robust classification. Finally, sequential routines of human faces are trained/tested using HMMs. Fig. 1 shows the system architecture flow of the proposed FER system.

2.1 Image preprocessing

To perform the preprocessing steps of sequential facial expressions images, we applied local histogram equalization method for controlling various resolution, light changing conditions and removing noisy effects [16]. While, to detect the face region, we applied skin color detection mechanism to locate the exact face region. Also, average intensity variation method $I(v)$ is used to capture face region $f(r)$ and maximize the distance between the face f and the background region $b(r)$.

$$I(v) = \alpha f + (1-\alpha)\, b(r), \quad \alpha \in [0 \le r \le 1] \tag{1}$$

$$f(r) = \int_0^{\infty} \left| I(v) - r_{in} \right|^2 dxdy + \int_0^{\infty} \left| I(v) - r_{out} \right|^2 dxdy \tag{2}$$

where r_{in} and r_{out} are the average intensities inside and outside the boundary region of facial image I. Then, a face alignment approach [17] is considered by fixing the vertex mask generation to match the facial components such as nose, lips, and mouth to its geometrical, and position constraints. Each face region is controlled by vertices movement design for fitness function and hold eigenface regions [17] to control the facial movements ΔMV as

$$\Delta MV = [m_t + m_{t-1}] \sum_{i=0}^{n} \left(f_i(r) - b_i(r) \right) \tag{3}$$

where m is the specific facial components along t and t-1 having (m_x, m_y) values with respect to facial vertex matrix. The interested areas of vertex face masks are determined by dividing the overall eigenface values. Finally, the facial image is detected, aligned properly and further fed for feature generation phenomenon.

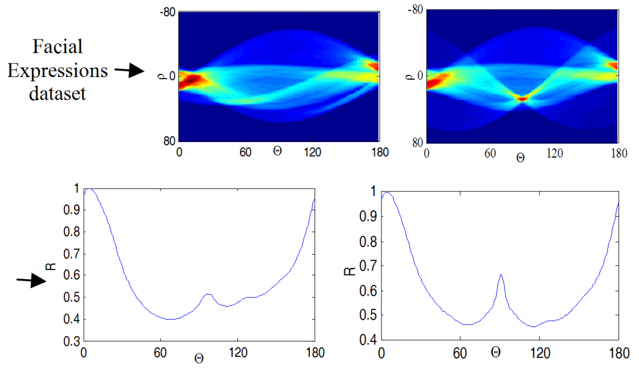

Fig. 2. Examples of 1D transform features using facial expressions dataset

2.2 1D Transform features

Radon transform is widely used in image applications and compute the total pixel intensities along lines of different directions [18]. However, center points of the facial images are selected as the reference points. To process the radon transform over facial images, a map between the image coordinate system $f(x, y)$ and the radon domain is created and defined as

$$RD(\rho,\theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\delta(\rho - x\cos\theta - y\sin\theta)\, dxdy \tag{4}$$

where RD is the line integral of the 2D radon function along a line from positive infinite to negative infinite values and the position of the arbitrary image pixels is determined by (ρ, θ). During R transform, the image data is represented in 1D profile which indicates the inner variations of facial expression. Hence, we used R transform of an image $f(x, y)$ as:

$$RT(\theta) = \int_{-\infty}^{\infty} RD^2(\rho,\theta)\, d\rho \tag{5}$$

where RT is defined as R transform and it is invariant to scaling and translation. Also, R transform possesses appearing properties which are important for shape representation during facial expression (see Fig. 2).

2.3 Dimension reduction and Feature discrimination

PCA is one of the famous predominant linear dimensionality reduction method which has been widely applied on datasets in all scientific domains [19]. The goal of PCA is to compress the features by extracting the most dominant information from the dataset. Whereas, to find PC, eigenvector with largest eigenvalue shows the axis of maximum variance and the second eigenvector shows the axis of second largest variance [20, 21]. We used 100 PCs from the R transform profile and the overall vector size become 1 x 100.

In order to discriminate features, we used Fisher's LDA method for encoding discriminatory information. LDA is processed by maximizing the ratio of determinant of the between scatter matrix S_b and the within-class scatter matrix S_w. The between class scatter matrix S_b and the within class scatter S_w can be defined as:

$$S_b = \sum_{j=1}^{c} I_j (\Psi_1 - \Psi_2)(\Psi_1 - \Psi_2)^T \tag{6}$$

$$S_w = \sum_{j=1}^{c} \sum_{\Psi \in c_j} (\Psi - \Psi_1)(\Psi - \Psi_1)^T \tag{7}$$

where I_j is the training samples in the j^{th} class, c is the number of distinct classes. Ψ_j is the mean vector of samples belonging to class j. These features are symbolized for codebook generation by considering k-mean algorithm.

2.4 Hidden Markov Model (HMM)

HMM is a set of statistical model having finite states used to characterize the transition probability and symbol observation probability of an image [22]. During HMM, we deal with different stages that includes a number of states, events, initial state distribution, state transition and discrete output matrix values [23]. To train each facial data using HMM, all parameters are set based on transition and observation symbols O probabilities for a slot (H_1, H_2,…, H_n) of trained models as

$$\alpha_t(i) = \pi_i b_i(O_1), \beta_t(i) = \sum_{j=1}^{r} a_{ij} b_j(O_{t+1}) \beta_{t+1}(i) \tag{8}$$

where r is the total states used in HMM model, α and β are the forward and background variables calculated from transition and observation matrix. All sequential data are stored as buffer handling concept [23, 24]. We used different number of states (i.e., 3 to 6) for training the HMMs and recognition (see Table 1). While, HMM is trained having a size of codebook of 64 [25].

During testing, trained HMMs are used for facial expression recognition (see Fig 3) having 4 states process.

Table 1. States of HMMs over facial datasets

Datasets	Different number of states for HMMs			
	3-state	4-state	5-state	6-state
Cohn-Kanade	89.24	96.75	82.04	69.38
AT&T	87.51	96.92	80.17	65.43

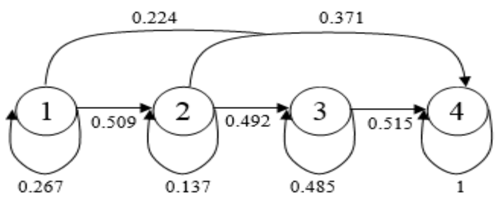

Fig. 3. Transition and observation probabilities of anger facial expression after training HMM

3. Experimental Results and Analysis

3.1 Experimental setting

To evaluate the performance of our proposed facial expression recognition system, we analyzed our experiments on Cohn-Kanade and AT&T datasets [26] having consecutive frames of different video sequences. These datasets consist of six universal expressions. During experimental results of both datasets, we used a cross-subject training/testing setup in which each subject taken out (i.e., leave-one-subject-out scheme) from the training set and repeat the experiment for each of them.

3.2 Features of PC and IC over facial datasets

Analyzing the characteristics and features performance of the proposed feature (i.e., 1D transform) with other conventional features (i.e., PC and IC features), the conventional features showed 2D shape representation while the proposed feature provides 1D transform. Fig. 4 shows the samples of PC basis images of facial images using all six typical facial expressions.

However, for the 3D plot representation of PCs features,

Fig. 4. Four PCs of all six facial expressions

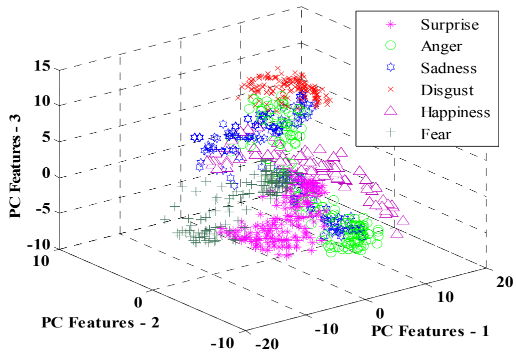

Fig. 5. 3D plot of LDA on PC features having six facial expressions images

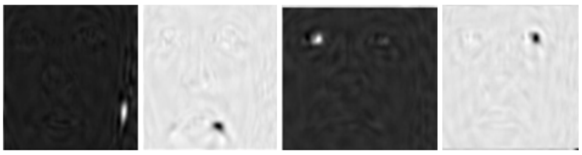

Fig. 6. Sample ICs of all six facial expressions using facial datasets

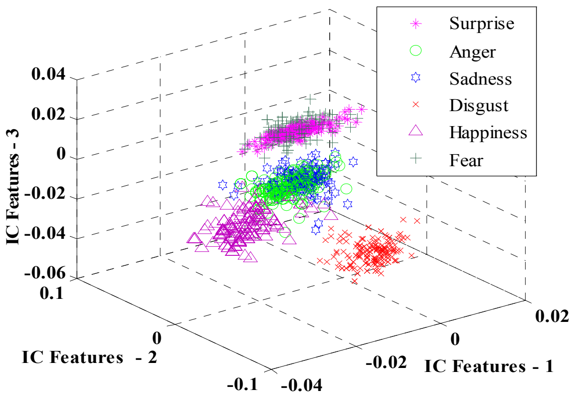

Fig. 7. 3D plot of LDA on IC features having all six facial expressions images

we introduced LDA for discrimination of each category as shown in Fig. 5.

As LDA on PC features applied on facial images, most of the facial expressions such as anger, sadness, happiness and fear are quite merged with each other whereas the overall features do not become prominent for FER.

Next, IC features are used to extract local characteristics of facial expression images. Fig. 6 shows the samples of IC basis images of facial images using all six typical facial expressions.

While, Fig. 7 describes the 3D plot representation of the facial images using IC features. In this Figure, the overall classification of features need further improvements. Thus, we proposed 1D transform features.

3.3 Features of 1D transform over facial datasets

Whereas, the feature representation using radon transform technique provides the 2D projections along with facial expression variations are shown in Fig. 2. Finally, Fig. 8 shows the significant discrimination of all six different facial expressions using 3D plot view.

3.4 Comparison of Recognition accuracy of all methods using Cohn-Kanade and AT&T datasets

During this section, we compared the recognition results of conventional systems along with the proposed system using both Cohn-Kanade and AT&T datasets. As shown in Table 2, the recognition rate using LDA on PCA is 86.67% acting as the lowest recognition rate.

Table 2. Recognition results of LDA on PC features

Expression	SE	AR	SN	DT	HS	FR
SE	85.0	9.5	0	0	4.0	1.5
AR	13.0	79.0	5.5	0	1.5	1.0
SN	5.5	0	88.0	3.0	0	3.5
DT	2.0	0	4.5	91.0	0	2.5
HS	1.5	0	2.5	0	92.5	3.5
FR	2.5	5.5	1.5	0	6.0	84.5
Mean Recognition Rate = 86.67%						

*SE=Surprise; AR=Anger; SN=Sadness; DT=Disgust; HS=Happiness; FR=Fear.

Table 3. Recognition results of LDA on IC features

Expression	SE	AR	SN	DT	HS	FR
SE	90.0	5.5	0	0	0	4.5
AR	6.0	92.0	0	0	2.0	0
SN	2.0	0	90.5	7.5	0	0
DT	0	0	4.5	95.5	0	0
HS	1.5	4.0	0	0	93.0	1.5
FR	0	0	4.5	2.0	5.0	88.5
Mean Recognition Rate = 91.58%						

Table 4. Recognition results with proposed features using Cohn-Kanade facial expression dataset

Expression	SE	AR	SN	DT	HS	FR
SE	94.5	2.5	0	0	3.0	0
AR	0	96.0	2.50	0	0	1.5
SN	0	0	98.5	0	0	1.5
DT	0	0	2.0	98.0	0	0
HS	0	0	2.5	0	97.5	0
FR	1.0	3.0	0	0	0	96.0
Mean Recognition Rate = 96.75%						

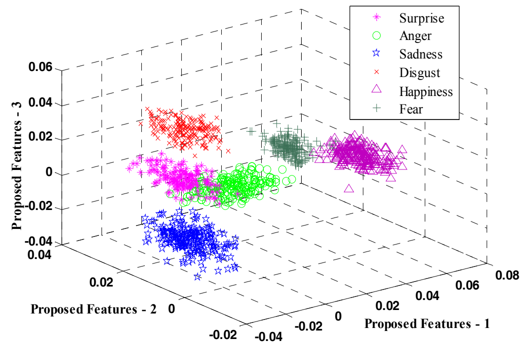

Fig. 8. 3D plot of LDA on 1D transform features

Thus, we applied the LDA on ICA technique to extract the ICs from the datasets of Cohn-Kanade dataset as shown in Table 3.

Considering the same setting in Table 4, we obtained best recognition rate using proposed system as 96.75%.

Similarly, to consider the AT&T dataset as input image sequences, we achieved 88.42% and 92.08% recognition as shown in Table 5 and 6.

Finally, Table 7 disclosed that the recognition rate of our proposed LDA based 1D transform features approach proved highest recognition rate of 96.92% upto all other conventional approaches using AT&T dataset.

Table 5. Recognition results with LDA on PC features

Expression	SE	AR	SN	DT	HS	FR
SE	82.5	8.0	0	6.5	0	3.0
AR	3.5	85.0	0	2.5	9.0	0
SN	0	6.5	91.5	0	2.0	0
DT	4.0	0	1.0	89.5	3.5	2.0
HS	0	2.5	0	3.5	94.0	0
FR	5.0	0	0	3.5	3.5	88.0
Mean Recognition Rate = 88.42%						

Table 6. Recognition results with LDA on IC features

Expression	SE	AR	SN	DT	HS	FR
SE	87.5	3.0	0	2.5	5.0	2.0
AR	4.5	89.5	3.5	0	0	2.5
SN	0	2.5	93.0	2.5	1.0	1.0
DT	3.5	2.0	0	92.5	2.0	0
HS	0	0	2.0	1.5	95.5	1.0
FR	2.0	1.0	0	2.5	0	94.5
Mean Recognition Rate = 92.08%						

Table 7. Recognition results with proposed features using AT&T facial expressions dataset

Expression	SE	AR	SN	DT	HS	FR
SE	96.5	0	1.0	1.0	0	1.5
AR	0	93.0	4.0	1.0	2.0	0
SN	0	2.0	97.0	0	1.0	0
DT	0	1.5	0	98.5	0	0
HS	1.0	0	0	0	99.0	0
FR	0	0	1.0	1.5	0	97.5
Mean Recognition Rate = 96.92%						

4. Conclusion

In this paper, we have presented 1D transform features for facial expression recognition along with HMM to model, train and recognize the time sequential facial images. The main aim of these features is to influence the global facial properties and highlight the local expression changes information. To the best of our knowledge, there has been no previous work regarding the time sequential facial expression recognition utilizing LDA on 1D transform features along with HMM. Our experimental results showed the promising performance of the proposed features method achieving recognition rates of 96.75% and 96.92% using Cohn-Kanade and AT&T facial expressions datasets.

Acknowledgements

The research was supported by the Implementation of Technologies for Identification, Behavior, and Location of Human based on Sensor Network Fusion Program through the Ministry of Trade, Industry and Energy (Grant Number: 10041629).

References

[1] M. Wimmer, B. MacDonald, D. Jayamuni and A. Yadav, “Facial expression recognition for human-robot interaction: a prototype,” *in Proceedings of RobVis Conference*, Auckland, pp.139-152, Feb. 2008.

[2] A. Jalal and S. Kim, “Global security using human face understanding under vision ubiquitous architect-ture system,” W*orld academy of science, engineering and technology*, vol. 13, pp.7-11, 2006.

[3] A. Jalal and I. Uddin, “Security architecture for third generation (3G) using GMHS cellular network,” *in Proceedings of ICET*, PK, pp.74-79, Nov. 2007.

[4] S. Kamal and A. Jalal, “A hybrid feature extraction approach for human detection, tracking and activity recognition using depth sensors,” *Arabian Journal for science and engineering*, 2015.

[5] A. Jalal and Y. Rasheed, “Collaboration achievement along with performance maintenance in video stream-ing,” *in Proceedings of ICL*, Austria, pp. 1-8, Sep. 2007.

[6] A. Jalal, S. Kamal and D. Kim, “A depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environ-ments,” *Sensors*, vol. 14, pp. 11735-11759, 2014.

[7] A. Jalal, N. Sharif, J. Kim and T. Kim, “Human activity recognition via recognized body parts of human depth silhouettes for residents monitoring services at smart homes,” *Indoor and built environ-ment*, vol. 22, pp. 271-279, 2013.

[8] A. Jalal and S. Kim, “Algorithmic implementation and efficiency maintenance of real-time environment using low-bitrate wireless communication,” *in Pro-ceedings of SEUS*, Korea, Apr. 2006.

[9] A. Kapoor, W. Burleson and R. Picard, “Automatic prediction of frustration,” *International Journal of human-computer studies*, vol. 65, pp. 724-736, 2007.

[10] A. Jalal, S. Lee, J. Kim and T. Kim, “Human activity recognition via the features of labeled depth body parts,” *in Proceedings of ICOST*, Italy, pp. 246-249, Jun. 2012.

[11] A. Jalal, J. Kim and T. Kim, “Development of a life logging system via depth imaging-based human activity recognition for smart homes,” *in Proceedings of SHB*, Korea, pp. 91-95, Oct. 2012.

[12] A. Farooq, A. Jalal and S. Kamal, “Dense RGB-D Map-Based Human Tracking and Activity Recognition using Skin Joints Features and Self-Organizing Map,” KSII transactions on internet and information systems, vol. 9, pp. 1856-1869, 2015.

[13] A. Jalal, Y. Kim, S. Kamal, A. Farooq and D. Kim, “Human daily activity recognition with joints plus body features representation using Kinect sensor,” *in Proceedings of ICIEV*, Japan, pp. 1-6, Jun. 2015.

[14] A. Jalal, and S. Kamal, “Real-Time Life Logging via a Depth Silhouette-based Human Activity Recognition

System for Smart Home Services," *in Proceedings of AVSS*, Korea, pp. 74-80, Aug. 2014.
[15] A. Jalal, and Y. Kim, "Dense Depth Maps-based Human Pose Tracking and Recognition in Dynamic Scenes Using Ridge Data," *in Proceedings of AVSS*, Korea, pp.119-124, Aug. 2014.
[16] A. Jalal and A. Shahzad, "Multiple facial feature detection using vertex modeling structure," *in Proceedings of ICL*, Austria, pp.1-7, Sep. 2007.
[17] A. Jalal and S. Kim, "Advanced performance achievement using multi-algorithmic approach of video transcoder for low bit rate wireless communication," *ICGST journal*, vol. 5, no. 9, pp. 27-32, 2005.
[18] A. Jalal, S. Kamal and D. Kim, "Depth Silhouettes Context: A new robust feature for human tracking and activity recognition based on embedded HMMs," *in Proceedings of URAI*, Korea, pp. 294-299, Oct. 2015.
[19] A. Jalal, S. Kamal and D. Kim, "Individual Detection-Tracking-Recognition using depth activity images," *in Proceedings of URAI*, Korea, pp. 450-455, Oct. 2015.
[20] A. Jalal, Y. Kim and D. Kim, "Ridge body parts features for human pose estimation and recognition from RGB-D video data," *in Proceedings of ICCCNT Conference*, China, pp. 1-6, Jul. 2014.
[21] A. Jalal, S. Kamal and D. Kim, "Depth Map-based Human Activity Tracking and Recognition Using Body Joints Features and Self-Organized Map," *in Proceedings of ICCCNT*, China, pp. 1-6, Jul. 2014.
[22] A. Jalal, S. Kamal and D. Kim, "Shape and motion features approach for activity tracking and recognition from Kinect video camera," *in Proceedings of WAINA Conference*, Korea, pp. 445-450, Mar. 2015.
[23] A. Jalal and M. Zeb, "Security enhancement for e-learning portal," *Journal of computer science and network security*, vol. 8, no. 3, pp. 41-45, 2008.
[24] A. Jalal and M. Zeb, "Security and QoS optimization for distributed real time environment," *in Proceedings of CIT Conference*, Japan, pp. 369-374, Oct. 2007.
[25] A. Jalal, J. Kim and T. Kim, "Human activity recognition using the labeled depth body parts information of depth silhouettes," *in Proceedings of SHB*, Korea, pp. 1-8, Oct. 2012.
[26] F. S. Samaria, and A. C. Harter, "Parameterization of a stochastic model for human face identification," *in Proceedings of WACV*, USA, pp. 138-142, Dec. 1994.

Ahmad Jalal He received his Ph.D. degree in the Dept. of Biomedical Eng. at Kyung Hee University, Korea. His research interest includes human computer interaction, image processing, and computer vision.

Shaharyar Kamal He received his Ph.D. degree in the Dept. of Radio & Electronics Eng. at Kyung Hee University, Korea. His research interest includes wireless communication, image processing.

Daijin Kim He received the Ph.D. degree in electrical and computer engineering from Syracuse University, Syracuse, NY. He is currently a Professor in the Department of Computer Science and Engineering at POSTECH, Pohang, Korea and Director of BK21+ POSTECH CSE Institute. His research interests include computer vision, human computer interaction, and intelligent systems.