

# Facial Expression Recognition Using 3D Convolutional Neural Network

Young-Hyen Byeon

Department of Control and Instrumentation Engineering,  
Chosun University  
Gwangju, Korea

Keun-Chang Kwak\*

Department of Control and Instrumentation Engineering,  
Chosun University  
Gwangju, Korea

**Abstract**—This paper is concerned with video-based facial expression recognition frequently used in conjunction with HRI (Human-Robot Interaction) that can naturally interact between human and robot. For this purpose, we design a 3D-CNN(3D Convolutional Neural Networks) by augmenting dimensionality reduction methods such as PCA(Principal Component Analysis) and TMPCA(Tensor-based Multilinear Principal Component Analysis) to recognize simultaneously the successive frames with facial expression images obtained through video camera. The 3D-CNN can achieve some degree of shift and deformation invariance using local receptive fields and spatial subsampling through dimensionality reduction of redundant CNN's output. The experimental results on video-based facial expression database reveal that the presented method shows a good performance in comparison to the conventional methods such as PCA and TMPCA.

**Keywords**—convolutional neural network; facial expression recognition; deep learning

## I. INTRODUCTION

HRI (Human-Robot Interaction) is a critical technology as evaluating, developing and designing interactional environments for intelligent system to make cognitive and emotional interaction through some communicational channels between human and robot. It is for synthetically understanding a user's intention and then responding [1]-[5].

Such HRI has basic difference in autonomy for robot and conventional HCI (Human-Computer Interface) to have, interactional bidirection and diversity of interaction or controlling level[6][7].

In a routine living with human for robot to give service should need ability to interact with human using same means. Likewise, for efficient interaction between human and robot should need system of C3 paradigm that consists of modules to develop convenience, cooperativeness and closeness of interaction between human and robot. Also, for efficient interaction between human and robot should need those technologies of various areas are converged. For example, it needs multimodal interactive method that can provide mediator of various communications like vision, auditory, feel and mediating interface, and multimodal technology that converges information inputted through various interactive channels. Besides, for performing proper works to requests from user and situation should need a series of cognitive course like recognition of situation, inference, decision making and plan. Also, it needs technologies to make proper emotional move

fitted to situation and for robot to be characterized under recognizing various emotional response like voice and expression [6]. We especially focus on video-based facial expression recognition technique.

People express their mind through gestures and facial expressions. The facial expression is the most useful, natural mean to notify their mind. Study about facial expression have been researched long ago by cognitive scientist and recently many researcher try to develop methods to recognize facial expression automatically and accurately [8][9].

There are methods to recognize facial expression such as comparing positions of eye, nose and mouth[10], optical flow extracting muscle's movement[11][12], PCA(Principal Component Analysis)[13] and LDA(Linear Discriminant Analysis)[9][14][15]. Ekman, a psychologist who researches facial expression, said human's facial expression disappears within seconds [16]. That is, facial expression recognition needs to work under keeping facial expression both long and short [17]. That means we need to study facial expression recognition using a stationary image and video having time base [18][19]. However, Methods referred above are a little difficult to use for video-based facial expression recognition, including successive frames with facial expression images.

On the other hand, Convolutional Neural Networks (CNN) has been successfully applied to face recognition from two-dimensional images [20]. The networks incorporate constraints and achieve some degree of shift and deformation invariance. This method has demonstrated to be successful in various fields such as character recognition [21], document recognition [22], object recognition [23], handwritten digit recognition [24], EEG signal classification [25], and facial expression recognition [26]. However, the conventional CNNs are currently limited to handling video-based images. Furthermore, because this network is trained with the usual backpropagation gradient descent procedure, it is not appropriate for video-based face images with redundant CNN's output.

Therefore, we design a 3D-CNN (3D Convolutional Neural Networks) by expanding 2D structure of CNN to 3D structure of CNN for video-based facial expression recognition. There has been no study that tried to apply 3D-CNN for video-based facial recognition. The experiment uses video-based facial expression database of CNU captured from video camera. The experimental results reveal that the 3d-CNN shows a good performance in comparison with the previous approaches with vector representation such as PCA and TMPCA [17].

Conventional PCA and TMPA are described in section 2. A 3D-CNN is explained in section 3 and experimental result is in section 4. We make conclusions in section 5.

## II. RELATED WORK

### A. Principal Component Analysis

PCA is a 2D statistical method that uses statistical properties to variance and is used to efficiently reduce dimensionality of input data that have high dimensionality. To summarize PCA, it is a method of dimensionality reduction by linearly projecting whole data of image to some eigenvectors that are the biggest variances of whole data of image [27][28].

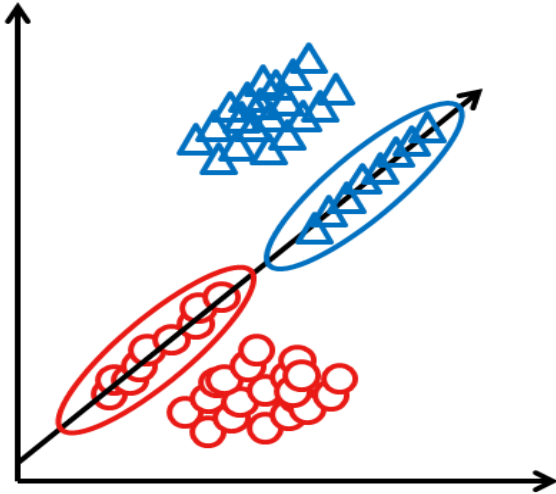


Fig. 1. Projection of PCA

Though the data's dimensionality is reduced to one dimensionality using PCA the data still can be separable to each class as shown in the Fig. 1. That is, PCA have merits that it seems to have effect of keeping information about distribution of input data, reducing computing power, reducing noise of data, and compressing data. The course of training PCA is described below, see table 1.

TABLE I. TRAINING OF PCA

1. Define of p units of vector of training image	$X = [x^1 x^2 \dots x^p]$	(1)
2. Subtraction of each image vector from averaging image	$\bar{x}^i = x^i - mean, mean = \frac{1}{p}\sum_{i=1}^p x^i$	(2)
3. $N \times N$ covariance matrix using p units of $\bar{x}^i$ vector	$\Omega = \bar{X}\bar{X}^T, \bar{X} = [\bar{x}^1 \bar{x}^2 \dots \bar{x}^i]$	(3)
4. Definition of eigenvalue and eigenvector about covariance matrix	$\Omega v_i = \lambda v_i$	(4)
5. Definition of feature vector about training image		

Here, an eigenvalue from covariance matrix means direction to make variance the biggest and an eigenvector matched to it means variability to certain direction. This eigenvector is the eigenface. Fig. 2 shows face images which

are linearly combined from an eigenface with a feature vector. The course of verifying PCA is described below, see table 2.

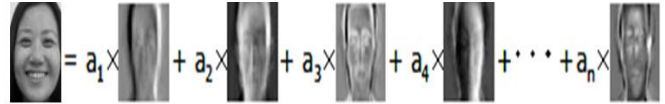


Fig. 2. Face images by linearly combining an eigenface with a feature vector

TABLE II. VERIFICATION OF PCA

1. Definition of p units of vector of verifying image	$Y = [y^1 y^2 \dots y^p]$	(5)
2. Subtraction of each image vector from averaging image	$\bar{y}^i = y^i - mean, mean = \frac{1}{p}\sum_{i=1}^p y^i$	(6)
3. Definition of feature vector using eigenvector about verifying image	$\tilde{y}^i = V^T \bar{y}^i$	(7)

Finally a face image is classified to minimum similarity by measuring a similarity between feature vector of image that is calculated above and feature vectors that exist for training. The method of measuring a similarity is shown in detail at section 4 [29].

### B. Multilinear Principal Component Analysis

MPCA (Multilinear Principal Component Analysis) is a method that gets covariance directly without transforming 1D vector. Tensors as A and components of tensors as N are normally represented. N is degree of tensor's target and each exponent defines a mode. Tensors under  $N > 2$  can be shown as a high degree of vector and normalized matrix. Component of tensor has exponent marked at bracket. Tensor having N degree is  $A \in R^{I_1 \times I_2 \times \dots \times I_N}$ . It is expressed by N exponents,  $i_n, n = 1, 2, \dots, N$  and each  $i_n$  allocates A's n-mode. Tensor A's n-mode multiplication by a matrix  $U \in R^{J_n \times I_n}$  at  $A \times_n U$  is shown as following.

$$(A \times_n U)(i_1, \dots, i_{n-1}, j_n, j_{n+1}, i_{n+1}, \dots, i_N) = \sum_{i_n} A(i_1, \dots, i_N) \cdot U(j_n, i_n) \quad (8)$$

Scalar multiplication of two tensors,  $A, B \in R^{I_1 \times I_2 \times \dots \times I_N}$ , is defined as following.

$$\langle A, B \rangle = \sum_{i_1} \sum_{i_2} \dots \sum_{i_N} A(i_1, i_2, \dots, i_N) \cdot B(i_1, i_2, \dots, i_N) \quad (9)$$

Tensor A's frobenius norm is defined as  $\|A\|_F = \sqrt{\langle A, A \rangle}$  and under standard multilinear algebraic expression, tensor A is expressed as following form of multiplication.

$$A = S \times_1 U^{(1)} \times_2 U^{(2)} \times \dots \times_N U^{(N)} \quad (10)$$

Where,  $U^{(N)} = (u_1^n u_2^n \dots u_n^n)$  is an orthogonal matrix of  $I_n \times I_n$ . In  $R^{I_1} \otimes R^{I_2} \dots \otimes R^{I_N}$  space, when M tensor's set is  $\{Y_m, m = 1, \dots, M\}$ , total variance is the following equations.

$$\Psi_Y = \sum_{m=1}^M \|Y_m - \bar{Y}\|_F^2 \quad (11)$$

$$\bar{Y} = \frac{1}{M} \sum_{m=1}^M Y_m \quad (12)$$

Here,  $\bar{Y}$  is an averaged tensor and total variance matrix of n-mode is the following equation.

$$C_{TY} = \sum_{m=1}^M (A_m^{(n)} - \bar{A}^{(n)})(A_m^{(n)} - \bar{A}^{(n)})^T \quad (13)$$

Where,  $A_m^{(n)}$  is  $A_m$ 's n-mode unfolded matrix. MPCA maximizes a scatter criterion based on tensor and this problem can be solved by separating it into N linear optimizations. When there are all different projection matrix,  $\tilde{U}^{(1)}, \dots, \tilde{U}^{(n-1)}, \tilde{U}^{(n+1)}, \dots, \tilde{U}^{(N)}$ , a  $\Psi_Y$  consists of eigenvectors corresponding the largest  $P_n$  eigenvalue in matrix of  $\{\tilde{U}^{(n)} \in R^{I_n \times P_n}\}$  is maximized as the following equation. Eigenvector is calculated using following equation.

$$\{\tilde{U}^{(n)}, n = 1, 2, \dots, N\} = \arg \max_{\Psi_Y} \quad (14)$$

$$\bar{U}^{(1)}, \bar{U}^{(2)}, \dots, \bar{U}^{(N)}$$

$$\Phi^{(n)} = \sum_{m=1}^M (X_{m(n)} - \bar{X}_{(n)}) \cdot \tilde{U}_{\Phi(n)} \cdot \tilde{U}_{\Phi(n)}^T \cdot (X_{m(n)} - \bar{X}_{(n)})^T \quad (15)$$

Where,  $\tilde{U}_{\Phi(n)} = (\tilde{U}^{(n+1)} \otimes \tilde{U}^{(n+1)} \otimes \dots \otimes \tilde{U}^{(n)} \otimes \tilde{U}^{(1)} \otimes \tilde{U}^{(2)} \otimes \dots \otimes \tilde{U}^{(n-1)})$ .

Gotten projection matrix  $\tilde{U}^{(n)}$  is seen as  $\prod_{n=1}^N P_n$ . But, because all of these are not useful for recognition some of these can be selected through discernment of class. Ordered feature vector by discernment of class is calculated as following [17].

$$Y_m = X_m \times_1 \tilde{U}^{(1)T} \times_2 \tilde{U}^{(2)T} \dots \times_N \tilde{U}^{(N)T} \quad (16)$$

### C. Tensor-based Multilinear Principal Component Analysis

Common facial expression recognition is to recognize facial expression every moment facial expression image comes. It is difficult to hold a same facial expression for general person. Such method is not appropriate in environment of recognizing facial expression on real-time. TMPCA (Tensor-based Multilinear Principal Component Analysis) is a method that considers several frames as a facial expression. For example, assume that 10 frames of which each image's size is  $64 \times 48$  comes from camera per second. In a case of TMPCA,  $1 \times 3072$  covariance matrix is generated for each frame. The total projection vector's size is  $10 \times 3072$ . 3D tensor is constructed by adding time axis on grayscale image commonly used in facial expression recognition and facial recognition is performed using MPCA. TMPAC is used to extract image's essential features by expressing 3D tensor axis of frame is added to direct tensor. It improves processing time and performance of recognition [17].

### III. 3D CONVOLUTIONAL NEURAL NETWORK

A 3D convolution is performed with 3D kernel and 3D data which 2D images are merged. A 3D convolution is expressed by the following equation (17) and Fig. 3.

$$o_{(1,1,1)} = a_{(1,1,1)}b_{(1,1,1)} + a_{(1,2,1)}b_{(1,2,1)} + a_{(1,3,1)}b_{(1,3,1)} \quad (17)$$

$$+ a_{(2,1,1)}b_{(2,1,1)} + a_{(2,2,1)}b_{(2,2,1)}$$

$$+ a_{(2,3,1)}b_{(2,3,1)} + a_{(3,1,1)}b_{(3,1,1)}$$

$$+ a_{(3,2,1)}b_{(3,2,1)} + a_{(3,3,1)}b_{(3,3,1)}$$

$$+ a_{(1,1,2)}b_{(1,1,2)} + a_{(1,2,2)}b_{(1,2,2)}$$

$$+ a_{(1,3,2)}b_{(1,3,2)} + a_{(2,1,2)}b_{(2,1,2)}$$

$$+ a_{(2,2,2)}b_{(2,2,2)} + a_{(2,3,2)}b_{(2,3,2)}$$

$$+ a_{(3,1,2)}b_{(3,1,2)} + a_{(3,2,2)}b_{(3,2,2)}$$

$$+ a_{(3,3,2)}b_{(3,3,2)} + a_{(1,1,3)}b_{(1,1,3)}$$

$$+ a_{(1,2,3)}b_{(1,2,3)} + a_{(1,3,3)}b_{(1,3,3)}$$

$$+ a_{(2,1,3)}b_{(2,1,3)} + a_{(2,2,3)}b_{(2,2,3)}$$

$$+ a_{(2,3,3)}b_{(2,3,3)} + a_{(3,1,3)}b_{(3,1,3)}$$

$$+ a_{(3,2,3)}b_{(3,2,3)} + a_{(3,3,3)}b_{(3,3,3)}$$

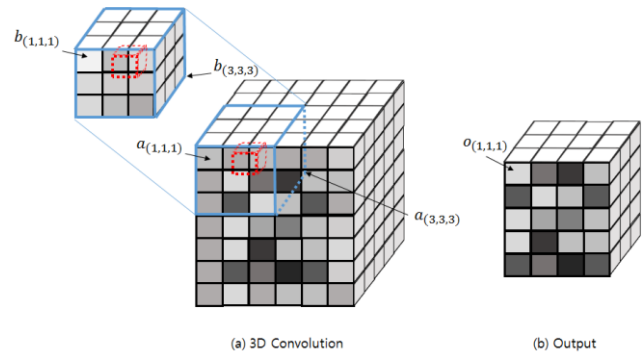


Fig. 3. Example of a 3D convolution

A subsampling value of a pixel is calculated by multiplying each pixel of kernel and image and accumulating them in an overlapped area. Here values of kernel are for calculating a mean. This process is conducted on whole image and the resulting image becomes small by abandoning some intersecting pixels. A subsampling achieves some degree of shift and deformation invariance. A subsampling is expressed by the following Fig. 4.

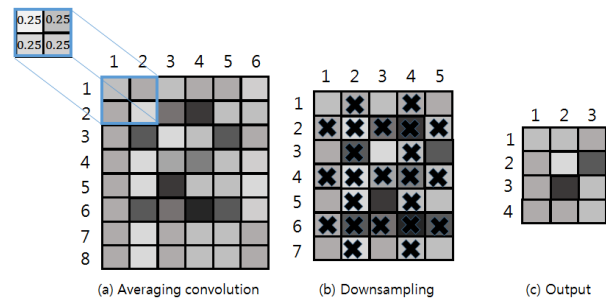


Fig. 4. Example of a subsampling

Fig. 5 shows a structure of 3D-CNN for facial expression recognition based on video. Here, the structure consists of 5 layers. First layer is for input, second layer is for convolution, third layer is for subsampling, fourth layer is for convolution and fifth layer is for subsampling. Initial values of kernels are random in specific range. In a first layer, a 3D data which consists of five frames on video enters as input of system. In a second layer having 3 maps, a convolution extracts features from the input.

In a third layer, a subsampling reduces the size of input image. In a fourth layer having 29 maps, a convolution extracts features from output of previous layer. In a fifth layer, a subsampling reduces the size of image from previous layer. Finally, a feature vector is created by making images arranged to single row on all maps.

Data used for input is facial expression images based on video which are overlapped successively along five frames. The size of data starts with 64x48x5 as input. At the second layer, the size of data turns 60x44x3 because a convolution is performed with a kernel of which size is 5x5x3. At the third layer, the size of data turns 30x22x3 because a subsampling is performed with a kernel of which size is 2x2x1 without time base. At the fourth layer, the size of data turns 26x18x1 because a convolution is performed with a kernel of which size is 5x5x3. At the fifth layer, the size of data turns 13x9x1 because a subsampling is performed with a kernel of which size is 2x2x1 without time base. This 13x9x1 can be a vector by making it have single row. That is, a map can have 117 feature values. So the last size of feature vector is 3393 because there are 29 maps [20][30].

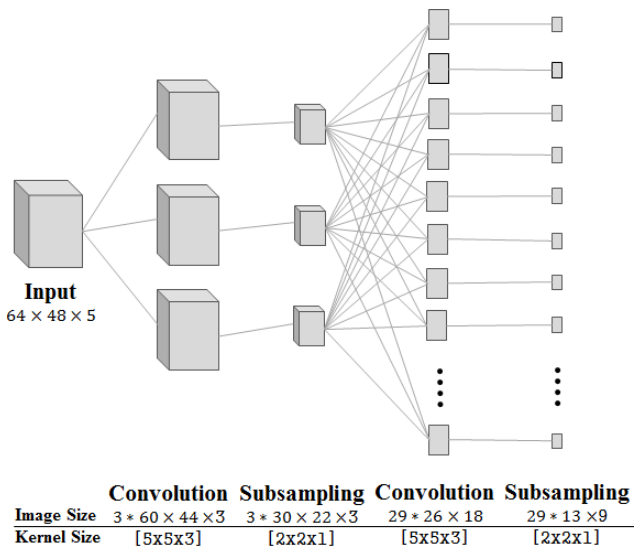


Fig. 5. A struct of 3D-CNN used in this study

A method for classification used in this study is Euclidean distance between vectors previously generated for learning and a vector generated for recognition. The close distance is accepted for a result of recognition. Euclidean distance is expressed by the following equation (18)[29].

$$d(X, Y) = \sum_{i=1}^n |x_i - y_i| \tag{18}$$

#### IV. EXPERIMENTAL RESULTS

To evaluate the performance of 3D-CNN, a database of CNU for facial expression is used. This database was captured from video camera. There are training and checking data which consist of 15 frames every facial expression. Six facial expressions like happiness, sadness, anger, surprise, disgust,

and fear were captured from 10 people. To analyze performance, we used a computer that had CPU 3.10GHz, Intel(R) Core(TM) i5 4440, memory 8GB and a software we use is Matlab R2013A. This experiment is performed with comparing PCA and TMPKA, conventional methods. Fig. 6



shows the images of facial expressions from CNU.

Fig. 6. Facial expression images captured by video camera

In the database of CNU for facial expressions, there are facial expressions of ten people. Every image of facial expressions is resized to 64x44. Such single images are grouped every five frames because a facial expression consists of 5 frames in the data of CNU for facial expressions. So the size is 64x44x5 also as an input. Fig. 7 shows an example of video data that five frames are linked successively. The structure of 3D-CNN we use is summarized at table 3[17][31].

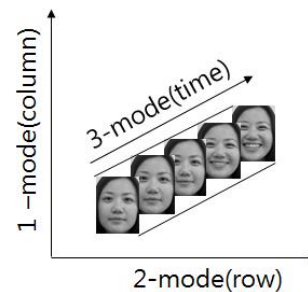


Fig. 7. Example of video data that five frames are linked successively

Feature vectors are obtained about all training data and all checking data. A label has the least value among distances of feature vectors between training and checking is considered as its class. Finally, the recognition rate is obtained by dividing the successful times by total times. For experiment more, we make changes on the number of maps at second and fourth layer. The number of map at the second layer is varied from 1 to 30 and the number of map at fourth layer is varied from 1 to 30. The result of experiment is summarized at table 4. The best recognition rate and its number of map are compared. Fig. 8 shows the images that are generated during 3D-CNN about first data. Those images were performed by histogram equalization.

TABLE III. STRUCTURE OF 3D-CNN

Layer	type	kernel size
1	Input	
2	Convolution	5x5x5
3	Subsampling	2x2
4	Convolution	5x5x6
5	Subsampling	2x2

TABLE IV. RESULT OF EXPERIMENT

	Number of map		Recognition rate
	Second layer	Fourth layer	
Checking	3	29	95%

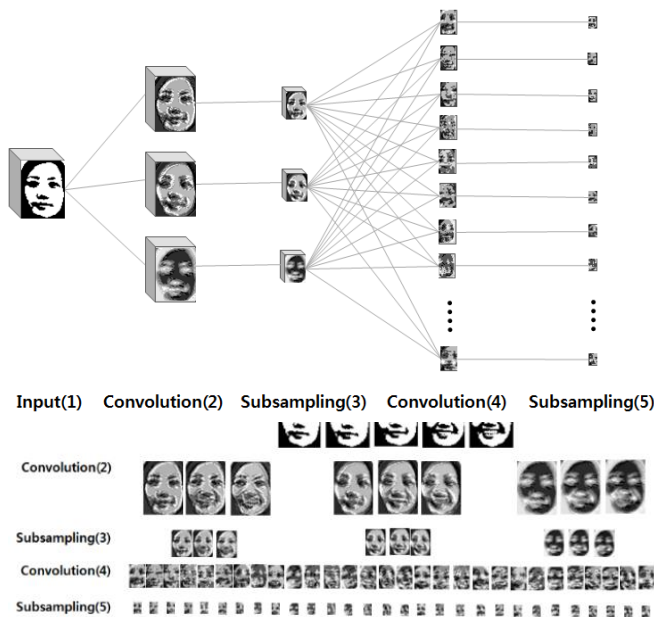


Fig. 8. Images generated during 3D-CNN

To compare 3D-CNN proposed in this paper with PCA and TMPCA in facial expression recognition, each recognition rate is shown in table 5. The performance of 3D-CNN is higher than performance of PCA as 6.7% up and performance of TMPCA as 4.44% up.

TABLE V. COMPARISON OF 3D-CNN WITH PCA AND TMPCA

	Recognition Rate
PCA[13]	88.3%
TMPCA[17]	90.56%
3D-CNN	95%

## V. EXPERIMENTAL RESULTS

We have designed 3D Convolutional Neural Networks for video-based facial expression images to perform human-robot interaction technique. Thus, the 3D-CNN has characteristics that can recognize simultaneously the successive frames with facial expressional images obtained through video camera. The experimental results reveal that the 3D-CNN shows different performance in varying the number of map and a good performance in comparison with the conventional methods such as PCA and TMPCA. Consequently, we hope that the presented method can have better performance. For more experiment, we will try to combine the 3D-CNN with other method that has possibility to enhance its performance.

### REFERENCES

- [1] A. S. Sekmen, M. Wilkes, K. Kawamura, "An application of passive human-robot interaction: human tracking based on attention distraction", IEEE Trans. on Systems, Man, and Cybernetics-Part A, vol. 32, no. 2, pp. 248-259, 2002.
- [2] X. Yin, M. Xie, "Finger identification and hand posture recognition for human-robot interaction", Image and Vision Computing, vol. 25, pp.1291-1300, 2007.
- [3] Y. Sugimoto, Y. Yoshitomi, S. Tomita, "A method for detecting transitions of emotional states using a thermal facial image based on a synthesis of facial expressions", Robotics and Autonomous Systems, vol. 31, pp. 147-160. 2000.
- [4] J. J. Lien, T. Kanade, J. F. Cohn, C. C. Li, "Detection, tracking, and classification of action units in facial expression", Robotics and Autonomous Systems, vol. 31, pp. 131-146. 2000.
- [5] G. Medioni, A. R. J. Francois, M. Siddiqui, K. Kim, H. Yoon, "Robust real-time vision for a personal service robot", Computer Vision and Image Understanding, vol. 108, pp. 196-203, 2007.
- [6] Korea Intellectual Property Office, Trend of Patent Application for Human-Robot Interaction, 2005.
- [7] Institute for Information Technology Advancement, Human-Robot Interaction Technology (Intelligent Service Robot), 2006
- [8] P. Ekman and W. V. Friesen, "Emotion in the human face system," Cambridge University Press, San Francisco, CA, scnd edition, 1982.
- [9] J. H. Rho, Y. H. Baek, S. R. Moon, Y. J. Kang, "A Study on Face Expression Recognition using LDA Mixture Model," Proceeding of the Spring Conference of the Korea Multimedia Society, pp. 50-54, 2006.
- [10] Z. Zang, M. Lyons, M. Schuster and S. Akamatsu, "Comparison between Geometry-Based and Gabor Wavelets-Based Facial Expression Recognition Using Multi-Layer Perceptron," Proceedings of Third IEEE International Conference on AutomaticFace and Gesture Recognition, pp. 454-459, 1998.
- [11] J. J. Lien, T. Kanade, J. Cohn, and C. Li, "Detection, Tracking, and Classification of Action Units in Facial Expression," Journal of Robotics and Autonomous Systems, July 1999.
- [12] K. Mase, "Recognition of facial expression from optical flow," IEICE Transactions on Information Systems, vol. J80-D, no. 6, June 1997.
- [13] M. Turk, A. Pentland, "Eigenfaces for recognition," Journal of Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.
- [14] P. Belhumeur, J. Hespanha, D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, 1997.
- [15] S. Balakrishnama, A.Ganapathiraju, "Linear Discriminant Analysis – A Brief Tutorial," Institute for signal and Information Processig, 1998.
- [16] P. Ekman and W. V. Friesen, "Unmasking the face," Malor Books Press, 2003.
- [17] K. C. Kwak, M. W. Lee, S. B. Pan, "Facial Expression Recognition by Tensor Representation from Video," The Journal of Korean Institute of Information Technology, vol. 10, no. 4, pp. 185-190, April 2012.

- [18] Y. H. Joo, K. H. Jeong, M. H. Kim, J. B. Park, J. Lee, Y. J. Cho, "Facial Image Analysis Algorithm for Emotion Recognition," Journal of Korean Institute of Intelligent Systems, vol. 14, no. 7, pp. 801-806, December 2004.
- [19] J. T. Joo, G. J. Park, G. E. Gho, H. C. Yang, G. B. Sim, "Emotion Recognition and Expression using Facial Expression," Proceedings of KFIS Spring Conference, vol. 17, no. 1, pp. 295-298, 2007.
- [20] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face Recognition: A Convolutional Neural-Network Approach", IEEE Trans. on Neural Networks, Vol. 8, No. 1, pp. 98-113, 1997.
- [21] G. Lv, "Recognition of multi-fontstyle characters based on Convolutional neural network", International Symposium on Computation Intelligence and Design, Vol. 2, pp. 223-225, 2011.
- [22] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition", Proceeding of IEEE , Vol. 86, No. 11, pp.2278-2324, 1998.
- [23] Y. Lecun, F. J. Huang, L. Bottou, "Learning methods for generic object recognition with invariance to pose and lighting", Proceeding of the Computer Vision and Pattern Recognition, vol. 2, pp. 97-104, 2004.
- [24] X. X. Niu, and C. Y. Suen, "A novel hybrid CNN-SVM classifier for recognizing handwritten digits", Pattern Recognition, Vol. 45, No. 4, pp. 1318-1325, 2012.
- [25] H. Cecotti, and A. Greaser, "Convolutional Neural Network with embedded Fourier Transform for EEG classification", International Conference on Pattern Recognition, pp. 1-4, 2008.
- [26] M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda, "Subject independent facial expression recognition with robust face detection using a convolutional neural network", Neural Networks, Vol. 16, No. 5-6, pp. 555-559, 2003.
- [27] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 711-720, 1997.
- [28] Y. H. Shin, J. S. Ju, E. Y. Kim, T. Kurata, A. K. Jain, S. Park, K. Jung, "Automatic Facial Expression Recognition using Tree Structures for Human Computer Interaction," Korea Society of Industrial Information Systems, vol. 12, no. 3, pp. 60-68, 2007.
- [29] M. Y. Lee, "Facial Expression Recognition by Tensor Representation from Video", Chosun University, 2012.
- [30] Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu, "3D Convolutional Neural Networks for Human Action Recognition", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 35, No. 1, pp. 221-231, 2013.
- [31] S. J. Han, K. C. Kwak, H. J. Go, S. S. Kim, M. G. Chun, "Facial Expression Recognition using ICA-Factorial Representation Method," Journal of Korean Institute of Intelligent Systems, vol 13. no. 3, pp. 371-376, june 2003.

#### AUTHORS PROFILE

Young-Hyen Byeon received the B.Sc and M.Sc. from Chosun University, Gwangju, Korea, in 2012 and 2014, respectively. He is currently pursuing a Ph.D. candidate. His research interests include human-robot interaction, computational intelligence, and pattern recognition.

Keun-Chang Kwak received the B.Sc., M.Sc., and Ph.D. degrees from Chungbuk National University, Cheongju, Korea, in 1996, 1998, and 2002, respectively. During 2003-2005, he was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. From 2005 to 2007, he was a Senior Researcher with the Human-Robot Interaction Team, Intelligent Robot Division, Electronics and Telecommunications Research Institute, Daejeon, Korea. He is currently the Associative Professor with the Department of Control and Instrumentation, Engineering, Chosun University, Gwangju, Korea. His research interests include human-robot interaction, computational intelligence, biometrics, and pattern recognition. Dr. Kwak is a member of IEEE, IEICE, KFIS, KRS, ICROS, KIPS, and IEEK.