Soft Comput (2017) 21:4995–5002
DOI 10.1007/s00500-016-2095-0

METHODOLOGIES AND APPLICATION

A face recognition system based on convolution neural network using multiple distance face

Hae-Min Moon[1] · **Chang Ho Seo**[1] · **Sung Bum Pan**[2]

Published online: 2 March 2016

Abstract The recognition technology that recognizes or discriminates certain individuals is very important for the security that provides intelligence services. Face recognition rate can vary depending on variability of the face itself as well as other external factors such as illumination, background, angle and distance of a camera position. The paper suggests a proper method for long-distance face recognition by resolving the change in recognition rate resulting from distance change in long-distance face recognition. For the long-distance face recognition test, face images by actual distance from 1 to 9 m away were obtained directly. Actual face images taken by distance were applied to resolve the issue rising from distance change and CNN was applied to extract overall features of face. The test showed that proposed face recognition algorithm that used CNN as feature extraction and face images by actual distance for training was found to show the best performance.

Keywords Long-distance face recognition · Multiple distance face · Intelligent robot service · Convolution neural network

Communicated by V. Loia.

✉ Sung Bum Pan
sbpan@chosun.ac.kr

Hae-Min Moon
bombilove@gmail.com

Chang Ho Seo
chseo@kongju.ac.kr

[1] Department of Convergence Sciences, Kongju National University, Kongju, Republic of Korea

[2] Department of Electronics Engineering, Chosun University, Gwangju, Republic of Korea

1 Introduction

The past robot market was focused on industrial robots as a means of productivity improvement and labor alternatives. But recently the demand for service robots to improve the quality of life is increasing. The most important issue in service robots is that they judge and act for themselves through interaction between human beings and robots. Therefore, the interest is rising in the intelligent robots that can interact with humans, recognize the changes of external environment and automatically move upon their own judgment. The user recognition technology that allows recognizing or discriminating a specific individual at a robot environment is absolutely necessary for security and surveillance service Chellappa et al (1995). The robot service through user recognition is now able to provide proper personal services quickly and precisely to specific users Kim et al (2015). As face recognition, in particular, can be executed without contact or cooperation and from a far distance, researches on long-distance recognition using faces are undergoing Moon et al (2015).

The face recognition technology in the robot environment detects the face that exists among the images input from the robot's camera and verifies the identity. The existing face recognition methods mainly consist of Principal Component Analysis (PCA) Turk and Pentland (1991) or Linear Discriminant Analysis (LDA) Belhumeur et al (1997) where statistic value of all faces is specifically recognized. Since the high-order relationships among the facial image pixels may contain much significant discriminant information, independent component analysis (ICA) is adopted to extract discriminant feature for face recognition Bartlett et al (2002). Recently, it is proved that this feature extractor dependency problem can be addressed by deep learning. Deep learning has a lot of interest in recent years as it can be considered as

the next generation of neural networks. Allowing computers to learn the world like human brain, which is at least learning step by step, from simple concepts to more complex concepts, is the motivation of studying neural networks and also the big challenge in this field (Bevilacqua et al 2008; Chen et al 2012). Convolution neural network (CNN) is comprised of one or more convolution layers and then followed by one or more fully connected layers as in a standard multilayer neural network. This is achieved with local connections and tied weights followed by some form of pooling which results in translation invariant features, so it has been successfully applied to handwritten numeral recognition (Niu and Suen 2012) facial expression recognition (Matsugu et al 2003), character recognition (Lv 2011) and document recognition (Yann et al 1998). CNN studies using two-dimensional and three-dimensional image information have been also conducted in face recognition field (Byeon and Kwak 2014; Lawrence et al 1997).

Unlike the conventional face recognition method that is performed in a fixed distance under cooperative position, it is difficult to ask for cooperative position in the robot environment as the distance between the robot and human is variable. When the existing face recognition technology is applied to the robot environment, the performance in a close distance may be great. However, it can be worse as the distance is farther away. Therefore, the image quality that deteriorates with distance should be considered to apply the face recognition technology to the robot environment. Recently, there are studies on the long-distance face recognition technology that uses a high-performance zoom camera that produces high-quality images even from a far distance (Chen et al 2013; Park et al 2013).

This paper proposes the long-distance face recognition method that can be applied to the robot and surveillance systems that require the user recognition. Under the actual robot environment, the recognition rate declines when the face image was obtained from afar since the distance between the camera and user is not fixed. Therefore, the paper tried to overcome the issue arising from distance change through the method that uses face images by distance for training and the method that standardizes the size of face images. Face images by distance can be obtained from actual distance by the user moving in person. The proposed method in comparison with the conventional method, which only acquires short distance face information, has merits in acquiring precise characteristics of individual faces in both short and long distance. Besides, as the surrounding lights change when the user face is obtained in an actual environment, the paper used histogram as a preprocessing. For the long-distance face recognition test, face images by actual distance from 1 to 9 m away were obtained directly. CNN was applied to face recognition and Euclidean distance was used as similarity measurement. The proposed method was confirmed through the tests indicate excellent performance in various distances, compared to traditional face recognition. It can be operable under environment with conventional low resolution image and strong real-time processing according to distance change. In addition, the conventional method needs the zoom camera to recognize face recognition but the proposed method does not require any additional equipment. The paper consists of Sect. 2 that introduces the proposed system-related research, Sect. 3 explains how to analyze face recognition rate by distance and how to obtain face images by distance. Our experimental results are presented in Sect. 4 and Sect. 5 concludes this paper.

2 Pre-processing and feature extraction for face recognition

The face recognition rate changes so much depending on external factors such as the variability of the face, illumination, background, angle and distance of a camera. The existing studies often removed the surrounding lights or used the face's skin color data to heighten recognition performance. Besides, improved algorithms have been suggested through improvement of face recognition algorithm. The paper uses histogram equalization as a preprocessing to reduce the illumination effect among others that influence the face recognition performance. It also used the method to configure the face images by distance for training to reduce the effect from distance change. Proposed system uses CNN to extract overall features of face for face recognition.

2.1 Histogram equalization

In face recognition which catches the distribution of gray values, the important information in the image, the typical face brightness normalized methods, includes histogram equalization, Modified Census Transform (MCT) conversion and Local Binary Pattern (LBP) conversion. Histogram is an available method Abdullah-Al-Wadud et al (2007). The configuration of histogram uses the gray values from 0 to 255 as an index and this method accumulates the frequency to the corresponding index table based on gray value of each pixel. Histogram equalization is to change the histogram of the image to appear equally in entire sections of the gray scale. Through this adjustment, the intensities can be better distributed on the histogram. This allows for areas of lower local contrast to gain a higher contrast. Histogram equalization accomplishes this by effectively spreading out the most frequent intensity values. When defining the function of converting the pixel value of the image as shown in the following (1), r is the input gray scale and s is the output gray scale value by the conversion function. In general, the conversion function T is assumed as a monotone increasing function,

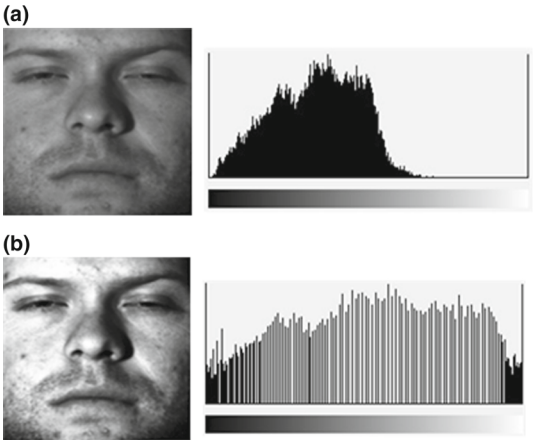

Fig. 1 Equalized image and corresponding histogram **a** original image, **b** histogram equalized image

and the histogram equalization can be expressed in the form of a monotone increasing conversion function.

$$s = T(r) \tag{1}$$

Based on the above, the histograms of input image and output image can be expressed as $P_r(r)$ and $P_s(s)$ in the form of probability density function Parzen (1962), respectively. The conversion function of histogram equalization is defined as a form corresponding to the value obtained by accumulation $P_r(r)$ and can be expressed as the following (2), where τ is a temporary variable for integration and the function adding the probability density function using the integration is called a cumulative distribution function. Figure 1 shows the histogram and histogram cumulative after going through histogram equalization. Figure 1a shows dark image in general because when looking at histogram distribution for original image, the dark value is partially included but the bright value is not included. Figure 1b shows the image resulting from performing histogram equalization to the original image, which can confirm that characteristics of histogram distribution is equally distributed from bright values to dark values compared to the original image.

$$s = T(r) = \int_0^r P(\tau)\mathrm{d}(\tau) \tag{2}$$

2.2 Convolution neural network

CNN has been used for handwriting recognition limitedly in the past, while the recent CNN is applied in various fields as its excellent performance in object recognition has been verified. CNN is a way to use the low-dimensional information by extending to high-dimension, making it easier to classify information. At this time, convolution is used in the process of extracting a feature from the expansion of the image. In two-dimensional image, CNN obtains the output image by adding the image and overlapped kernel parts after multiplying them when the center of kernel with the size of odd times lies in the pixel of the image.

Figure 2 shows typical CNN structure. Currently, CNN is designed with a total of five layers: The first layer is an input step of the image, the 2nd layer is 1st convolution step, the 3rd layer is 1st sub-sampling step the 4th layer is 2nd convolution step and 5th layer is 2nd sub-sampling step. The initial kernel value is set as a random value of particular area. Once an image is entered in the 1st layer, it extracts features into 6 maps by convoluting from this image in the 2nd layer. The sizes of 6 maps are reduced through sub-sampling in the 3rd layer. Through repeated process, the image is configured as a single vector at the last layer, and when all 12 maps are patched together, the feature vector for input is obtained.

For size changes of the image, the input image in the 1st layer is 28×28 size, and when the convolution of kernel size of 5×5 is performed in the 2nd layer of convolution, the convolution uses completely overlapped kernel and image. Therefore, the output image is 24×24 size since the result image is 'the original image size-(kernel size-1)'. Pooling which is a form of non-linear down-sampling is proceeded in the 3rd layer. Pooling partitions the input image into a set of non-overlapping rectangles and, for each such sub-region, outputs the maximum. The function of the pooling layer is to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network. The pooling operation provides a form of translation invariance. At this time, since the kernel size is 2×2, the image size is reduced by half as 12×12. It is possible to design CNN structure variously. For example, the number of convolution layer or sub-sampling layer can be increased or reduced and the number of feature map in the convolution layer can also be changed.

3 Proposed face recognition using multiple distance face image

Figure 3 is the face recognition process where face images extracted by distance from 1 to 9 m away are used. The input face image is standardized through interpolation and histogram equalization is applied as a preprocessing that is robust to illumination change. Face recognition uses CNN to extract overall features of face. The extracted feature data are saved at DB and used for face recognition by Euclidean distance similarity measurement. Face images extracted from various distances can have different face image sample dimensions depending on the distance. To resolve the issue, the paper uses the face size normalization as shown. The process of face image standardization per distance is shown below. When face images per by distance of various sizes are input, the input images are scaled to fit the reference face

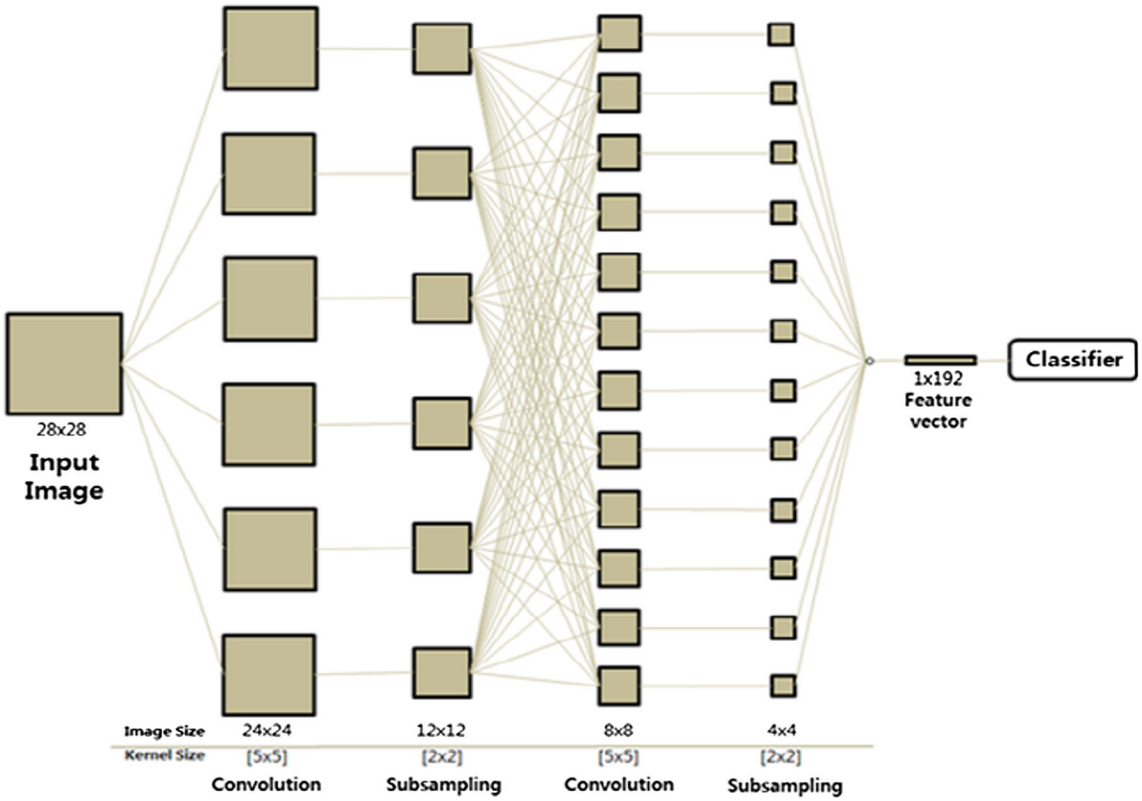

Fig. 2 The architecture of convolution neural network

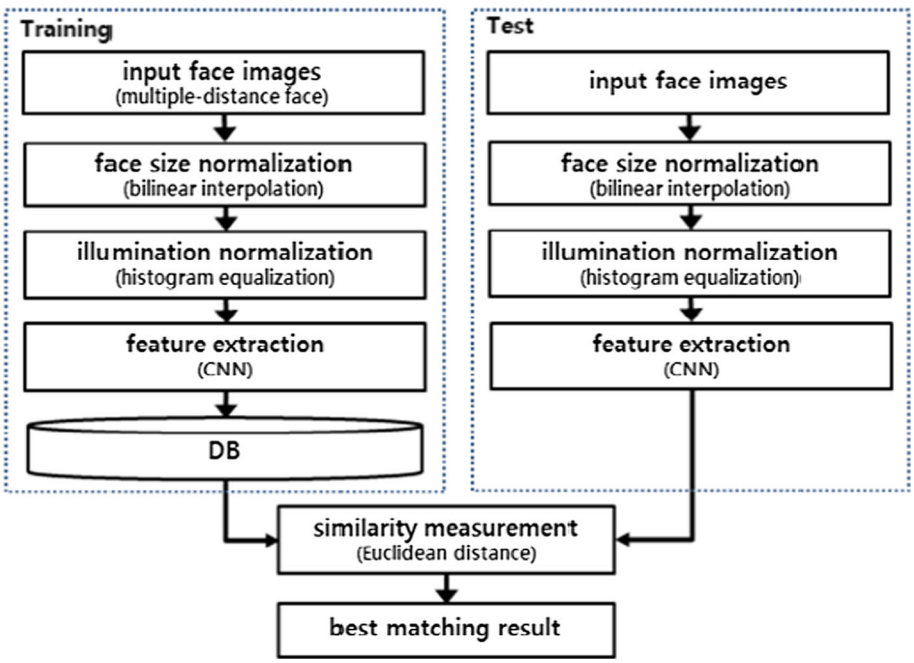

Fig. 3 The overall flow of proposed CNN-based face recognition

size of face images used for training. If the reference size used for training has 1 m as its standard, the size of reference face image is 60 × 60. When the size of input face image is 60 × 60, equalization comes next and when it is smaller or bigger than 60 × 60, it is scaled to 60 × 60 through bilinear interpolation (Moon et al 2015; Gonzalez 2009). All face images entered through the process are standardized to 60 × 60 that is the current reference face size.

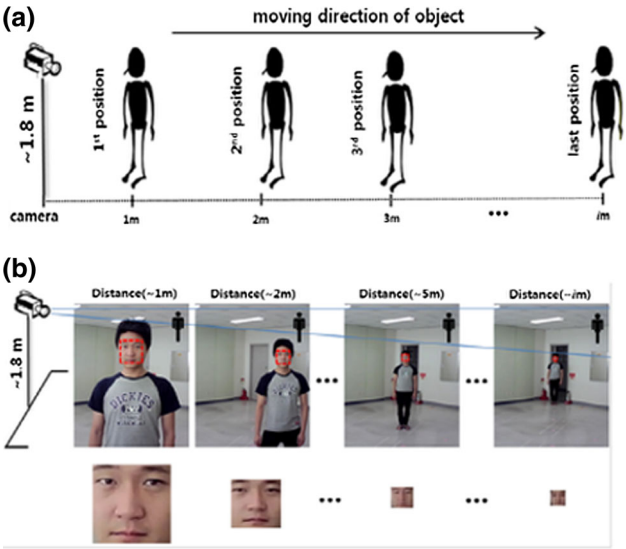

Fig. 4 How to obtain by distance face **a** method to obtain the face image by actual distance through the camera, **b** example of multiple distance image

3.1 How to obtain face images by distance for training images

The proposed long-distance face recognition method uses face image by distance as a training data. A method to use the face image by distance as a training obtains the face image by moving from 1 m to the maximum distance where the face can be detected from the camera directly by the user. Figure 4a is a method to obtain the face image by actual distance through the camera, by moving from 1 m to *i*-m in the 1 m interval directly by a person while the position of camera is fixed. Figure 4b is an example of obtaining face image by distance from 1 m to *i*-m. At this time, the position of camera is installed at 1.8 m high from the ground. When comparing face images by each distance, the sizes of extracted face images are different. A method using face image by actual distance has the advantage of obtaining the accurate short and long-distance face image. However, it has the disadvantage that the additional cooperation from the user is required compared to the method using the single distance face image.

3.2 Extracting feature vector of the face image using CNN

Figure 5 shows a process of extracting feature vector of the face image for training in the proposed face recognition using CNN. At this time, CNN is composed of five layers. Zero layer where the original image by distance with different sizes are entered, is not included in the structure of CNN. The 1st layer, as a input layer, normalizes the size of face image by each distance into the same size. The 2nd layer is a convolution layer, 3rd layer is a sub-sampling layer, 4th layer is convolution layer and the 5th layer is the sub-sampling layer. Data used in the training per person use total of 9 face images including 60 × 60 size of face image by distance. Changes in the image size are set to produce 4 output maps from 1 input map by performing convolution with 5 × 5 kernel size in the 2nd layer when the input images of 9@60 × 60

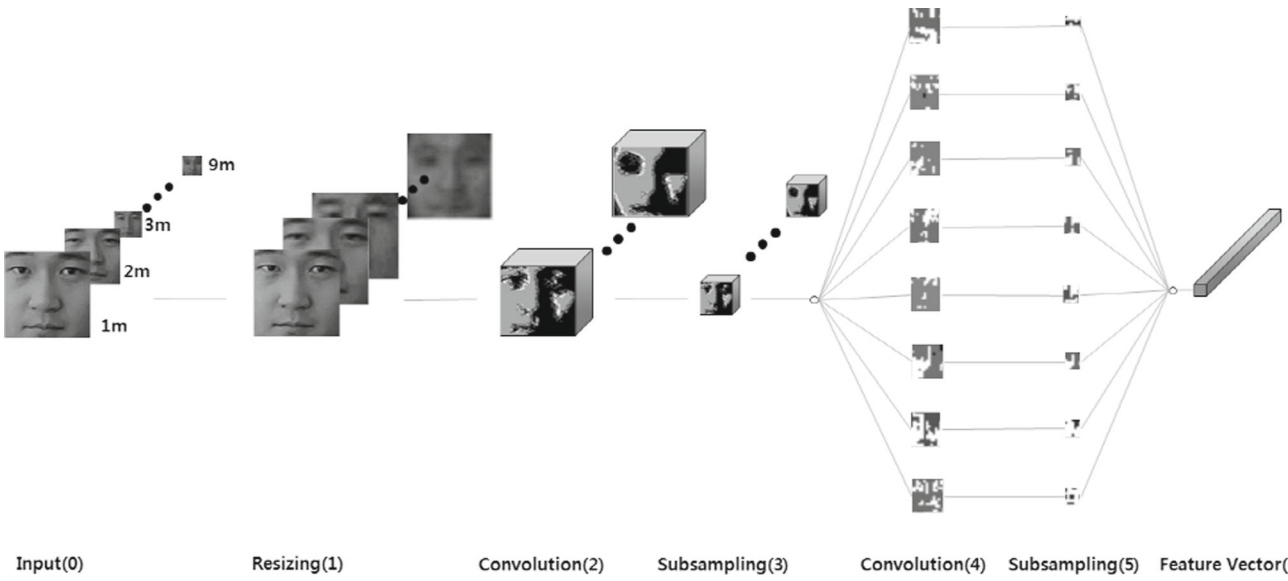

Fig. 5 Example of extracting feature vector of the face image by CNN for training

size are entered in the input layer. That is, the output of final image is 9@56 × 56 × 4 size. Then, in the 3rd layer, the sub-sampling is performed to reduce the size of the image while having invariant to movement, rotation and size changes. At this time, since the kernel size is 2 × 2, the image size is reduced by half as 9@28 × 28 × 4. In the 4th layer, the convolution with kernel size 5 × 5 is performed, and the output image is 9@24 × 24 × 8. By performing sub-sampling of 2 × 2 kernel size in the 5th layer, the output image is reduced by half as 9@12 × 12 × 8. When configuring this image as a single vector, it becomes 9@1152 × 1 dimensional vector. Therefore, the feature vector per person used in a training is 9@1152 dimension.

4 Experimental results

4.1 Face database

Traditionally, there are face recognition experiments, which commonly used the face DB including Yale DB, MIT Face DB and FERET DB. The conventional face DB includes lighting, twisted face and facial expression changes as external changes, but there is no DB considering the face changes according to the distance changes. In this paper, the face DB is configured directly and used by considering the face recognition situation in the actual indoor environment. In addition, this experiment configures DB by extracting the face region directly with the assumption that faces are all detected from the input images regardless of the distance. If the face is extracted manually, it has a feature of extracting the face region delicately than the automatic face detection method. In this experiment, the original image is used as it is without considering the twisting or rotation of the extracted face image.

IPES-1280 face DB is composed of face images taken in the interior environment for 12 candidates. Table 1 shows the configuration of IPES-1280 face DB and Fig. 6 is the normalized image based on 1 m face size. IPES-1280 face DB is acquired by a frame partitioning through video. Each image includes the distance and lighting changes and the distance changes from 1 to 9 m. Lighting changes include the changes of indoor lighting that varies depending on the distance. At this moment in time, the illumination change occurs according to the position change between human and indoor light because of fixed indoor lights as shown in 8 m of Fig. 6. During each shot, people have been asked in their free talking, rotate and twisting the head from 0 to −20 degrees, again to 0, then to +20 and back to 0 degrees with different facial expressions. The test image per one candidate is total 270 images by 30 images per distance, and the total number of test images for entire candidates is 3240 images.

Table 1 IPES-1280 FACE DB

Category	Specification
Captured environment	1–9 m distance change
	Face position change
	Facial expression change
	Illumination change at a distance
Image acquiring method	Frame division from video
Image resolution	1280(W) × 720(H)
Distance change	1–9 m (1 m interval)
Total person	12
The number of face images per person	1–9 m: 30 images each

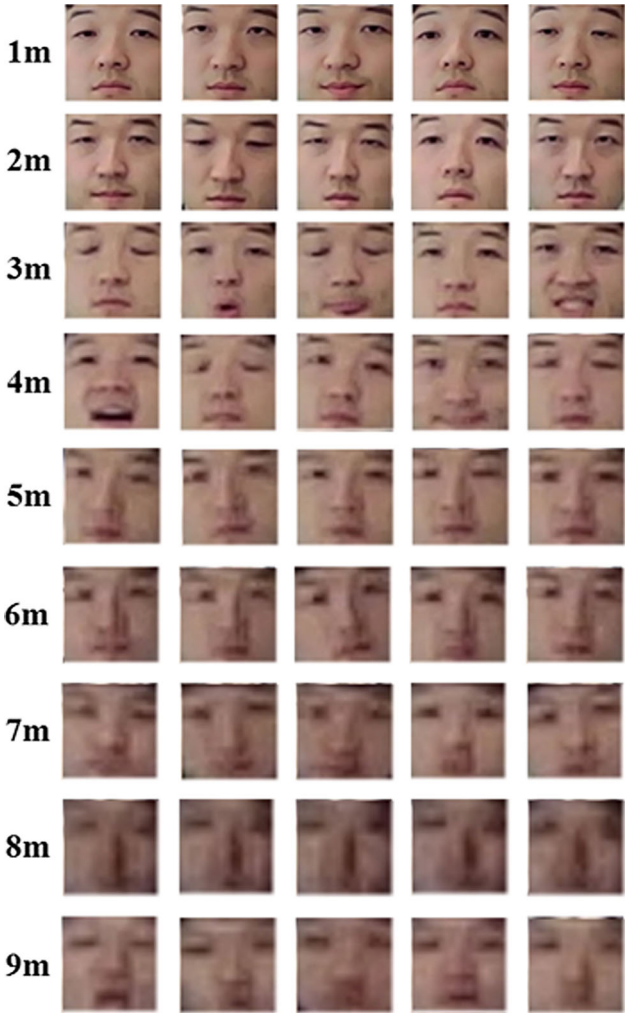

Fig. 6 IPES-1280 face DB

4.2 Performance evaluation

The proposed method compares PCA, LDA and CNN that use the existing single distance face image as a training with

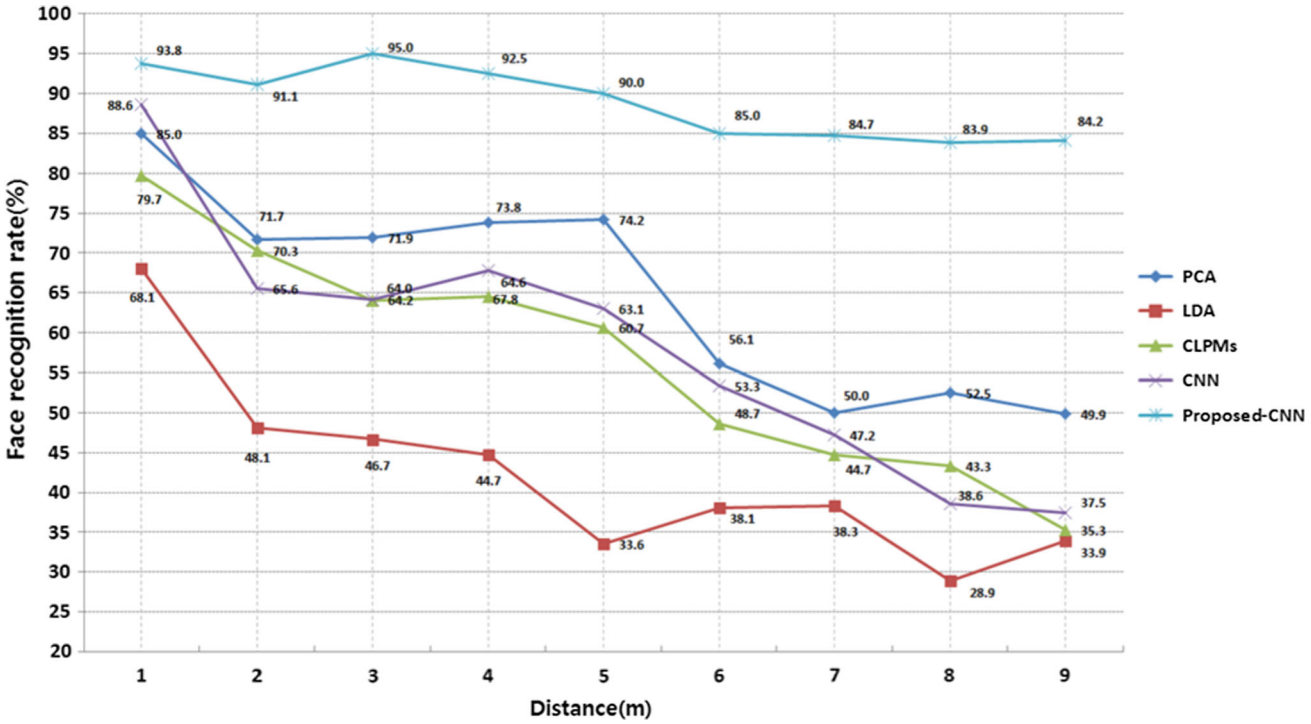

Fig. 7 The face recognition rate by distance in IPES-1280 face DB

CLPMs performance which is a face recognition method using structural features Li et al (2010). CLPMs shows the best performance by an average of 90.1 % in the low-resolution face recognition method Wang et al (2014). The proposed method uses the bilinear interpolation for face size normalization by distance and the face recognition method uses CNN. Face recognition, which is 1:N search method rather than 1:1 authentication, uses the method to classify the results for verified images of first face images which are the most similar one among face images stored in DB. A method for classification used in this study is Euclidean distance between vectors previously generated for learning and a vector generated for recognition Duda et al (2012). The close distance is accepted for a result of recognition. Euclidean distance is expressed by the following (3):

$$d(X, Y) = \sum_{i=1}^{n} | x_i - y_i | \tag{3}$$

Figure 7 shows the face recognition rate by distance in accordance with the face recognition method. If only a single distance face image is used as a training image, a method that uses the existing OCA shows the best performance in the overall average of 65.0 %, followed by a method using CNN which shows excellent performance as average of 58.4 %. When comparing the performance of the overall face recognition rate, the proposed CNN-based face recognition method using face image by distance as a training showed the best performance as average of 88.9 %. However, the proposed method requires additional user cooperation to obtain the face images by actual distance.

5 Conclusions

The paper performed the tests using various training image composition methods, preprocessing and face recognition methods to analyze the long-distance face recognition method applicable to a robot system environment. For long-distance face recognition test, face images per actual distance were obtained directly from 1 to 9 m away. The paper used face images by actual distance to resolve the issue arising from distance that affects face recognition rate. Besides, face image size normalization used bilinear interpolation while histogram equalization was applied as image preprocessing. CNN were used for face recognition while Euclidean distance was used for similarity measurement. The proposed method showed the best performance with an average of 88.9 % among face recognition method. In general, since the long distance face recognition is conducted in an uncooperative user environment, if only one image is used, there is a limit in the recognition performance. In the future, our plan is to study a method that can improve the recognition performance by processing multiple images at once in the uncooperative environment.

Acknowledgements The work was supported by Next-Generation Information Computing Development Program through the National

Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology, Republic of Korea (2011-0029927) and the Ministry of Trade, Industry and Energy(MOTIE) and Korea Institute for Advancement of Technology (KIAT) through the Promoting Regional specialized Industry.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Abdullah-Al-Wadud M, Kabir MH, Dewan MAA, Chae O (2007) A dynamic histogram equalization for image contrast enhancement. IEEE Trans Consum Electron 53(2):593–600

Bartlett MS, Movellan JR, Sejnowski TJ (2002) Face recognition by independent component analysis. IEEE Trans Neural Netw 13(6):1450–1464

Belhumeur PN, Hespanha JP, Kriegman D (1997) Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell 19(7):711–720

Bevilacqua V, Cariello L, Carro G, Daleno D, Mastronardi G (2008) A face recognition system based on pseudo 2d hmm applied to neural network coefficients. Soft Comput 12(7):615–621

Byeon YH, Kwak KC (2014) Facial expression recognition using 3d convolutional neural network. Int J Adv Comput Sci Appl 5(12):107–112

Chellappa R, Wilson CL, Sirohey S (1995) Human and machine recognition of faces: a survey. Proc IEEE 83(5):705–741

Chen CH, Yao Y, Chang H, Koschan A, Abidi M (2013) Integration of multispectral face recognition and multi-ptz camera automated surveillance for security applications. Cent Eur J Eng 3(2):253–266

Chen X, Liu W, Lai J, Li Z, Lu C (2012) Face recognition via local preserving average neighborhood margin maximization and extreme learning machine. Soft Comput 16(9):1515–1523

Duda RO, Hart PE, Stork DG (2012) Pattern classification. Wiley, New Jersey

Gonzalez RC (2009) Digital image processing. Pearson Education India

Kim HJ, Kim D, Lee J, Jeong IK (2015) Uncooperative person recognition based on stochastic information updates and environment estimators. ETRI J 37(2):395–405

Lawrence S, Giles CL, Tsoi AC, Back AD (1997) Face recognition: a convolutional neural-network approach. IEEE Trans Neural Netw 8(1):98–113

Li B, Chang H, Shan S, Chen X (2010) Low-resolution face recognition via coupled locality preserving mappings. IEEE Signal Process Lett 17(1):20–23

Lv G (2011) Recognition of multi-fontstyle characters based on convolutional neural network. In: 2011 Fourth International Symposium on Computational Intelligence and Design (ISCID), IEEE, vol 2, pp 223–225

Matsugu M, Mori K, Mitari Y, Kaneda Y (2003) Subject independent facial expression recognition with robust face detection using a convolutional neural network. Neural Netw 16(5):555–559

Moon HM, Shin J, Shin J, Pan SB (2015) User authorization method based on face recognition for auto network access in home network system. Res Briefs Inf Commun Technol Evol 1(2015):1–13

Niu XX, Suen CY (2012) A novel hybrid cnnsvm classifier for recognizing handwritten digits. Pattern Recognit 45(4):1318–1325

Park U, Choi HC, Jain AK, Lee SW (2013) Face tracking and recognition at a distance: a coaxial and concentric ptz camera system. IEEE Trans Inf Forensics Secur 8(10):1665–1677

Parzen E (1962) On estimation of a probability density function and mode. Ann Math Stat 33(3):1065–1076

Turk M, Pentland A (1991) Eigenfaces for recognition. J Cogn Neurosci 3(1):71–86

Wang Z, Miao Z, Wu QJ, Wan Y, Tang Z (2014) Low-resolution face recognition: a review. Vis Comput 30(4):359–386

Yann L, Leon B, Yoshua B, Patrick H (1998) Gradient-based learning applied to document recognition. Proc IEEE 88(11):2278–2324