A POSE ROBUST MULTI-VIEW FACE RECOGNITION SYSTEM USING PLANE OF POSE TOLERANCE

Bongjoe Kim(*), Hyoungchul Shin(**) and Kwanghoon Sohn(**)

Biometric Engineering Research Center(*),
Dept. of Electrical & Electronic Eng., Yonsei University(**)
134 Shinchon-dong, Seodaemun-gu Seoul, 120-749, Korea
E-mail: khsohn@yonsei.ac.kr

ABSTRACT

In this paper, we propose a pose robust multi-view face recognition system using Plane of Pose Tolerance (PPT). The proposed system defines the PPT which represents the tendency of pose variation well presented in multiple images. Compared with the traditional multi-view face recognition system, the proposed system is more robust and accurate especially to the error caused by pose estimation since it uses the property of multiple images. We obtain 91% face recognition rate for the proposed system and accomplished 15% improvement when compared with the traditional one.

***Index Terms*—**Plane of Pose Tolerance, multi-view

1. INTRODUCTION

In recent years, a number of face recognition systems have been developed. Most research results are focused on fontal face. Although they show good performance for frontal face, they have limited performance for the recognition of pose varied faces. Due to head rotation, the facial components undergo considerable changes of position and it leads to poor recognition performance. To cope with pose problem, a number of face recognition methods have been proposed, and it can be divided into two categories. 1) single-view based method 2) multi-view based method. In single-view based methods, they focus on extracting invariant features with respect to change of pose [1] or synthesizing a frontal face from the input image [2]. Such methods work well for small pose variation, but they fail under large pose variation because the important facial components are invisible. Thus face recognition system based on more than a single-view is emerging topic.

In the multi-view method, some of multi-view methods make a single classifier to deal with all views of faces [3]. Obviously, it leads to poor performance in most cases because one classifier can not sufficiently describe the change of appearance caused by pose variation. On the other hand, most of multi-view methods are to build several classifiers, each of them corresponding to a specific view [4]. Although this system performs better than a single classifier system, the recognition process is more complicate and even more time consuming. The previously mentioned multi-view systems have two problems. First, they need a lot of training faces with pose variation. Acquisition of faces with pose variation is a difficult work. Second, they do not use multiple images at all, thus they merely repeat the same process performed in single-view face recognition system. It neglects the property of multiple images. To overcome these problems, we propose a pose robust multi-view face recognition system, where training faces are easily acquired by means of 3D face data. Recognition is performed by using PPT which efficiently represents the tendency of pose variation presented in multiple images.

In the rest of this paper, section 2 briefly describes the background of the proposed system. In section 3, we explain the way of making database, and describe the method of measuring error using PPT. Section 4 describes the experimental results applying the proposed system. We draw our conclusion in section 5.

2. BACKGROUND

2.1. Database for multi-view system

Many face databases have been constructed in specially designed studios with various poses. However, none of these databases yet satisfies a large variation of poses. It is primarily caused by the difficulty of data collection of face images which satisfy the large variation of poses. Many approaches have been researched to obtain a database which satisfies a large variation of poses. Some approaches such as novel view generation with given limited view were proposed [5]. These approaches have a problem of obtaining a highly accurate face image. There are other approaches such as 3D morphable models [6]. The 3D morphable model can describe the 3D characteristic of human face for recognition purpose. It shows quite good results. However, due to their heavy computational cost, it is still practically unavailable.

2.2. Multi-view face recognition

One of the most significant problems in the face recognition is the variation of pose. It is not surprising that the performance of face recognition system drops significantly when large pose variations are presented in the input image. This is the basic reason why single-view face recognition systems have a fundamental limitation under pose varying environment. Thus a multi-view face recognition system has been of great interest in recent years. In the traditional multi-view systems, they merely repeat the same recognition process performed in single-view systems as follows:

$$d_m = \sum_{l=1}^{L} w^l d_m^l \tag{1}$$

, where d_m is the sum of distances between features of input images and features of the m^{th} subject in the database. L is the number of input images and w^l is a weighting factor. d_m^l is the distance between a feature of the l^{th} input image and corresponding feature of the m^{th} subject in the database. Finally, recognition is performed by selecting minimum d.

Unlike the traditional multi-view systems, some researchers proposed new approaches. They attempted to use the property presented in multiple images. Yamaguchi et al. [7] presented a method for face recognition from sequences by building a subspace for the detected faces on the given sequence and then matching the subspace with prototype subspace. Yongmin Li et al. [8] introduced an approach that models the dynamic of human faces from video sequences in a consistent spatio-temporal context and then matches the object trajectory to a set of identity model trajectories.

3. MULTI-VIEW FACE RECOGNITION WITH PPT

3.1. Generation of faces with pose variation

It is an arduous work to acquire a lot of training faces for a multi-view face recognition system. However, we easily solve this problem by means of normalized 3D face data. Fig. 1 shows a normalized 3D face data. When we obtain the 3D face data from 3D sensor, we need preprocessing such as noise filtering and normalization. 3D face data include both geometric and corresponding texture information. Due to wide and different range of geometric information, face recognition rate decreases without adequate normalization. For robust face recognition, the normalization of 3D face data is especially required. In this paper, we perform 3D face normalization proposed in [9].

Given a 3D face F, which is defined by a point set of Cartesian coordinates, we consider the range of coordinates on x-, y- and z-axis being infinite. However, we normalize these data to a defined range for each axis. All the faces that we consider are in this normalized space and are

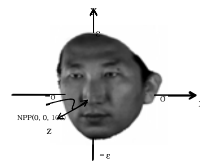

Fig. 1 Normalized 3D face data

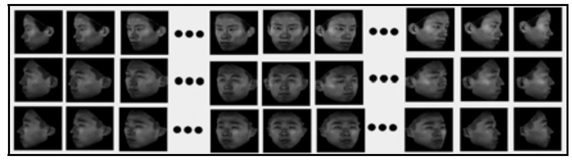

Fig. 2 Generated face images with pose variation

proportionally located based on the average face data in the limited range of $[-\sigma, \sigma]$, $[-\varepsilon, \varepsilon]$, $[0, Z]$ for x-, y- and z-axis, respectively. We especially locate the nose peak point as a fixed point on z-axis and the minimum point of the nose ridge on y- and z-plane. We first normalize 3D faces with depth information (z value) and then proportionally adjust the x and y range. We acquire the limited range for each axis as follows:

$$F(x_i, y_i, z_i) = \left(\frac{F_x - F_{\min\_x}}{F_{\max\_x} - F_{\min\_x}} \times \sigma, \frac{F_y - F_{\min\_y}}{F_{\max\_y} - F_{\min\_y}} \times \varepsilon, \frac{F_z - F_{\min\_z}}{F_{\max\_z} - F_{\min\_z}} \times Z \right) \tag{2}$$

, where $F(x_i, y_i, z_i)$ is normalized input data point, F_x, F_y and F_z are input data points for each axis. $F_{\max\_x}$, $F_{\max\_y}$ and $F_{\max\_z}$ are maximum values, and $F_{\min\_x}$, $F_{\min\_y}$ and $F_{\min\_z}$ are minimum values for the each x-,y- and z-axis, respectively. By projecting this normalized 3D face onto a 2D image plane, we generate sets of 37 2D face images for each person from $-90^\circ \sim +90^\circ$ in horizontal rotation with an interval of 5°. Fig. 2 shows the example of generated face images. Since these images have accurate change of facial appearance caused by rotation of head, the database constructed from these images increase the recognition rate.

3.2. Recognition using PPT

In this paper, we use Principal Component Analysis (PCA) to extract features of input image [10]. After extracting features by using the first 3 eigenvectors, a face image is converted into a vector with 3 components. Thus, a face image is represented as a point in three-dimensional space. Three-dimensional feature space can not sufficiently describe face images. However, three-dimensional feature space can identify face images using the tendency of pose variation presented in multiple images. To make a database, sets of 37 face images of a person are projected into the three-dimensional feature space and each set forms a trajectory. Each point on a trajectory corresponds to a certain pose of a face. In the traditional multi-view system, the recognition process typically follows three steps. First, they estimate poses of input faces, and extract features.

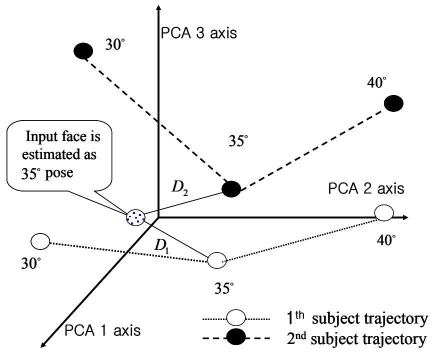

Fig. 3 Method of measuring error in the traditional system

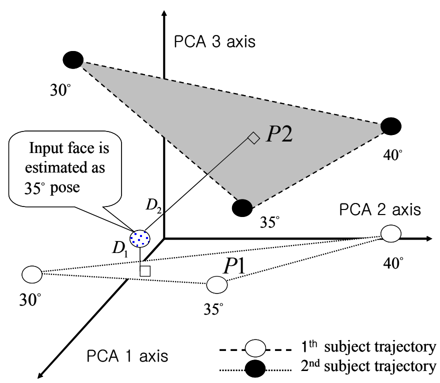

Fig. 4 Method of measuring error in the proposed system

Finally they measure sum of distances between features of input images and features which correspond to input pose in the database. Fig. 3 illustrates a method of measuring error in the traditional multi-view system. The input face estimated as 35° pose variation is projected into three-dimensional space. Then the Euclidean distance between point of input face and corresponding point of each subject is calculated shown as D_1, D_2 in Fig. 3. Finally, recognition is performed by selecting minimum D. However, this approach neglects the property of multiple images and does not utilize neighboring images at all. Additionally, it is assumed that there is no error at the stage of pose estimation. It may lead to poor recognition rate. One possible solution is to make a mean point by averaging the point which corresponds to input pose and neighboring two points on the subject's trajectory. Recognition is performed by matching points of input images to mean points on the subject's trajectories. It is simple and intuitive, but it merely averages the property of multiple images.

We propose a new solution which efficiently utilizes the property of multiple images well. The proposed solution is to make a plane with the point corresponding to input pose and neighboring two points on the subject's trajectory. We call this plane as PPT. Recognition is done by matching points of input images to theses planes. Since PPT contains the tendency of pose variation, this plane might have the tolerance about pose estimation error. The error criterion

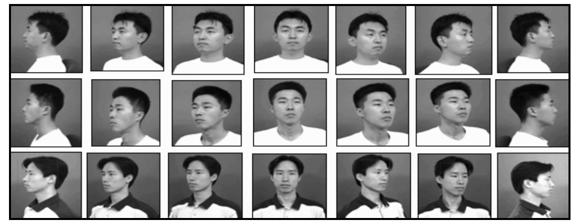

Fig. 5 Example of input face images

using PPT is the distance between the input images and the m^{th} subject as follows:

$$d_m = \sum_{l=1}^{L} \left| \frac{\overrightarrow{N_m^l} \bullet \overrightarrow{R_m^l X_{input}^l}}{\left|\overrightarrow{N_m^l}\right|} \right| \quad (3)$$

, where d_m is sum of distances between input images and the m^{th} subject in the database. L is the number of input images. X_{input}^l is a point of the l^{th} input image in the feature space. $\overrightarrow{N_m^l}$ is the normal vector of PPT corresponding to the l^{th} input image in the m^{th} subject's trajectory, and R_m^l is an arbitrary point on this PPT. Finally, recognition is performed by selecting minimum d. Fig. 4 illustrates the method of measuring error in the proposed system. The input face estimated as 35° pose variation is projected into three-dimensional space. Then we make planes using 30° point, 35° point, and 40° point on each subject's trajectory shown as *P1* and *P2* in Fig. 4. The Euclidean distances between the points of input faces and the corresponding planes of each subject are calculated shown as D_1, D_2. Finally, recognition is performed by selecting minimum D.

4. EXPERIMENT RESULTS

We utilize Visual studio 6.0 and OpenGL for simulation. We use 3D face data set of Biometric Engineering Research Center (BERC). These 3D face data are acquired by Cyberware Model 3030PS/RGB laser scanner. We use 3D face data of 100 persons. Each 3D face data is projected into 2D space to generate 37 face images with pose variation. The training set in this paper contains 3700 face images taken from 100 subjects, 37 views of each. We use 7 face images captured by multi camera system as input images for recognition. Since we focus on recognition stage, pose estimation and preprocessing are performed by semi manual. Fig. 5 shows the example of input images. The size of all the images is 120×100.

To demonstrate the effectiveness of the proposed system, the recognition rate based on (3) is compared with the recognition rate obtained from (1). From Table 1, it is evident that the recognition rate for the proposed system increases by 15% more than for the traditional system. Fig. 6 shows the experimental result of the 16th subject for both the proposed system and the traditional system. The

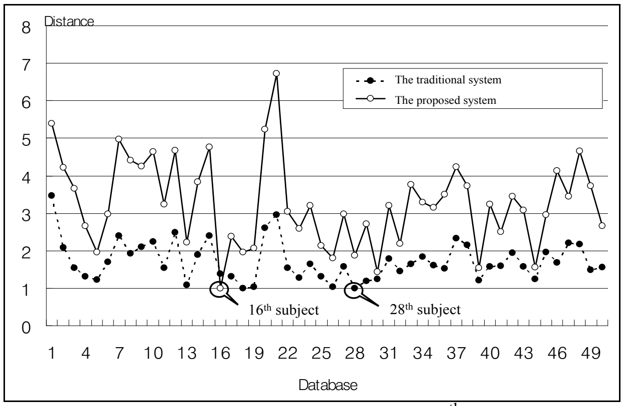

Fig. 6 Recognition result for the 16th subject

proposed system using PPT recognizes it correctly. On the other hand, the traditional system fails to recognize. It is noticeable that the variance of distance of the proposed system is much larger than that of the traditional system. It means that the proposed system decreases the distance for the true subject, however, it increases the distances for other subjects. We analyze the pose robustness of the proposed system based on the sensitivity to the error of pose estimation. The robustness is evaluated in terms of the recognition rate under the existence of pose estimation error. As shown in Table 2, the recognition rate of the proposed system is decreased by 2% under $\pm 5^{\circ}$ error of pose estimation and 8% under $\pm 10^{\circ}$ in comparison with the ground truth pose. The recognition rate of the traditional system is decreased by 14% under $\pm 5^{\circ}$ error of pose estimation and 22% under $\pm 10^{\circ}$ in comparison with the ground truth pose. The recognition rate of the traditional system steeply decreases under the existence of pose error. However, the proposed system is more robust than the traditional system because the proposed system utilizes the tendency of pose variation based on the concept of the proposed PPT.

5. CONCLUTION

In this paper, we propose a pose robust multi-view face recognition system using Plane of Pose Tolerance. Acquisition of multiple images and utilization of the property of multiple images are the main problems in the traditional multi-view face recognition system. To overcome these problems, we use normalized 3D face data to acquire multiple images. And we propose new method of measuring error using PPT. We have shown the effectiveness of the proposed system through the experimental results. Recognition rate remarkably increases by 15% in comparison with the traditional system. Especially, the proposed system is very stable against the influence of head rotation because the PPT contains the tendency of pose variation. Further work is to develop a multi-view system which can handle up-down pose variation.

Table 1 Recognition rate using different multi-view system

	Traditional	PPT
Recognition Rate	76% (24 person fail)	91% (9 person fail)

Table 2 Recognition rate using different multi-view system under existence of error of pose estimation

	Ground-truth pose	$\pm 5^{\circ}$ pose error	$\pm 10^{\circ}$ pose error
Traditional	76% (24 person fail)	62% (38 person fail)	54% (46 person fail)
PPT	91% (9 person fail)	89% (11 person fail)	83% (17 person fail)

ACKNOWLEDGMENTS

This work was partially supported by the Korea Science and Engineering Foundation (KOSEF) through the Biometrics Engineering Research Center (BERC) at Yonsei University. This work has been also partially supported by Yonsei University Institute of TMS Information Technology, a Brain Korea 21 program, Korea.

6. REFERENCES

[1] L. Wiskott, J.-M. Fellous, N. Kruger, and C. von der Malsburg, “Face recognition by elastic bunch. graph matching,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), pp.775-779, 1997.

[2] Andreas Lanitis, Christopher J. Taylor, Timothy F. Cootes. “Automatic face identification system using flexible appearance models,” *Image Vision Comput,* 13(5), pp.393-401, 1995.

[3] R.Gross, I.Matthews, and S.Baker. “Eigen Light-Fields and Face Recognition Across Pose,” *Int. Conf. on Automatic Face and Gesture Recognition,*, pp.3-9, 2002.

[4] A. Pentland, B. Moghaddam and T. Starner, “View-based and modular eigenspaces for face recognition,” *Inf. Conf. On Computer Vision and Pattern Recognition (CVPR),* pp.84-91, 1994.

[5] T. Vetter and T. Poggio, “Linear object classes and image synthesis from a single example image,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19, pp.733-742, 1997.

[6] Blanz, V. and Vetter. T, “A Morphable Model for the Synthesis of 3D faces,” *SIGGRAPH'99 Conference Proceeding*, pp.187-194, 1999.

[7] O. Yamaguchi, K. Fukui, and K. Maeda, “Face recognition using temporal image sequence,” *In IEEE International Conference on Automatic Face & Gesture Recognition*, pp.318-323, 1998.

[8] Li, Y., Gong, S. and Liddell, H., “Recognising Trajectories of Facial Identities Using Kernel Discriminant Analysis,” *Image and Vision Computing 21* (13-14), pp 1077-1086, 2003.

[9] H. Song, S. Lee, and K. Sohn, “3D Sensor Based Face Recognition,’ *Applied Optics*, vol. 44, no.5, pp.677-687, 2005.

[10] M.Turk and A.Pentland, “Eigenfaces for recognition,” *Journal of Cognitive Neuroscience,* 3, pp.72-86, 1991.