GENERATING OPTIMAL FACE IMAGE IN FACE RECOGNITION SYSTEM

Yingchun Li, Guangda Su,Yan Shang

Electronic Engineering Department, Tsinghua University, 100084, Beijing, China
liyingchun@mail.tsinghua.edu.cn, Sugd@ee.tsinghua.edu.cn, y-shang@tsinghua.edu.cn

ABSTRACT

In this paper, a new method of generating optimal face image is proposed for improving the performance of automatic face recognition system. In order to overcome the influences of variations in face pose and expression, it is essential to generate optimal face image for the face to be recognized. The optimal face image is the one which has the minimum pose rotation and neutral expression in image sequences. Through estimating face pose and expression and combining tensors decomposing in multi-linear algebra, we can generate optimal face images from three-route video inputs of CCD in parallel detecting system. A new module of optimal face image generation is built between face detecting module and face recognition module. The proposed method overcomes the problem which large pose angles affect face recognition rate. The recognition rate is improved after applying the proposed method.

Index Terms—face recognition, databases, CMU PIE, tensor decompose, pose estimation

1. INTRODUCTION

Automatic face detection and recognition have been an active research area in the last two decades. There are urgent needs for face recognition in many practical applications, such as security monitoring, surveillance system and biometrics identified system. The majority of research has so far focused on frontal face and neutral expression face recognition. Recognizing faces reliably across changes in pose and expression has proved to be a much harder problem [1].

The most existing face recognition systems consist of single route image acquisition module with one camera. They follows the strategy of detection face images in complicated background, normalizing the detected face image and transmitting them to recognition module to be recognized. In real surveillance applications, because of the irregular head movement, most of the face images are of various pose and expression. The recognition rate is relatively low in such dynamic system. The evaluation of FRVT shows the level of performance for face verification of the best systems to be on par with face recognition for frontal faces. With increasing of the pose angle, the recognition rate decreases. When the pose angle is larger than 30 degree, the recognition rate decreases greatly [2]. This paper proposed a new method to improve the recognition rate by selecting and generating optimal face image from serial face images. Meanwhile, a new face recognition system with three route parallel modules is constructed. In order to get optimal face, it is necessary to estimate face pose and expression from the detected face images.

Pose estimation techniques can be classified into two main categories: model-based approaches and appearance-based approaches. The former uses an explicit model of the face and recover the face pose based on the assumed model. A set of feature correspondences are established to estimate face pose [3]. The latter directly uses image pixels or features and assumes that there exists a mapping relationship between face pose and certain properties of the facial image, which is constructed from a large number of training images [4]. The popular eigenface approach is extended to handle multiple views and compared the performance of a parametric eigenspace with view-based eigenspace. Some researchers used model shape and texture nonlinearities across views in full 180 degree rotations. However, no face recognition experiments were performed. 3D can show face deep and topology information but at the cost of time [5]. It is not suitable for real-time applications.

This paper is organized as follows. Section 2 introduces structure of the new face recognition system with optimal face image generation module. A new system module is built between face detecting module and face recognition module which is called optimal face generation module. The pose estimate method is explained in Section 3. Section 4 give a detailed explanation of the optimal face generation algorithm. Through combining the pose estimate method and adopting three routes parallel detecting system, we can select the optimal face image with little view rotation (when have appropriate view) or generate the face

image with frontal view (when have no appropriate view). Experiments and conclusions are given in Section 5. The face recognition rate is improved after applying the proposed method and new system module.

2. THE SYSTEM STRUCTURE

The proposed method is applied in parallel multi-route surveillance system. Its flow chart is shown in Fig.1. The system consists of four parts: multi-route face image acquisition module, multi-route face image detection module, optimal face generation module and face recognition module.
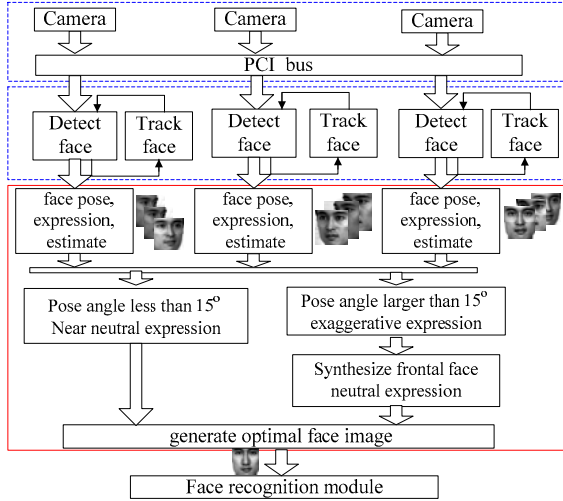

Fig.1. Face recognition system.

The face image acquisition module consists of three cameras, image collective cards and PCI bus. It utilizes cameras simultaneously capturing different pose angles from different face views in yaw. The image capturing device is shown in Fig. 2.

The three image streams are transmitted to the face detecting modules. Three cameras arranged in a row with equal intervals can extend detecting views and capture multiple face views in parallel. In Fig.2, we see that camera 1 can capture image with little pose angle even if the head is in a large rotation to camera 2. These performances are in need of real-time robust face detection module. The optimal face generation module is to select or synthesize best fit face to satisfy needs of recognition module by estimating the face pose and expression. FRVT shows that the recognition rate is relatively high when the face pose angle is between 15° and -15° and it greatly decreases after the pose angle larger than 15° [3]. Three cameras can enlarge rotation angles to 45° towards right and left. This structure is able to solve the problem of face recognition with large pose angles.

In the meantime, the optimal face generation module selects and synthesizes neutral facial expression from multiple detected face images. At last the optimal face image is acquired to be recognized in the following recognition module. This constructs a new face recognition system.
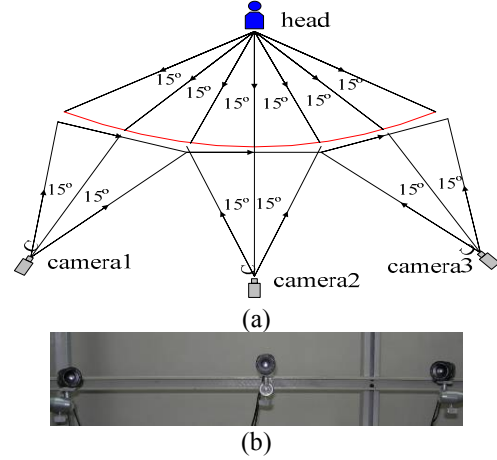

Fig. 2. Three-route face images captured device. (a)the configuration of cameras. (b) the image capturing device.

3. POSE ESTIMATION

In order to acquire optimal face with optimal face generation module, the first step is to estimate pose of the face images. The face images are preprocessed such as aligned, normalized, selected features. The system can automatically detect face locations [6] of three feature points (two eyes and chin tip) and geometrically aligns the image to candidate poses training and test of the subjects. The examples of training images in TH database (built by ourselves) are shown in Fig.3.

Fig.3. The training face images in TH database.

It is essential to extract features from images utilizing the composite PCA (principle component analysis) and projecting face images to the eigenspace. Given a set of samples $X_i \in \mathbb{R}^N$ represented face images by column vectors. The transformation matrix can be formed by using eigenvectors which normalized to unit matrix T. The projection of X_i into the N-dimensional subspace can be expressed as

$$\alpha = \{\alpha_1, \cdots, \alpha_N\} = X_i^T \cdot T \tag{1}$$

The shape feature is shown in Fig.4. The feature points can give geometric characteristic. AB and $A^{'}B^{'}$ are the distance between two eyes when pose angle is 0 and β degree respectively. Set radius as 1.

$$A'B' = \sin(\theta + \beta) + \sin(\theta - \beta) = 2\sin\theta\cos\beta \qquad (2)$$

Since distance $AB = 2\sin\theta$, then pose angle

$$\beta = \arccos(\frac{A'B'}{AB}) \qquad (3)$$

Set weights of two weight parameters α and β after two groups of features are gained. The new eigenvector ξ is

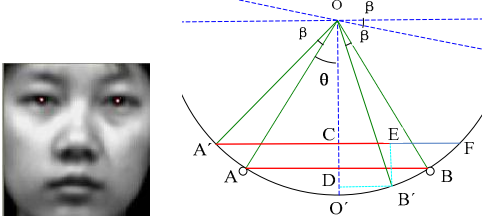$$\xi = p\cdot\alpha + q\cdot\beta \text{ , where } p + q = 1 \qquad (4)$$

Fig.4 Shape feature points and the configuration of pose variance.

SVM (support vector machine) is to find the optimal linear hyperplane which the expected classification error for unseen test samples is minimized [7]. According to the structural risk minimization principle, a function that classifies the training data accurately will generalize best regardless of the dimensionality of the input space.

Each training sample x_i is associated with coefficient a_i. Those samples whose coefficient a_i is nonzero are Support Vectors (SV) of the optimal hyperplane. $f(x)$ is an optimal SVM classified function. $y_i \in (+1,-1)$.

$$f(x) = \sum_{vector} y_i a_i K(x_i, x) + b \qquad (5)$$

Where the K is a kernel function. Here we use linear kernel, $\phi(x_i) = x_i$, then $K(x_i, x_j) = x_i \cdot x_j = x_i^T x_j$.

The PCA projection values of samples to eigenspace were as SVM input parameters and the optimal hyperplane that correctly separates data points were found. Combining the PCA and SVM classifier, we can draw better classification results. Then the pose angle will be acquired.

Besides pose estimation, the human facial expressions are also estimated. After estimate the pose and expression, we can generate optimal face by multi-linear mappings over a set of vector spaces.

4. CONSTRUCTION OF OPTIMAL FACE GENERATION

Tensor is a multidimensional generalization of a matrix. A n-mode SVD (Singular Value Decomposition) can decompose an n-dimensional tensor $\mathbb{F}$ into the n-mode product of N-orthogonal spaces. The notion of tensor can be applied to a face image in the following way [8]. Consider a set of U_{pixels} images of $U_{peoples}$ people's faces, each image in U_p poses, with U_e expressions. The facial image tensor can be decomposed into

$$\mathbb{F} = \mathbb{Z} \times_1 U_{pixels} \times_2 U_{peoples} \times_3 U_p \times_4 U_e \qquad (6)$$

Each mode matrix represents a parameter. The columns of U_p and U_e respectively span the space of pose and expression parameters. The columns of U_{pixels} span the image space. These are the eigenfaces obtained by PCA on the whole data set. The core tensor $\mathbb{Z}$ governs the interactions between these mode matrices and represents only the principal axes of variation over all images.

Each person can be represented by the same coefficient vector regardless of pose and expression. It is expressed as

$$\mathbb{F} = \mathbb{S} \times_2 U_{peoples} \qquad (7)$$

$$\mathbb{S} = \mathbb{Z} \times_1 U_{pixels} \times_3 U_p \times_4 U_e \qquad (8)$$

$\mathbb{S}$ defines different bases for each combination of pose and expression and $U_{peoples}$ contains the coefficients. Tensor $\mathbb{S}$ can be indexed into a particular pose and expression to obtain a subtensor $\mathbb{S}_{p,e}$. Then the subtensor $\mathbb{S}_{p,e(peoples)}$ was flattened along identity (people) mode, which column vectors span the space of images under particular pose and expression. Subtensor $\mathbb{S}_p$, $\mathbb{S}_e$ are in the maximum projection direction of which $\mathbb{S}_{p,e}$ projects in various types of eigenspace. Given an input image, the estimate value is computed by projecting into correlative subtensor such as pose, expression, and then identify them relatively.

The frontal faces or the faces which pose angle is less than 15° (radian is $\pi/12$) are the optimal face with optimal pose. If the pose angle of face image is over 15°, the frontal view can be synthesized by several multi-view face images in morphable model method. The optimal pose view U_{op_p} is expressed as

$$U_{op\_p} = \frac{1}{2}\left\{ U_p\left[sgn\left(\frac{\pi}{12} - |U_p|\right) + 1 \right] + U_{ps}\left[sgn\left(|U_p| - \frac{\pi}{12}\right) + 1 \right] \right\}$$
$$(9)$$

Where sgn is sign function, U_p is pose parameter, U_{ps} is synthesis parameter.

The same idea is also applied in expression estimate. Neutral expressions or little variation expressions are the optimal expression. So the faces with exaggerated expression can be synthesized to normal expression to generate optimal expression U_{op_e}.

The optimal face image is the well-posed (the yaw pose angle is within 15°) image with neutral facial expression or little variation expression. The optimal face B_{op_face} is expressed as

$$B_{op\_face} = \mathbb{Z} \times_1 U_{peoples} \times_3 U_{op\_p} \times_4 U_{op\_e} \qquad (10)$$

The optimal face image can reduce high dimensions which caused by the effects of uncertain factors such as pose and expression. It can simplify the design of the classifiers and improve their performance.

5. EXPERIMENTAL RESULTS AND CONCLUSION

We used three databases in our system across pose experiments, still face images of TH database, CMU PIE database [9] and video images of TH database. Fig.5 and Fig.6 show parts of results of pose estimation of still images in CMU PIE database and video images in TH database. The comparison results of the three tests are in Table 1.

Table 1. Comparison of pose estimation in three databases.

Test Database	images	Correct	Correct rate
TH database	1080	1059	98.1%
CMU PIE	400	386	96.9%
TH video	395	379	95.9%

Adopting the method of generating optimal face image can acquire a best fit face image among serial images. The recognition rate of still images is higher than that of video images. Because the training images are original from TH database, the quality of still image is better than that of video images. The recognition rate of still images in TH database is higher than that of in CMU PIE database. We compare another performance of the system. The test is between three-route parallel face recognition system with optimal face image generation module and single-route face recognition system. The test results are shown in Table 2.

Table 2. Comparison of recognition rates in face system.

Recognition rate Standard	No optimal face generation module	Adding optimal face generation module
TH Standard	82.57%	85.71%
FRVT Standard	78.29%	80.80%

We test recognition rates in dynamic face system according to TH (Tsinghua) standard and FRVT standard respectively. We have tested 656 individuals in pose angle no more than 60° (towards left and right). The experiment results show that the system with optimal face generation module gets higher recognition rate and outgoes the system without optimal face generation module no matter which standard is adopted. FRVT reported the test in database of 87 individuals which only have pose variance. When the pose view reached 30° or 45°, its recognition rate decreased to 45%. But our proposed method overcomes the problem which large pose angles affect face recognition rate. It helps to increase recognition rate and improve performance of system.

The key of face recognition is to acquire optimal face image. The aim of generating optimal face image is to form a fit image for recognition which has minimum pose angle and neutral expression in a serial dynamic face images.
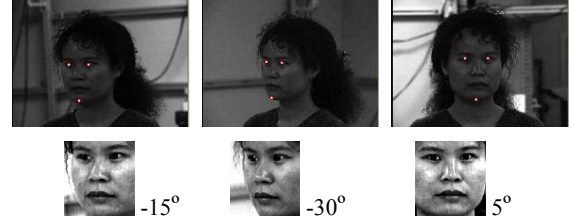

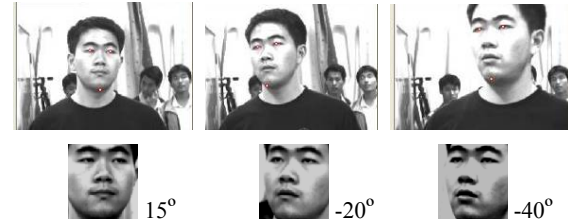**Fig.5.** Pose estimation results in CMU PIE database.

Fig.6. Pose estimation results in surveillance system.

In fact, the optimal face image generation module is a new integrated preprocessing module in face recognition system. Its idea can extend to other factors such as illumination and resolution and so on. The experiment results show that proposed method is feasible and effective. The optimal face image generation module is indispensable in the face recognition system. Future research will focus on improving its calculating speed to meet the needs of the real-time application.

6. REFERENCES

[1] Y. Wang and T. Tan, "Combining Face and Iris Biometrics for Identity Verification", *Proceedings of Audio and Video-Based Biometric Person Authentication*, pp. 805-813, 2003.
[2] P.J.Phillips, P.Grother and R.J.Micheal, Face Recognition Vendor Test 2002, *Evaluation Report*, 2003.
[3] Y.Gao, M.Leung and W.Wang, "Fast face identification under varying pose from a single 2-D model view", *IEE Proc. Vis. Image Signal Process*. Vol. 148 (4), 2001.
[4] Q. Ji, "3D Face pose estimation and tracking from a camera", *Image and Vision Computing*, Vol. 20 (7), pp.499-511, 2002.
[5] M. Turk, A. Pentland. "Face Recognition Using Eignefaces", In Proc. *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 586-591, 1991.
[6] H.Gu, G.Su and C. Du, "Feature Points Extraction from Faces", *Image and Vision Computing*, (IVCNZ'03), pp:154-158,2003.
[7] Vapnik, V., "Statistical Learning Theory", John Wiley & Sons, New York, 1998.
[8] M. Alex, O. Vasilescu and D.Terzopoulos, "Multilinear Analysis of Image Ensembles: TensorFaces", *ECCV 2002, Lecture Notes in Computer Science*, Vol. 2350, pp.447-460, 2002.
[9] T.Sim, S.Baker and M. Bsat, "The CMU Pose, Illumination, and Expression Database", *IEEE TPAMI,* Vol. 25(12), pp.1615-1618, 2003.