

# Frontal Face Synthesis Based on Multiple Pose-Variant Images for Face Recognition

Congcong Li, Guangda Su, Yan Shang, and Yingchun Li

Electronic Engineering Department, Tsinghua University, Beijing, 100084, China  
lcc@mails.tsinghua.edu.cn

**Abstract.** Pose variance remains a challenging problem for face recognition. In this paper, a stereoscopic synthesis method for generating a frontal face image is proposed to improve the performance of automatic face recognition system. Through this method, a frontal face image is generated based on two pose-variant face images. Before the synthesis, face pose estimation, feature point extraction and alignment are executed on the two non-frontal images. Benefited from the high accuracy of pose estimation and alignment, the composed frontal face retains the most important features of the two corresponding non-frontal face images. Experiment results show that using the synthetic frontal image achieves a better recognition rate than using the non-frontal ones.

**Keywords:** Face recognition, pose estimation, face alignment, stereoscopy, texture synthesis.

## 1 Introduction

In recent years, techniques on face recognition have been developed a lot. Many algorithms have achieved satisfying recognition performance in controlled conditions under which faces are in frontal pose with harmonious illumination and neutral expression. However, there are still many open problems when face recognition technology is put into real world applications. The face recognition vendor test (FRVT) 2002 reports [1] that recognition under illumination, expression and pose variations still remains challenging. Results show that recognition rate decreases sharply when face rotates to a large angle.

Aiming to improve face recognition performance under pose variance, face synthesis has been considered a lot. There have been considerable discussions of synthesizing face image in novel views which can be roughly divided into two categories: those based on 3D head model and those based on 2D image statistics [2].

One common way to synthesize novel views of a face is to recover its 3D structure. Many current algorithms utilize a morphable 3D model to generate face images of novel views from a single image. These methods face a problem in common: when only one face image is available, the texture within occluded region becomes undefined. Vetter et al. [3, 4] use the linear object class approach to deal with this

problem. It is assumed that a new face's texture can be represented as a linear combination of the texture from a group of example faces in the same view, and the combination coefficients can be used to synthesize the face image in another view. However, another difficulty appears. Since the generated texture is a linear combination of the textures in training database, some individual characteristic information would be lost, such as little scar, beauty spot, and so on.

Methods based on 2D images can also produce promising results for face synthesis. Stereopsis and projective geometry are often utilized. The utilization of these techniques highly depends on whether most face region is visible in all source images. While building statistical models for face images, AAM (Active Appearance Models) [5] and ASM (Active Shape Models) [6] are widely used. Both methods can help extract corresponding feature points for different facial images.

In this work, we present a whole scheme from pose-variant image capture to novel frontal image synthesis and recognition. This scheme generates a frontal facial image based on multiple pose-variant facial images. Pose estimation algorithm and ASM algorithm for face alignment are both improved to ensure the synthetic accuracy. The synthesis stage is divided into shape reconstruction and texture synthesis. Shape reconstruction is carried out mainly based on stereoscopy and partly assisted by a trained model. The proposed scheme solves the following problems:

1. It transforms the 2-D rotated facial images to a 2-D frontal view which is proven more suitable for face recognition by our experimental results.
2. Since the synthesis is based on more than one image, the texture contained in the two input images can cover most of the face so that details are retained;
3. It overcomes the computationally extensive problem of synthesis, and thus is suitable for real-time face identification applications;

The rest of the paper is organized as follows: Section 2 provides an overview of the proposed method; Section 3 introduces briefly the pose estimation method and Section 4 introduces the face alignment method, both of which affect the final accuracy of face reconstruction. Section 5 demonstrates the synthetic process, including shape reconstruction and texture synthesis. Experiment results are given in Section 6, and this paper is concluded in section 7.

## 2 Overview of the Proposed Scheme

Aiming to generate a frontal-pose face image for recognition, an integrated scheme is designed, including image collection, image preprocessing, pose estimation, face alignment, facial shape reconstruction and texture synthesis, as shown in Fig. 1.

The Multi-Channel Image Input & Process Module by Tsinghua University captures face image and performs face detection parallelly in four channels. Face synthesis will be carried on based on the input images with the help of pose estimation and face alignment.

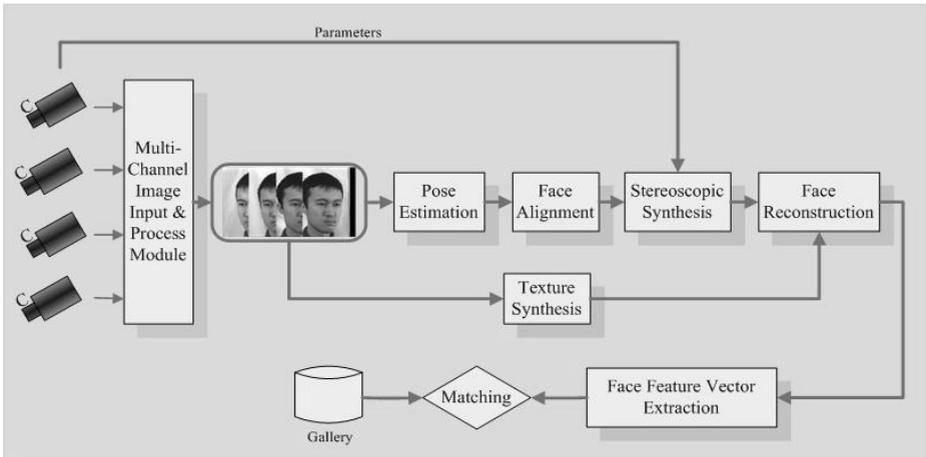


Fig. 1. Overview of the proposed scheme

### 3 Pose Estimation

In order to reconstruct a frontal face, the first step is to estimate poses of the input face images. The face images are preprocessed through feature-positioning and normalization. They are rectified and normalized geometrically according to the automatically located key-point positions. In the geometric normalization step, not only the eyes but also the chin middle point are located automatically. Then each face image is scaled and rotated so that the eyes are positioned in a horizontal line and the distance between the chin point and the center of the eyes equals a predefined length. After that, the face image is cropped to a given size. The examples of training images in TH face database [7] are shown in Fig.2. We collect 780 images of 60 people with left-right poses ranging from -45 degree to 45 degree at intervals of 15 degree and with up-down poses ranging from -20 degree to 20 degree at intervals of 10 degree.

It is essential to extract features from images utilizing the composite PCA (principle component analysis) and projecting face images to their eigenspace. Given a set of samples  $X_i \in \mathbb{R}^N$  represented face images by column vectors. The transformation



Fig. 2. The training face images in TH database and their corresponding normalized images

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

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

The shape feature is shown in Fig. 3. The feature points represent geometric characteristic.  $AB$  and  $A'B'$  are the distance between two eyes when pose angle is 0 and  $\beta$  degree respectively. Set radius as 1.

$$A'E' = A'B' = \sin(\theta + \beta) + \sin(\theta - \beta) = 2 \sin \theta \cos \beta \tag{2}$$

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

$$\beta = \arccos\left(\frac{A'E'}{AB}\right) \tag{3}$$

Set weights of two weight parameters  $\alpha$  and  $\beta$  after two groups of features are gained. The new eigenvector  $\xi$  is

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

SVM (support vector machine) is used to find the optimal linear hyperplane by which the expected classification error for unseen test samples is minimized [8]. 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.  $y_i \in (+1, -1)$ .

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

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

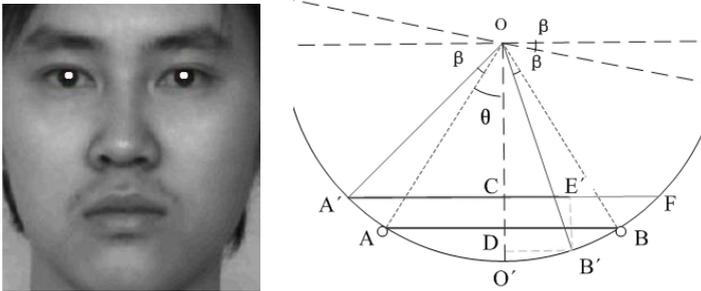


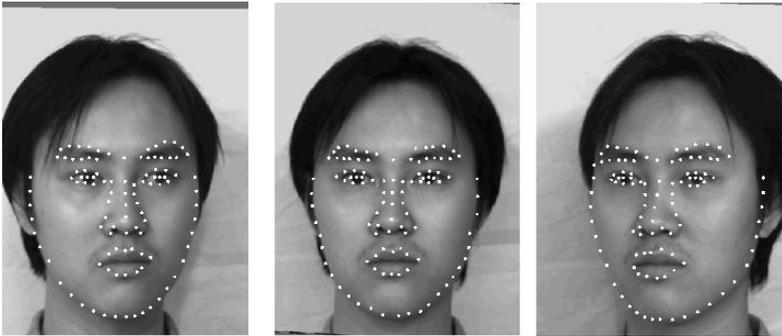
Fig. 3. Shape feature points and the configuration of pose variance

The PCA projection values of samples to eigenspace are used as SVM input parameters and the optimal hyperplane that correctly separates data points is found. Combining the PCA and SVM classifier, we can draw better classification results, thus more accurate pose angle estimation.

## 4 Face Alignment

An improved ASM (active shape models) method is chosen to extract the face feature points in this paper. It is hard for the conventional ASM to get accurate result on each feature point; what's more, the performance depends heavily on the initial positions of the landmarks. According to the facial structure, the edge information and facial part information are introduced to the matching process of ASM, which improved the performance of ASM [9, 10].

The face images are first normalized and totally 105 feature points extracted by the improved ASM algorithm are selected to represent the face shape feature, as shown in Fig. 4.



**Fig. 4.** Feature points extracted by the improved ASM algorithm

Although the improved ASM algorithm provides fine alignment, the accuracy of points can still be ameliorated. For images with different left-right poses, feature points with the same definition in different images should have the same y-coordinate value. However, we found the ASM results did not always meet this principle, although they together can always well describe contours of facial parts.

Therefore, to ensure the corresponding relation between feature points in different images, feature points of some facial parts like face contour are connected and fitted by polynomial curve. As shown in Fig. 5, half of the contour is fitted at two stages. At the first stage shown in the left, feature points are roughly fitted and the point parallel to the corner of mouth is considered to be a subsection point. Then the contour is fitted separated by two curves as shown in the right. Then the contour feature points are adjusted on the curves in well-proportioned distribution. Besides fitting the contour, similar operations are also carried onto the other facial parts. The alignment of the face feature points provides an important basis for shape reconstruction of frontal face, which will be mentioned in the next section.

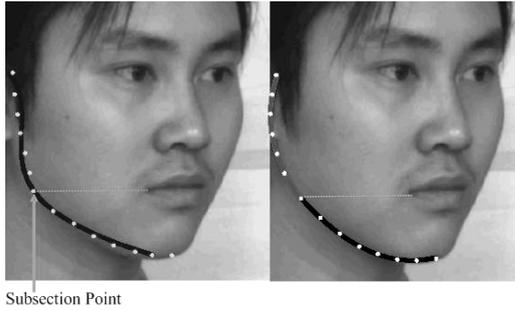


Fig. 5. A two-stage fitting to the contour with polynomial curves

## 5 Face Synthesis

A face image can be separated into shape information and texture information. If we have these two kinds of information, we can reconstruct a facial image.

### 5.1 Shape Reconstruction

The shape of the novel frontal facial image is reconstructed based on the aligned feature points extracted from the source images. Two-view stereoscopy is introduced into the shape reconstruction, as shown in the Fig. 6 below.

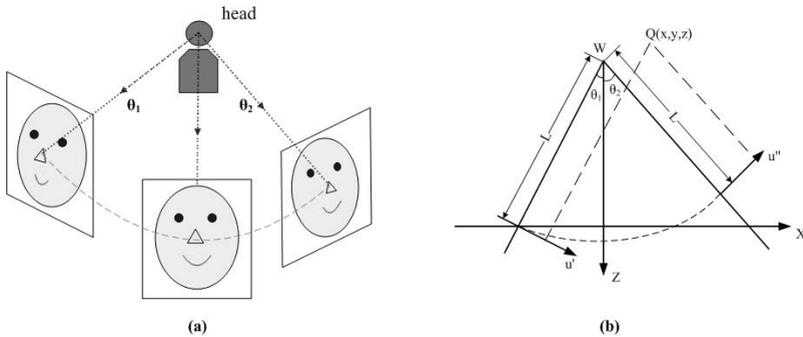


Fig. 6. Demonstration of the stereoscopy of shape reconstruction

For those feature points that are occluded in the source image with large rotation, we utilize a mean 3D model generated by 30 training subjects to reproduce their positions. We assume the shape of a human face is symmetric in horizontal direction. So we can generate positions of feature points in the occluded part from the viewable part in the same image with the help of this model. Although the model-based generated feature point positions are only approximations to their actual positions, the accuracy of these positions can be improved by a series of detailed iterative measures.

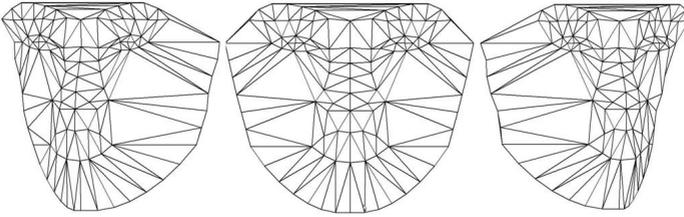


Fig. 7. A mean 3D model in our experiment

### 5.2 Texture Synthesis

For each source image and the reconstructed image, after feature points are positioned, triangularization is introduced to depart the face into multiple triangles. The triangularization follows the same principle and is realized by applying Delaunay triangularization [11] to a standard frontal shape. The standard frontal image can be a mean shape from a series of training images. Then the triangle based affine transform is used to span the source face images to fit the destination shape. Equation (6) describes this affine transform process. Here  $(x', y')$  is the corresponding coordinate in the destination image of a point  $(x, y)$  in the source image. Since the coordinates of vertexes of the corresponding triangles in the source and destination images are known, the affine transform in Equation (6) can be solved.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} O_x \\ O_y \end{bmatrix} \tag{6}$$

Fig.8 is an example of the texture synthesis. The left two are the source images with feature points extracted by ASM; the middle is the destination shape and its triangulation; the right one is the pure texture for the destination image.

For those source images with very small rotation, there is very little occluded region. So the whole texture of each source image is used to synthesize a frontal facial texture and then the mean of them is considered to be the texture for destination image. However, for those source images with large rotation, the situation will be different. For each source image, we use the non-occluded half to generate

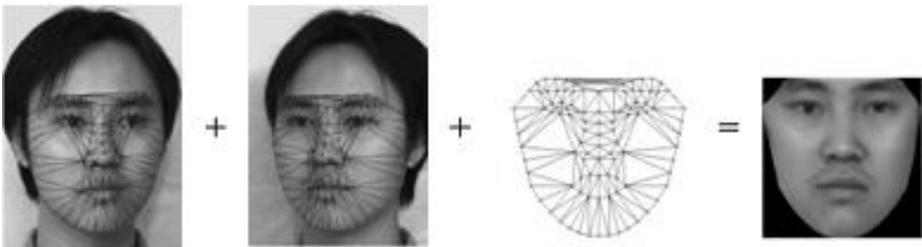


Fig. 8. Process of texture synthesis

respectively texture of half the destination image, and a smooth filter is applied to the boundary in the middle.

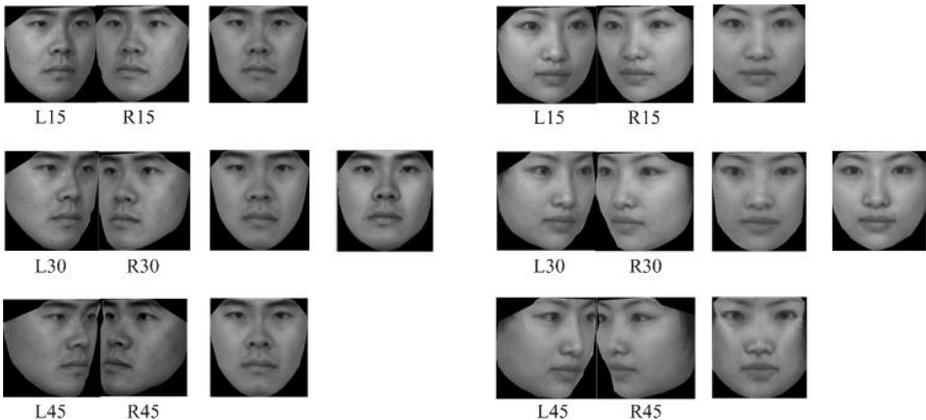
## 6 Experiment Results

In this section we describe our experiment results on TH face database [7]. The proposed synthesis method is used to generate frontal facial images based on different sets of facial images with different poses. Then the performance of face recognition based on the synthetic images is tested and compared with that based on the original face-rotated images, as shown in Table 1 on the next page.

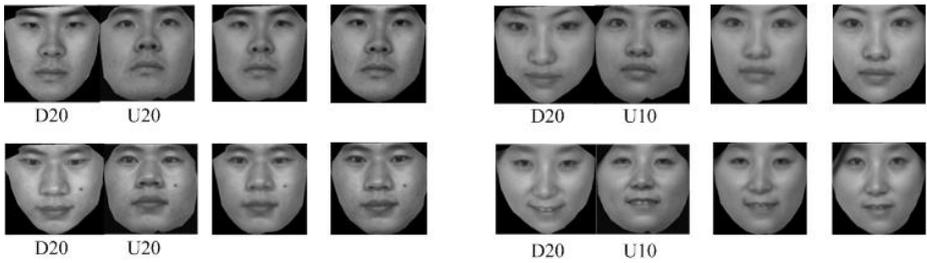
Experiments are based on an identification task. The task is completed by common Gabor-PCA algorithm, which is trained by using another 600 frontal images from TH face database. The recognition rate means the fraction of probes that have ranked first in the identification task. The gallery consists of images of 2000 people with one frontal image per person from TH face database. Probe sets include image sets of different facial poses and the synthetic face image sets. Each probe set contains 200 face images and would be used to test recognition performance respectively.

In our experiments, the mean time cost of the proposed method is tested. The one-time completion of the whole scheme is 0.98 second averagely. This indicates that this method can be used for real-time identification applications.

Fig. 9 and Fig. 10 show some examples of the facial synthetic results. Fig. 9 gives two examples of face synthesis based on images with left-right rotation angles (yawing angles). Fig. 10 gives three examples of face synthesis based on images with up-down rotation angles (pitching angles). For each person, the first two columns are original images with rotation; the third column is the synthetic frontal facial image generated from the two images in the same row by our method; the fourth row is the real frontal facial image for the same person which is presented here for comparison.



**Fig. 9.** Examples of face synthesis based on images with different yawing angles (left-right rotation angles)



**Fig. 10.** Examples of face synthesis based on images with different pitching angles (up-down rotation angles)

**Table 1.** Recognition performance of the original image sets and the synthetic image sets

Original Image Set	Recognition Rate	Original Image Set	Recognition Rate	Synthetic Image Set	Recognition Rate
L 15	71.4%	R 15	70.1%	L 15+R 15	<b>88.5%</b>
L 30	58.2%	R 30	57.8%	L 30+R 30	<b>76.5%</b>
L 45	25.0%	R 45	25.2%	L 45+R 45	<b>52.6%</b>
L 15	71.4%	R 30	57.8%	L 15+R 20	<b>85.6%</b>
L 15	71.4%	R 45	25.2%	L 15+R 45	<b>79.0%</b>
L 30	58.2%	R45	25.2%	L 30+R 45	<b>62.7%</b>
U 10	62.5%	D 10	62.8%	U 10+D 10	<b>79.9%</b>
U 20	45.0%	D 20	42.3%	U 20+D 20	<b>66.7%</b>
U 10	64.0%	D 20	42.3%	U 10+D 20	<b>70.9%</b>

Table 1 shows the recognition performance of different probe sets. In Fig. 9, Fig. 10 and in Table 1, “L xx” means the images in the set has a face pose with xx degree’s rotation to its left. Similarly, “R” means “Right”, U means “Up” and “D” means “Down”. The name “L xx + R yy” in the fifth column means that images in the set are synthesized through the corresponding images in “L xx” set and “R yy” set. From Table.1, we can see that the recognition rates in the last column are much higher than the other two in the same row, which indicates that the frontal face synthesis did help to improve the recognition performance significantly.

## 7 Conclusions

In this paper, we proposed a scheme for pose-variant face recognition. In order to overcome the difficulty brought by the non-frontal face images, a stereoscopic synthesis method is presented to generate a frontal face image based on two

pose-variant face images, which are captured at the same time. To ensure the accuracy of shape reconstruction, we introduced an eigenspace analysis and SVM classification combined method for facial pose estimation and an improved ASM method for facial feature point extraction and alignment. With more than one input images, the whole face texture is nearly totally covered so that the synthesized frontal face can retain the individual texture characteristics. Although a small quantity of unavoidable estimation and alignment errors may affect the final reconstruction accuracy, experiment results show that most of the information important for recognition has been retained and helps to improve the recognition performance. Without expensive time cost, this method is suitable for real-time identification applications.

## References

1. Phillips, P.J., Grother, P., Ross, J., Blackburn, D., Tabassi, E., Bone, M.: Face Recognition Vendor Test 2002: Evaluation Report. (March 2003)
2. Du, Y., Lin, X.: Multi-view face image synthesis using factorization models. In: International Workshop on Human-Computer Interaction (2004)
3. Vetter, T., Poggio, T.: Linear object classes and image synthesis from a single example image. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7), 733–742 (1997)
4. Cootes, T., Edwards, G., Taylor, C.: Active appearance models. *IEEE Trans. Pattern Analysis and Machine Intelligence* 23(6), 681–685 (2001)
5. Cootes, T., Taylor, C., Cooper, D., et al.: Active shape models ~ their training and application. *Computer Vision and Image Understanding* 61(1), 38–59 (1995)
6. Vetter, T.: Synthesis of novel views from a single face image. *International Journal of Computer Vision* 28(2), 103–116 (1998)
7. Li, C., Su, G., Meng, K., Zhou, J.: Technology Evaluations on TH-FACE Recognition System. In: Zhang, D., Jain, A.K. (eds.) *Advances in Biometrics*. LNCS, vol. 3832, pp. 589–597. Springer, Heidelberg (2005)
8. Vapnik, V.: *Statistical Learning Theory*. John Wiley & Sons, New York (1998)
9. Gu, H., Su, G., Du, C.: Feature Points Extraction from Faces. In: *Image and Vision Computing (IVCNZ'03)*, pp. 154–158 (2003)
10. Du, C., Su, G., Lin, X., Gu, H.: An Improved Multi-resolution Active Shape Model for Face Alignment. *Journal of Optoelectronics. Laser* 15(12), 706–710 (2004) (in Chinese)
11. Hassanpour, R., Atalay, V.: Delaunay Triangulation based 3D Human Face Modeling from Uncalibrated Images. In: *Int. C. Computer Vision and Pattern Recognition*, p. 75. IEEE Comput. Soc. Press, Los Alamitos (2004)