3D Face Reconstruction and Dynamic Feature Extraction for Pose-Invariant Face Recognition

Xiaohu Shao, Xi Zhou, Cheng Cheng
Automated Reasoning and Cognition Key Laboratory of Chongqing
Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences
Chongqing, China
{shaoxiaohu, zhouxi, chengcheng}@cigit.ac.cn

Tony X. Han
Electrical and Computer Engineering Department
University of Missouri
Columbia, USA
hantx@missouri.edu

***Abstract*—In this paper, we present a pose invariant face recognition framework leveraged on 3D face reconstruction and dynamic feature extraction. First, we synthesize the virtual frontal face from the probe face based on 3D face reconstruction. In the initialization of reconstruction, a tree-structured model is applied to detect landmark points from a 2D image and a hierarchical gaussianization (HG) based method is used for pose estimation. Second, in the recognition step, we present a dynamic feature extraction method to improve the recognizer, which measure the similarity between the synthesized the virtual face and the probe face. Recognition experiments are carried on the Multi-PIE, and CIGIT Face databases. The experimental results show that our system significantly improves the accuracy of face recognition, especially for the faces with extreme poses.**

Keywords-3D face reconstruction; face recognition; dynamic feature extraction

I. INTRODUCTION

Face recognition plays an important role in pattern recognition and computer vision applications. It remains a difficult problem due to the variations in pose, illumination and expression. More specifically, different poses of the same face have dramatically different appearances, causing fatal problems to most of current face recognition systems.

In order to solve the aforementioned problems, many approaches have been explored: The pose insensitive feature-based methods are widely used; they try to extract specific features which are invariant or insensitive to different poses. Wiskott et al [1] collapse face variance of pose and expression by extracting concise face descriptions in the form of image graphs. Gross et al [2] develop the theory of appearance-based face recognition from light-field, which leads directly to a pose-invariant face recognition algorithm that uses as many images of the face are available. Lai et al [3] use wavelet transform and multiple view images to determine the reference image representation.

Generating virtual frontal faces is another promising method for pose-invariant face recognition. Chai et al [4] use locally linear regression (LLR) to generate the virtual frontal view from a given non-frontal face image. Li et al [5] propose a method for cross-pose face recognition using a regression with a coupled bias-variance tradeoff. Berg [6] takes advantage of a reference set of faces to perform an "identity-preserving" alignment, warping the faces in a way that reduces differences due to pose and expression. Hu et al [7] reconstruct a 3D face model from a single frontal face image, and synthesize faces with different PIE to characterize face subspace. Wang [8] proposes a fully automatic, effective and efficient framework for 3D face reconstruction based on a single face image in arbitrary view. Asthana et al [9] build a 3D Face Pose Normalization system which improves the recognition accuracy of face variation up to ± 45 degrees in yaw and ± 30 degrees in pitch angles.

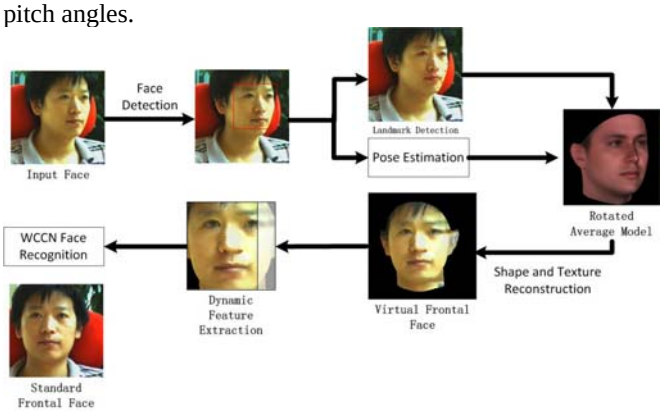

Figure 1. Framework of our pose-invariant face recognition system

Inspired by the above approaches, we present a pose-invariant face recognition framework leveraged on 3D face reconstruction and dynamic feature extraction. First, we synthesize the virtual frontal face from the probe face based on 3D face reconstruction. In the initialization step of face reconstruction, a tree-structured model is applied to detect landmark points from a 2D image and a hierarchical gaussianization (HG) based method [10, 11] is used for pose estimation. Second, in the recognition process, we present a dynamic feature extraction method to improve the Within-Class Covariance Normalization (WCCN) recognizer [12, 13], which measure the similarity between the synthesized the virtual face and the probe face. The framework of our pose-invariant face recognition system is shown in Figure 1.

The remainder of this paper is organized as follows. Section II introduces the initialization of 3D face reconstruction briefly. Reconstructing face shape and texture model is detailed in section III. Section IV describes the method of dynamic feature extraction according to face poses. Section V and VI provide the experimental results and conclusion.

II. Initialization of 3D Face Reconstruction

Landmark detection and Pose Estimation for a 2D face image are very important for the initialization of 3D Face Reconstruction. When the location of landmarks and pose of face are not correctly provided, it will make reconstruction strategy inevitably converging to bad local minima [7, 9].

Following the previous work [9], we use Haar-like and cascade AdaBoost to train a face detector [15], which is able to detect most of the faces with yaw angles from -45° to 45° . We apply Zhu's method [14] to locate key facial points. This method is based on a tree-structured model. It not only outperforms the state-of-the-art on laboratory databases, but also works well on wild images.

We use an image descriptor, hierarchical gaussianization (HG) [10, 11] for pose estimation in complex environment. We train our pose estimation model on a mixture dataset including images with yaw angles ranging from -45° to +45° with a stride of 5° .

III. 3D Face Reconstruction

Hu's method [7] only reconstructs face model frontal face image, but fails on non-frontal face. We make some improvements on their method, and build a pose-invariant system. These improvements are detailed as follows.

A. Shape and Texture Reconstruction

The geometry of a 3D face is denoted as a shape vector:

$$S = (x_1, y_1, z_1, ..., x_n, y_n, z_n)^T \in \Re^{3n},$$

where (x_i, y_i, z_i) is the 3D coordinate of the i-th vertice of total n vertices. A common assumption is that new shapes of 3D faces can be modeled as a linear combination of the average shape and the principal components [16, 17]. Then a new face shape S_{new} can be expressed as:

$$S_{new} = \overline{S} + \sum_{i=1}^{m} p_i s_i , \quad (1)$$

where $\overline{S}$ is the average shape, and $\vec{s} = (s_1, s_2, ..., s_m)^T \in \Re^m$ is the coefficients of the shape eigenvectors. $\vec{p} = (p_1, p_2, ..., p_m) \in \Re^{3n \times m}$ is the matrix of the first m principal components of the shape.

In the landmark detection, T landmarks are selected for a 2D face image. These key vertices are projected into 2D coordinates, and used for shape construction. This method are described briefly as follows [7]:

Let the $S_{2D} = (x_1, y_1, ..., x_n, y_n)^T \in \Re^{2n}$ be the set of X , Y coordinates of those landmarks on the 2D face. A new 2D shape $S_{2D,new}$ can be expressed as:

$$S_{2D,new} = \overline{S_{2D}} + \sum_{i=1}^{m} p_{2D,i} s_i , \quad (2)$$

where $\overrightarrow{p_{2D}} = (p_{2D,1}, p_{2D,2}, ..., p_{2D,m}) \in \Re^{2t \times m}$ is the X , Y coordinate of $\vec{p}$. The new shape model S_{new} can be reconstructed when geometry coefficient $\vec{s}$ is obtained. More details can be found in [7].

Because the aforementioned shape reconstruction method assumes that input image is a frontal face, there is no pre-processing step before projecting for different face poses. When a non-frontal face appears, $\overline{S_{2D}}$ should be the face projected from $\overline{S}$ with the corresponding pose. In order to get a correct $\overline{S_{2D}}$, We rotate the mean 3D model $\overline{S}$ using our pose estimation and project them into 2D coordinate space as follows:

$$PRS_{new} = PR(\overline{S} + \sum_{i=1}^{m} p_i s_i), \quad (3)$$

where P is the projection matrix, and R is the rotation matrix calculated from the pose estimation. Then a new equation can be obtained:

$$S^R_{2D,new} = \overline{S^R_{2D}} + \sum_{i=1}^{m} p^R_{2D,i} s_i . \quad (4)$$

After the shape reconstruction, we project the 2D image orthogonally to the 3D geometry to generate the face texture, and use the interpolate method to complete the texture reconstruction [7, 8].

B. Frontal Face Generation

After rotating the model S back by multiplying R^{-1} , we could get the frontal image reconstructed from the non-frontal image. The results of reconstruction are shown in Figure 2.

Figure 2. Frontal Face Generated from Multi-PIE and CIGIT Face databases: From top to bottom are initial images, frontal faces using 3D reconstruction and ground truths.

IV. DYNAMIC FEATURE EXTRACTION

Although our 3D reconstruction system can handle most of non-frontal images, it fails in a few cases where face images are captured with extreme illumination and pose variation. These factors decrease the performance of landmark and pose estimation, leading to a poor initialization for shape reconstruction.

We noticed that part of faces occluded in the 2D images sometimes fails in texture reconstruction, while other parts not occluded are precisely reconstructed. What we need to do is to decrease the importance of the mistaken parts while increasing the importance of the correct parts of the reconstructed frontal faces. So we present a dynamic feature extraction according to faces' pose to avoid the faultiness of texture reconstruction.

A. The Similarity of Features between Non-frontal and Frontal Faces

In order to test our idea, we first calculate feature distances between non-frontal and frontal faces. We call the initial feature static feature, and call the feature with weights dynamic feature. We divide a face image into $N \geq 2$ patches averagely in horizontal direction. For each patch feature vector $\mathbf{f}_i$ of every patch $\mathbf{P}_i$, we multiply a weight ω_i ($0 \leq \omega_i \leq 1$), which represents the importance of the feature vector. A demonstration of the dynamic feature extraction is shown in Figure 3.

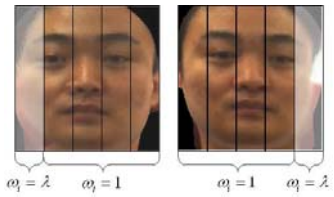

Figure 3. Dynamic feature extraction for faces with yaw angle of -30° and 30° when $N = 5$.

In this subsection, we select features of faces with 45° and 0° as the test database, and set sum of patches N to 10. When we decrease more weights of the mistaken parts, the similarity of features between faces with 45° and 0° should become larger. As mistaken parts of faces with 45° often appear in the right side, we set $\boldsymbol{\omega} = \{\omega_1, \omega_2, ..., \omega_N\}$ in the first test:

$$\omega_i = \begin{cases} 1, i < n_1 \\ \lambda_1, i \geq n_1 \end{cases}, \quad (5)$$

where $\lambda_1 \in [0,1)$, $n_1 \in [N/2, N]$. In the second test, we set:

$$\omega_i = \begin{cases} 1, i > n_2 \\ \lambda_2, i \leq n_2 \end{cases}, \quad (6)$$

where $\lambda_2 \in [0,1)$ $n_2 \in [1, N/2]$. We use the inner product to measure the similarity of feature vectors.

B. The Optimization of Weights of Dynamic Feature Extraction

The experiments above confirm our assumption, and we optimize n and λ for each pose of faces to reach the maximum face recognition rate.

During the training, we use equation (5) and (6) to obtain weights ω of features with positive and negative pose angles independently. We set $N = 10$, and $\lambda \in \{0, 0.5\}$. By trying all pairs of the two parameters on faces with each pose except frontal faces, we find the suitable weights $\boldsymbol{\omega}$ for each pose, which make the classification rate maximum. When calculating the distance between a probe image and the frontal image, the weights $\boldsymbol{\omega}$ of the latter should be set as equal to that of the former.

V. EXPERIMENTS

A. Database

Multi-PIE Database: Multi-PIE [18] has been widely used in the research of face recognition. In our experiment, we use the faces with pitch angle from [-45°, 45°], neutral expression and five illumination cases for training and testing. There are 7, 377 faces detected from 8, 716 images by using our face detector.

CIGIT Face Database: CIGIT Face database contains around 800, 000 images of more than 400 people under different viewpoints, occlusions, expression and illuminations. These images are captured by 91 cameras, with pitch angle ranging from 0°to 30°, yaw angle ranging from -90°to 90°. In our experiment, we select the face images with pitch angle from [0°, 10°], yaw angle ranging from [-45°, 45°] and normal expression for training and testing. There are 12, 507 faces successfully detected from 13, 017 images.

We select faces of 20 people as the training set, and use others for test. In the test, we use the frontal image with normal illumination as the gallery image, and the others are used as probe images.

B. Feature Extraction and Classification

TABLE I. COMPARISON OF MEAN RECOGNITION RATES FOR MULTI-PIE AND CIGIT FACE DATABASES.

Method	Mean rate (%)	
	Multi-PIE	CIGIT Face
Baseline	90.6	93.1
3D	92.7	94.8
3D+Dynamic Features	94.8	97.0

We evaluate our system on the two databases described above. Hierarchical gaussianization (HG) and WCCN [12, 13] are chosen for feature extraction and face recognition. During the dynamic feature extraction, we calculate the HG

feature according to the face angle. The proposed algorithm is systematically evaluated by comparing it with the algorithm that does not use the 3D reconstruction. The results are shown in Table 1 and Figure 4.

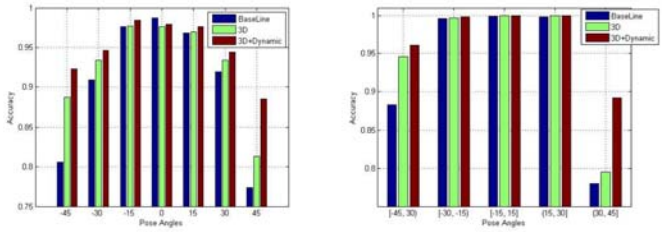

Figure 4. Comparison of recognition rates of each pose for Multi-PIE (Left) and CIGIT Face (Right) databases. Baseline: HG + WCCN method, 3D: Using 3D reconstruction based on baseline, Dynamic: Using dynamic feature extraction based on 3D method.

From the results, we conclude that our system performs better than the algorithms that do not use the 3D reconstruction, especially for the faces with pose angle larger than 30°. Furthermore, by applying dynamic feature extraction, our system achieves a higher recognition rate. Considering HG feature and WCCN are such efficient for the recognition of frontal faces (rate≈100%), we only get a limited improvement on faces with small yaw angles.

VI. Conclusion and Future Work

In this paper, we propose an efficient and automatic system of face recognition robust to pose variation. Compared with previous work, our work has the following contributions: 1) By applying robust algorithms of accurate landmark detection and pose estimation, we synthesize the virtual frontal face from the probe face based on 3D face reconstruction. 2) To reduce the noisy effect induced by the texture reconstruction, we present a dynamic feature extraction method to help the recognition system to measure the similarity between synthesized the virtual face and the probe face. Based on these work, our face recognition system can handle faces of different poses in more complicated environment.

In the future, we will improve methods of landmark detection and pose estimation, and also focus on texture reconstruction in large poses, mirror method and Poisson Editing [19] in our future work.

Acknowledgment

This work was supported by the Project from Committee on Science and Technology of Chongqing 2011 (No. cstc2011ggC40009) and 2012 (No. cstc2012gg-sfgc0079).

References

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

[2] R. Gross, I. Matthews, and S. Baker, "Appearance-based face recognition and light-fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 449-465, 2004.

[3] J. H. Lai, P. C. Yuen, and G. C. Feng, "Face recognition using holistic Fourier invariant features," *Pattern Recognition*, vol. 34, pp. 95-109, 2001.

[4] X. Chai, S. Shan, X. Chen, and W. Gao, "Locally linear regression for pose-invariant face recognition," *IEEE Transactions on Image Processing*, vol. 16, pp. 1716-1725, 2007.

[5] A. Li, S. Shan, and W. Gao, "Coupled Bias-Variance Tradeoff for Cross-Pose Face Recognition," *IEEE Transactions on Image Processing*, vol. 21, pp. 305-315, 2012.

[6] T. Berg and P. N. Belhumeur, "Tom-vs-Pete Classifiers and Identity-Preserving Alignment for Face Verification," *Proceedings of the British Machine Vision Conference (BMVC)*, 2012.

[7] Y. Hu, D. Jiang, S. Yan, and L. Zhang, "Automatic 3D reconstruction for face recognition," *In Proceedings of International Conference on Automatic Face and Gesture Recognition*, pp. 843-848, 2004.

[8] C. Wang, S. Yan, H. Li, H. Zhang, and M. Li, "Automatic, effective, and efficient 3D face reconstruction from arbitrary view image," *Advances in Multimedia Information Processing-PCM*, pp. 553-560, 2005.

[9] A. Asthana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. Rohith, "Fully automatic pose-invariant face recognition via 3D pose normalization," *In Proceedings of International Conference on Computer Vision (ICCV)*, pp. 937-944, 2011.

[10] S. Yan, X. Zhou, M. Liu, M. Hasegawa-Johnson, and T. S. Huang, "Regression from patch-kernel," *In Proceedings of International Conference on Computer Vision and Pattern Recognition（CVPR）*, pp. 1-8, 2008.

[11] X. Zhou, N. Cui, Z. Li, F. Liang, and T. S. Huang, "Hierarchical gaussianization for image classification," *In Proceedings of International Conference on Computer Vision*, pp. 1971-1977, 2009.

[12] X. Zhou, K. Yu, T. Zhang, and T. S. Huang, "Image classification using super-vector coding of local image descriptors," *In Proceedings of International Conference on Computer Vision*: Springer, pp. 141-154, 2010.

[13] X. Zhou, J. Navrdtit, J. W. Pelecanos, G. N. Ramaswamy, and T. S. Huang, "Intersession variability compensation for language detection," *In Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4157-4160, 2008.

[14] X. Zhu and D. Ramnan, "Face Detection, Pose Estimation, and Landmark Localization in the Wild," *In Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR)* 2012.

[15] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *International Conference on Computer Vision and Pattern Recognition*, pp. I-511-I-518, 2001.

[16] I.T.Jollife, "Principal Component Analysis," in *Springer-Verlag* New York, 1986.

[17] L.Sirovich and M.Kirby, "Low-dimensinal procedure for the characterization of human faces," *Jounal of the Optical Society of America A*, vol. 4, pp. 519-554, 1987.

[18] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *In Proceedings of International Conference on Automatic Face & Gesture Recognition*, pp. 1-8, 2008.

[19] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Transactions on Graphics (TOG)*, vol. 22, pp. 313-318, 2003

.