Face Recognition Across Poses Using A Single 3D Reference Model

Gee-Sern Hsu*, Hsiao-Chia Peng
National Taiwan University of Science and Technology
No. 43, Sec.4, Keelung Rd., Taipei, 106, Taiwan
*jison@mail.ntust.edu.tw

Abstract

Approaches for cross-pose face recognition can be split into 2D image based and 3D model based. Many 2D based methods are reported with promising performance but can only work for poses same as those in the training set. Although 3D based methods can handle arbitrary poses, only a small number of approaches are available. Extended from a latest face reconstruction method using a single 3D reference model, this study focuses on using the reconstructed 3D face for recognition. The reconstructed 3D face allows the generation of multi-pose samples for recognition. The recognition performance varies with poses, the closer the pose to the frontal, the better the performance attained. Several ways to improve the performance are attempted, including different numbers of fiducial points for alignment, multiple reference models considered in the reconstruction phase, and both frontal and profile poses available in the gallery. These attempts make this approach competitive to the state-of-the-art methods.

1. Introduction

The approaches for face recognition across poses can be generally split into two categories, one is based on 2D images [10, 5, 3], and the other based on 3D models [2, 6, 13]. More advancements have been made on the former which appear to outnumber the latter considerably [12]. However, most 2D approaches suffer from the limitation that they only work for poses same as those in the training set. Because 3D information is considered fundamental for recognition across arbitrary poses, more 3D based methods are yet to be developed.

In the 2D image based methods, the Eigen Light-Feilds (ELF) [5] assumes that the pixel intensities correspond to the radiance of light emitted from the face along certain rays in space, and estimates the basis set of the radiance values at each pose using samples of the same pose in the training set. The eigen light-field is defined on this basis set, which allows the gallery and probe faces represented in ELF coefficients, and recognition can be performed by matching these coefficients. It is an effective method dealing with poses, but suffers from the requirements that the probe images must align with light-field vectors. The Tied Factor Analysis (TFA) is proposed in [10] which decomposes a face into a latent variable (or factor) in the identify space, a pose-dependent mapping from identity to observation, a pose-dependent mean and a noise. Since the pose-dependent mapping and mean are independent of the subject, they can be obtained from a training set. Given a non-frontal face with a known pose, its corresponding frontal pose can be estimated using the learned frontal pose mapping and mean, and then matched against those in the gallery. This method requires manual annotation of local features for pose-specific alignment. The performance degrades, sometimes significantly, when local features fail to be accurately localized. A stereo matching approach with epipolar geometry is applied in [3] to evaluate the similarity between two faces of different poses. Given three or four corresponding feature points on both faces, two sets of scanlines with epipolar constraints can be determined, and a stereo matching cost can be computed and optimized to reveal how well the two faces match to each other. The regression-based method in [8] estimates the coefficients of linear combinations of 2D faces in the training set for approximating the linear subspaces for 3D face. To reduce the high variances in the estimated coefficients, the method exploits the regressors with local Gabor features for bias-variance balancing. Although these 2D-based methods report performances better than many 3D-based ones, all of them and many other 2D methods suffer from the limitation that they only work for poses available in the training set, making them ineffective in some practical applications.

In 3D model based approaches, the morphable model [2] uses the prior knowledge, including the 3D face shapes and textures, collected from hundreds of 3D facial scans to build a 3D model for a given 2D image. Although considered as an effective solution for recognition under pose and illumination variations, it is expensive in storage and computation because of the storage of the hundreds of 3D scans

and the search for the correspondences to the reference model. A similar approach but modified with automatic feature localization is given in [6], which reports a satisfactory performance for poses less than 45°, but degrades significantly for large poses. Because the conventional PCA is used after synthesizing the views to match against the probe, we consider this a baseline for 3D methods in our performance evaluation. The Heterogeneous Specular and Diffuse (HSD) [13], one of the latest approaches, allows both specular and diffuse reflectance coefficients to vary spatially to better accommodate surface properties of real faces. A few face images under different lighting conditions are needed to estimate the 3-D shape and surface reflectivity using stochastic optimization. The resultant personalized 3-D face model is used to render novel gallery views of different poses for cross-pose recognition.

Our method extends the latest work on 3D face reconstruction proposed by Kemelmacher-Shlizerman and Basri [7] to tackle cross-pose recognition. It is 3D model-based in nature, but different from [2, 6, 13] and others in that it exploits a single 3D reference model and recovers the 3D shape of a 2D face image in the gallery. It consists of the following steps: (1) 3D reconstruction based on the reference model, (2) model-based synthesis of novel poses, and (3) pose-oriented feature extraction and matching. The proposed method is a low-cost alternative to many 3D model-based approaches that require a large number of 3D scans.

The rest of the paper is organized as follows: The preparation of the 3D reference model and the model-based reconstruction are presented in Sec. 2. Although the reconstruction part is mostly based on [7], our interpretation from a different viewpoint can be easier for implementation. The recognition using the Uniform Local Binary Pattern (ULBP) features extracted from the model-based synthesized novel views is described in Sec. 3. Sec. 4 presents experimental results with three ways for performance improvement, including additional fiducial points for pose alignment, an additional 3D reference model considered in the reconstruction phase, and the addition of the profile pose to the gallery set. A performance comparison with the aforementioned 2D-based and 3D-based methods is also given in Sec. 4, followed by a conclusion in Sec. 5.

2. Reconstruction Using A Single Reference Model

We reformulate the problem as a constrained minimization so that the well-known scheme with Lagrange multipliers can be applied. We also make some minor modifications to the original algorithm in [7], making our reconstruction somewhat different from theirs, although the results are similar. Nevertheless, the investigations that we have added to the reconstruction phase include the rendering of a smooth surface from the noisy data of a 3D face scan for the reference model, and the study on different numbers of fiducial points used for the alignment between the 2D image and 3D reference model. The former is presented in Sec.2.1, and the latter in Sec.4 with experimental results.

2.1. Reference Model Surface Rendering and Parameter Estimation

This step is not described explicitly in [7], but considered an essential part of the reconstruction when making a raw 3D face scan good as the reference model. Instead of using samples from the USF database as the reference models as in [7], we select samples from the FRGC database [9] because of its popularity. Each FRGC 3D face scan consists of a range image and a texture image that we can use to estimate the surface normal $\vec{n}_r(x, y)$ and albedo $\rho_r(x, y)$, which are required for the reconstruction of other faces.

We applied the Moving Least Squares (MLS) [1] to smooth $z_{r,0}$, the raw depth data of the reference model, so that the measurement noise in $z_{r,0}$ can be removed and the smoothed surface z_r can best approximate $z_{r,0}$. Given a subset of $z_{r,0}$ in the form of point clouds, denoted as $\mathbf{P}_k=\{\vec{p}_i\}_{i=1,\cdots,N_k}$, the goal is to determine a novel set of points, $\mathbf{R}_k=\{\vec{r}_i\}_{i=1,\cdots,N_k}$, on a low-order polynomial that minimizing the distance between $\mathbf{P}_k$ and $\mathbf{R}_k$. The smoothed surface z_r can then be obtained from $\{\mathbf{R}_k\}_{\forall k}$. Modified from the MLS reported in [1] for better efficiency, our method is composed of the following step,

1. Use $\mathbf{P}_k$ to determine a local plane H_0 with origin $\vec{q}_0$ and normal $\vec{n}_0$ so that the following weighted sum can be computed,

$$\sum_{i=1}^{N_k} \left(u_0(x_i, y_i) - \mu_{i,0}\right)^2 \phi\left(\|\vec{p}_i - \vec{q}_0\|\right) \qquad (1)$$

   where $u_0(x_i, y_i)$ is the distance of $\vec{r}_i$ to H_0 with the location of its projection onto H_0 given by (x_i, y_i); $\mu_{i,0}$ is the distance of $\vec{p}_i$ to H_0, i.e., $\mu_{i,0} = \vec{n}_0 \cdot (\vec{p}_i - \vec{q}_0)$; and $\phi(\cdot)$ is a Gaussian function so that the points closer to $\vec{q}_0$ are weighted more. Assuming that $\mathbf{R}_k$ are described by a low-order polynomial in terms of the coordinates (x_i, y_i) on H_0, i.e., $\vec{r}_i = f(x_i, y_i|\Lambda_0)$ and $u(x_i, y_i) = \vec{n}_0 \cdot (f(x_i, y_i|\Lambda_0) - \vec{q}_0)$, where $f(x_i, y_i|\Lambda_0)$ is a polynomial surface with parameter Λ_0 that defines the local geometry of $\mathbf{R}_k$.

2. Because H_0 can be uniquely defined given $\vec{q}_0$ and $\vec{n}_0$, one can change them to $\vec{q}_1$ and $\vec{n}_1$ and obtain a novel plane H_1. Given that the order of the polynomial $f(x_i, y_i|\Lambda_0)$ is fixed (so that the number of parameters of $f(x_i, y_i|\Lambda_0)$ is fixed), a parameter estimation problem can be defined as the minimization of the weighted sum as:

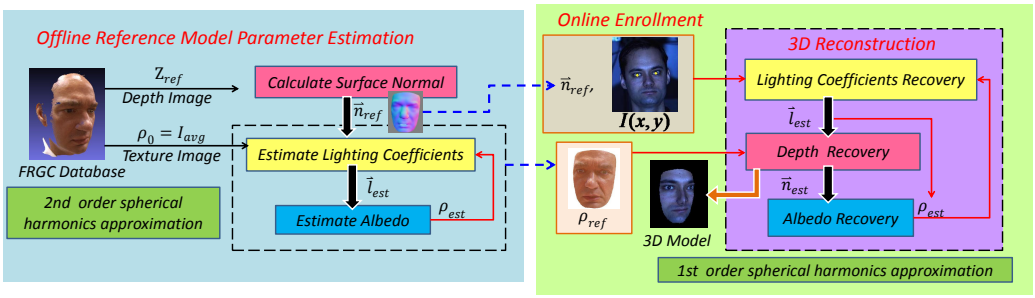

Figure 1. Work flow of the reference model parameter estimation and face model reconstruction.

$$\Lambda_k^*, \vec{n}_k^*, \vec{q}_k^* = \underset{\Lambda,\vec{n},\vec{q}}{\operatorname{argmin}} \sum_{i=1}^{N_k} \left(u(x_i, y_i) - \mu_i\right)^2 \phi\left(\|\vec{p}_i - \vec{q}\|\right) \tag{2}$$

The above can be repeated on other subsets $\{\mathbf{P}_k\}_{\forall k}$ for estimating $\{\Lambda_k, \vec{n}_k, \vec{q}_k\}_{\forall k}$ and $\{\mathbf{R}_k\}_{\forall k}$. A key issue in this scheme is the initial estimates of $\vec{n}_0$ and $\vec{q}_0$. A few possible ways are given in [1]; however, from our experiments we found that the minimum principal component extracted from $\mathbf{P}_k$ offers a good estimate of $\vec{n}_0$ and the centroid of $\mathbf{P}_k$ can be appropriate as $\vec{q}_0$. To extract the principal components, one needs to solve the eigenvectors of the covariance C_k,

$$C_k = \frac{1}{k} \sum_{i=1}^{k} (\vec{p}_i - \vec{q}_0) \cdot (\vec{p}_i - \vec{q}_0)^T \tag{3}$$

where $\vec{q}_0$ is the centroid of $\mathbf{P}_k$, and considered as the origin of the initial plane H_0. $\vec{n}_0$, the normal vector of H_0, is given by the eigenvector of C_k associated with the lowest eigenvalue. Following the above approach, the surface normal $\vec{n}_r$ can be obtained from the estimated polynomials $f(x_i, y_i|\Lambda_k)$. Given $\vec{n}_r$ and the associated 2D image I_r, ρ_r can be estimated using the method presented in the next section with some simplification, as described at the end of Sec.2.2.

2.2. Irradiance Evaluation using Constrained Minimization

The goal in this section is to estimate the 3D shape model of any given 2D face image $I(x, y)$ using the depth $z_r(x, y)$, surface normal $\vec{n}_r(x, y)$ and albedo $\rho_r(x, y)$ of the reference model. Assuming that the face surface is Lambertian, $I(x, y)$ can be decomposed as

$$I(x, y) = \rho(x, y)\vec{h}(x, y) \cdot \vec{n}(x, y) = \rho(x, y)R(x, y) \tag{4}$$

where $\rho(x, y)$ is the surface albedo at the point (x, y), $\vec{h}(x, y) \in R^3$ is the lighting cast on (x, y) with intensity on each of the three directions, $\vec{n}(x, y)$ is the face surface normal at (x, y), and the reflectance $R(x, y) = \vec{h}(x, y) \cdot \vec{n}(x, y)$.

For simplicity of notation, the coordinates (x, y) is dropped in the rest of the paper, and $\vec{n}(x, y)$, for example, is written as $\vec{n}$. With a few assumptions [7], the reflectance can be approximated using spherical harmonics,

$$R(x, y) \approx \vec{l} \cdot \vec{Y}(\vec{n}) \tag{5}$$

where $\vec{l}$ is the lighting coefficient vector and $\vec{Y}(\vec{n})$ is the spherical harmonic vector, which, in the second order approximation, takes the following form:

$$\begin{aligned} \vec{Y}(\vec{n}) \quad = \quad & [c_0, c_1 n_x, c_1 n_y, c_1 n_z, c_2 n_x n_y, c_2 n_x n_z, c_2 n_y n_z, \\ & c_2(n_x^2 - n_y^2)/2, c_2(3n_z^2 - 1)/2\sqrt{3}]^T \end{aligned} \tag{6}$$

where $c_0 = 1/\sqrt{4\pi}$, $c_1 = \sqrt{3}/\sqrt{4\pi}$, $c_2 = 3\sqrt{5}/\sqrt{12\pi}$.

The difference between (4) and (6) is that the lighting intensity and direction are all merged into $\vec{h}$ in (4), separated from $\vec{n}$, but in (6) they are split into the lighting vector $\vec{l}$ and the spherical harmonics $\vec{Y}(\vec{n})$, which is solely dependent on the components of $\vec{n}$, namely n_x, n_y and n_z.

The core problem can now be formulated as the minimization of $||I - \rho\vec{l} \cdot \vec{Y}(\vec{n})||$ over ρ, $\vec{l}$ and $\vec{n}$. The solution in [7] uses the depth z_r, the surface normal $\vec{n}_r$ and the albedo ρ_r of the reference model for initialization, making the problem solvable by regularization. Because of a better computational efficiency, we choose DoG (Difference of Gaussian) instead of LoG (Laplacian of Gaussian) adopted in [7] in the minimization.

$$\min_{\vec{l},\vec{z},\rho} \int (I - \rho\vec{l} \cdot \vec{Y}(\vec{n}))^2 + \lambda_1 (D_g * d_z)^2 + \lambda_2 (D_g * d_\rho)^2 dxdy \tag{7}$$

where $d_z = z(x, y) - z_r(x, y)$, $d_\rho = \rho(x, y) - \rho_r(x, y)$, and D_g* denotes the convolution with the DoG; λ_1 and λ_2 are constants. Although this is not described explicitly in [7], the formulation in (7) can be better interpreted as the minimization of $||I - \rho\vec{l} \cdot \vec{Y}(\vec{n})||$ subject to the constraints $D_g * d_z \approx 0$ and $D_g * d_\rho \approx 0$. Such a formulation allows the interpretation of λ_1 and λ_2 as the Lagrange multipliers. Assuming that I is aligned to the reference model, the reconstruction tackles the minimization in (7) by first solving for

the spherical harmonic coefficients $\vec{l}$ using the references z_r and ρ_r, then the depth $z(x, y)$, and then the albedo $\rho(x, y)$.

The alignment between I and the reference model needs corresponding fiducial points on both I and the reference model. We applied the method in [4] for automatic detection of facial features, and adjusted the results manually in case the method failed to perform ideally. Given a set of fiducial points that splits I and the reference face into corresponding local regions, perspective and affine transforms are then applied to fit each local region of the reference model to the corresponding region in I.

The minimization (7) is also used for computing ρ_r given I_r and $\vec{n}_r$. In such a case, there are no constraint terms in (7), and one can use the average of 2D faces in the gallery as the initial guess of the albedo, $\rho_r^{(0)}$, to solve the lighting coefficients $\vec{l}^{(0)}$ and search for the desired ρ_r iteratively. The overall reference model parameter estimation and gallery face reconstruction are summarized in the workflow in Fig. 1.

3. Recognition Across Pose

3.1. Generation of Model-based Training Images

We assume a common scenario that the gallery has one frontal face image per subject for enrollment, and the probe set contains face images of other poses for recognition. A couple issues must be solved for this scenario: the generation of images good for training from the reconstructed 3D face, and the estimate of the pose of a given probe so that its matching to the gallery can be fast. To constrain the scope of this paper from covering facial feature localization, which can be solved by many algorithms, e.g., [4], we assume that the fiducial points on a probe can be available using these algorithms or manual annotation.

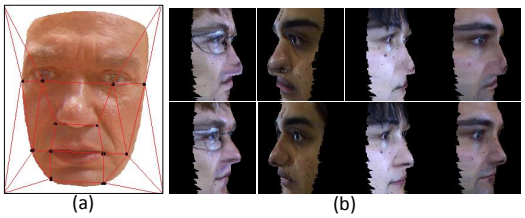

Figure 2. (a) 12 fiducial points for piece-wise warping. (b) Reconstruction with alignment using 3 fiducial points in the upper row, and 12 fiducial points in the bottom row.

Each frontal face image in the gallery is taken as the $I(x, y)$ in (7) for making its corresponding 3D face. The alignment between $I(x, y)$ and the reference model is performed using a set of fiducial points. Our experiments reveal that the fiducial-points-based alignment makes a strong impact on the reconstruction and recognition performance. Fig. 2 shows the reconstruction using 3 and 12 fiducial points. This, however, does not imply that more fiducial points always lead to better reconstruction. This issue is discussed along with experimental results in Sec. 4.

Following the approach presented in Sec. 2, one can obtain a 3D reconstructed face for each gallery image. The surface smoothing and rendering in Sec.2.1 is performed on each reconstructed face to obtain the finalized surface for each face. A weakly perspective transformation with a rotation matrix R_s and a translation vector t_s specified for pose $\mathbf{p}_s$ is then applied on the 3D facial surface to render its 2D projection on the image plane as the training image with pose $\mathbf{p}_s$.

3.2. Pose-Oriented Recognition

Although one can generate training images of arbitrary poses using the above approach, we consider it a better option to generate pose-oriented clusters of training images. Take the pose subset in the CMU PIE database [11] as an example, which is used in our experiments for performance evaluation. The pose subset offers 13 poses in total, 9 taken from horizontal views with yaw angle roughly $22.5°$ apart (so the central one corresponding to the frontal), 2 taken from the vertical views with pitch angle $22.5°$ up and down, and the rest 2 taken from *surveillance* views with yaw angle $67.5° = 3 \times 22.5°$ to both side and pitch angle $22.5°$ down. When generating the training set, each of these poses is considered as the center of a pose-oriented cluster, and four neighboring poses are synthesized and added to the cluster, including $10°$ up and down and $10°$ to the left and to the right. Instead of using the PIE original pose tags, such as c02, c37, ..., we use the approximated pose angle with an alphabet in the front to denote its direction. For example, R67.5° refers to 67.5° to the right, U22.5° is 22.5° upward and D22.5° is 22.5° downward. All synthesized face images are normalized in size to either the distance between the eyes and mouth when the poses are primarily caused by horizontal rotations, or to the distance between both eyes when the poses are caused by vertical rotations. The ULBP (Uniform Local Binary Pattern) features are extracted from each training image, and stored in the associated pose cluster for each subject. Compared with other forms of LBP, the ULBP gave the most consistent result in our experiments. When extracting the features, each 128×128 face image is split into 4×4 blocks, and the 59 dimensional ULBP vector extracted[1] from each block is cascaded into a $4 \times 4 \times 59 = 944$ dimensional feature vector. Following

[1]The ULBP is extracted from each pixel with associated histogram obtained as the feature vector from each 32×32 block.

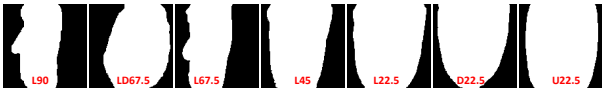

Figure 3. Partial masks of different poses.

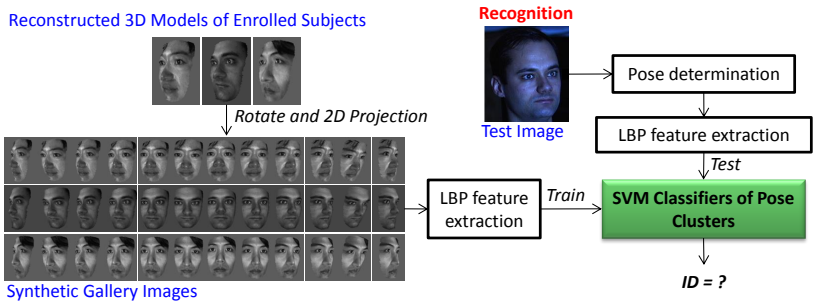

Figure 4. Workflow of cross-pose recognition across poses given the reconstructed 3D gallery face.

the above procedure, each person in the gallery can have 13 pose clusters, and each cluster has five synthesized images with ULBP feature extracted.

Given a probe with fiducial points available for size normalization and pose matching to the pose cluster in the gallery set, its region of interest is first obtained when imposed with a pose-oriented mask and then the ULBP feature is extracted. The pose-oriented masks, obtained from the averages of the training face images from the reconstructed faces, aim at blocking out non-facial region in the probe image. In our experimental setup with PIE, the frontal pose is used as the gallery and the rest 12 poses as the probe set. Considering the symmetry of the pose-oriented masks between 5 pose pairs left to right, 7 pose-oriented masks are shown in Fig. 3. We use SVM with linear kernel as the classifier as other kernels have not revealed better performance in our experiments. The overall workflow is shown in Fig.4.

4. Experiments

All experiments were run on a Linux PC with 2.6GHz and 4G DDR3. We used OpenCV library for image processing, CLAPACK (http://www.netlib.org/clapack) for solving optimization and Freeglut (http://freeglut.sourceforge.net) for handling 3D models of different poses. The Point Cloud Library (http://pointclouds.org) was used for preprocessing on both the reference and reconstructed models. The reference models were taken from the FRGC database and normalized to 250×300 in size. The face images in the training and testing sets were all scaled to 128×128. The performance was evaluated on the PIE pose subset, which has 68 subjects and 13 poses. The frontal pose of each subject was used in the gallery for enrollment and the rest poses in the probe for testing. This protocol is common for 3D-based methods. Most 2D-based methods need a "pose training set", which is often composed of all poses of half of the subjects, i.e., 34 subjects, for learning the relationship between the 13 poses. The frontal of the other 34 subjects are used as the gallery and the rest poses used as the testing set.

Because reconstruction takes most of the processing time, and the large the given image I, the longer the reconstruction takes. A few scales were tested, and although large scales generally led to better reconstruction and recognition performance, the scale factor 0.3 was selected for a balance between processing time and performance. Three issues were studied: additional fiducial points for local correspondences and pose alignment, an additional 3D reference model considered in the reconstruction phase, and the addition of profile pose to the gallery set.

Additional Fiducial Points for Pose Alignment

Different numbers of fiducial points yield reconstructed faces with different details. Four cases with 3, 6, 12, 15 and 23 fiducial points were considered. The fiducial points were used to split the face into local regions. Perspective transform and affine transform were then applied to fit a gallery image to the reference model in one region after another. A case with 12 fiducial points is shown in Fig. 2. We found that the reconstruction results vary with not just the number of fiducial points, but also the locations of the fiducial points. The fiducial points at eye regions were better located right below the eyes, rather than on the center of the eyes, as shown in Fig. 2. Although more fiducial points led to better reconstruction, we found that this was only valid for cases with less than 12 fiducial points. When the number exceeded 12, the performance barely improved, and it even slightly degraded when using 23 fiducial points.

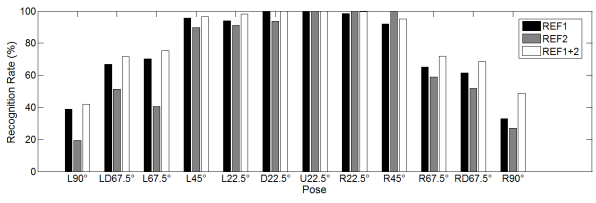

Figure 5. Performance comparison of single and double reference models.

Additional Reference Model

The default reference model was arbitrarily selected, as shown in the previous figures. We selected an additional one with different gender and age. Following the same ap-

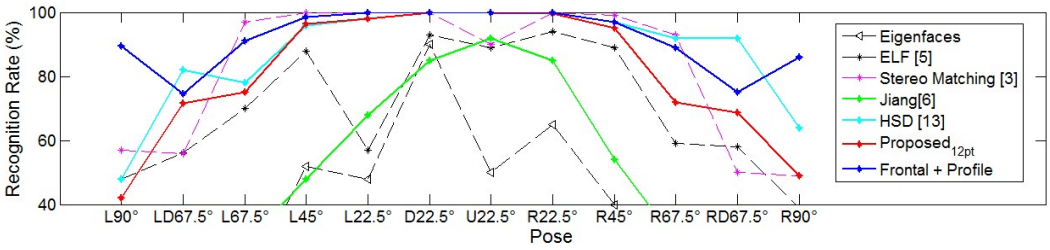

Figure 6. Recognition comparison on PIE database with ambient light on.

proach, each gallery image had two reconstructed models, generating an additional set of pose clusters for training. The comparison between these two cases is shown in Fig. 5, where the case with both REF1 and REF2 models outperforms the case with either one alone. Because different reference models allow different details generated on the reconstructed face given the same 2D image, each gallery face with multiple reconstructed faces would improve the overall recognition performance, especially for large poses.

Addition of Profile Pose in the Gallery

A common scenario in forensic and law enforcement applications considers both frontal and profile poses available in the gallery. This scenario was included in our experiments. We compared the proposed methods with the aforementioned settings to several state-of-the-art approaches on the PIE pose subset. Two reference models were used, one frontal pose with 12 fiducial points and the other of profile pose with 9 fiducial points. The recognition performance is shown in Fig. 6, together with the case with only one frontal pose in the gallery with 12 fiducial points.

Considering the poses less than $67.5°$, the best are stereo matching [3], HSD [13] and the proposed with 12 fiducial points and with both frontal and profile poses in the gallery. Although the performance of the proposed with 12 fiducial points drops at $90°$, the case with both frontal and profile poses available performs exceptionally well, giving an indication on the further improvement to our method. Because the 3D reference model can also offer fiducial points on the profile pose, which has not been considered yet in our study, it is believed that the proposed method can be further advanced taking such an advantage.

5. Conclusion

3D-based approaches for cross-pose recognition deserve special attention as 2D-based ones are mostly limited to the poses same as those in the training set. This work extends a latest work in [7] on 3D face reconstruction to recognition, and studies the impacts made by multiple reference models, fiducial points for alignment and others. The smoothed surface rendering, which is an important part for reconstruction but missing in [7], has been elaborated. Experiments on the PIE database show that this method can be competitive to the state of the art with multiple reference models considered in the reconstruction phase, an appropriate set of fiducial points selected for pose alignment, and, if allowed, the addition of profile pose to the gallery.

References

[1] M. Alexa, J. Behr, D. Cohen-Or, S. Fleishman, D. Levin, and C. T. Silva. Computing and rendering point set surfaces. *IEEE Trans. Vis. Comput. Graph.*, 9(1):3–15, 2003.

[2] V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *TPAMI*, 25(9):1063–1074, Sep. 2003.

[3] C. D. Castillo and D. W. Jacobs. Using stereo matching for 2-D face recognition across pose. In *CVPR*, pages 1–8, 2007.

[4] L. Ding and A. M. Martínez. Features versus context: An approach for precise and detailed detection and delineation of faces and facial features. *TPAMI*, 32(11):2022–2038, 2010.

[5] R. Gross, I. Matthews, and S. Baker. Appearance-based face recognition and light fields. *TPAMI*, 26:449–465, 2004.

[6] D. Jiang, Y. Hu, S. Yan, L. Zhang, H. Zhang, and W. Gao. Efficient 3D reconstruction for face recognition. *Pattern Recognition*, 38:787–798, June 2005.

[7] I. Kemelmacher-Shlizerman and R. Basri. 3D face reconstruction from a single image using a single reference face shape. *TPAMI*, 33(2):394–405, Feb. 2011.

[8] A. Li, S. Shan, and W. Gao. Coupled bias-variance tradeoff for cross-pose face recognition. *IEEE Transactions on Image Processing*, 21(1):305–315, 2012.

[9] P. Phillips, P. J.Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *CVPR*, volume 1, pages 947–954, 2005.

[10] S. J. Prince, J. H. Elder, J. Warrell, and F. M. Felisberti. Tied factor analysis for face recognition across large pose differences. *TPAMI*, 30:970–984, June 2008.

[11] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression (PIE) database. In *Proc. IEEE Conf. Automatic Face and Gesture Recognition*, pages 46–51, 2002.

[12] X. Zhang and Y. Gao. Face recognition across pose: A review. *Pattern Recognition*, 42:2876–2896, Nov. 2009.

[13] X. Zhang and Y. Gao. Heterogeneous specular and diffuse 3-D surface approximation for face recognition across pose. *IEEE Trans. Inf. Forensics and Security*, 7(2):1952–1961, 2012.