Automatic 3D Reconstruction for Face Recognition[•]

Yuxiao Hu[†], Dalong Jiang[*], Shuicheng Yan[‡], Lei Zhang[†], Hongjiang zhang[†]
[†]*{i-yuxhu, leizhang, hjzhang}@microsoft.com, Microsoft Research, Asia, Beijing, China, 100080*
[*]*dljiang@ict.ac.cn, Inst. of Computing Tech., Chinese Academy of Sci., Beijing, China, 100080*
[‡]*scyan@math.pku.edu.cn, School of Mathematical Sci., Peking University, Beijing, China, 100871*

Abstract

An analysis-by-synthesis framework for face recognition with variant pose, illumination and expression (PIE) is proposed in this paper. First, an efficient 2D-to-3D integrated face reconstruction approach is introduced to reconstruct a personalized 3D face model from a single frontal face image with neutral expression and normal illumination; Then, realistic virtual faces with different PIE are synthesized based on the personalized 3D face to characterize the face subspace; Finally, face recognition is conducted based on these representative virtual faces. Compared with other related works, this framework has the following advantages: 1) only one single frontal face is required for face recognition, which avoids the burdensome enrollment work; 2) the synthesized face samples provide the capability to conduct recognition under difficult conditions like complex PIE; and 3) the proposed 2D-to-3D integrated face reconstruction approach is fully automatic and more efficient. The extensive experimental results show that the synthesized virtual faces significantly improve the accuracy of face recognition with variant PIE.

1. Introduction

Recognition of faces in digital photographs remains a challenging problem despite of over three decades of research efforts [1]. Face Recognition Vendor Test in 2002 (FRVT2002) [2] evaluated the state of art algorithms and systems by large-scale, real-world test datasets. The results indicate that face recognition accuracy on frontal face with indoor lighting has reached about 90%. However, face recognition among different pose, illumination and expression (PIE) is still far from satisfactory.

In order to deal with the aforementioned problems, expansion based methods and normalization based methods have been explored in previous works. Expansion based methods tries to utilize more samples which cover different PIE to enhance the representation capability of the face gallery. View-based method [17] has shown its efficiencies, but it needs sufficient gallery samples. To enlarge the training set and improve its representative ability, variant analysis-by-synthesis methods are put forward. Photometric stereo technologies [6] are used to recover the illumination or relight the sample face images. Shape from shading [7] has been explored to extract 3D geometry information of a face and to generate virtual samples by rotating the result 3D face models. It requires that the face images are pixel-wise precisely aligned, which is difficult to be implemented in practical face recognition applications. In contrast to expansion based methods, normalization based methods either tries to normalize probe samples to a unified PIE which is the same or similar to the gallery samples to ensure the generalization capability of the classifier trained on the gallery samples [3], or tries to extract specific features which are invariant or insensitive to different PIE [4]. In the work of Lam *et al.* [5], face samples with out-of-plane rotation are warped to frontal faces according to a cylinder face model, but this method requires heavy manual labeling work. This 2D based methods do not consider the specific structures of human faces, thus frequently leads to the worse performance on face samples with different pose. The most related work to this paper is proposed by Vetter *et al.* [8, 9]. They presented a 3D alignment algorithm to recover the shape and texture parameters of a 3D morphable model. However, the 3D face alignment requires manual initialization and the speed (one minute for a face image) is not able to meet the requirement of most real face recognition systems.

In this paper, an efficient and fully automatic 2D-to-3D integrated face reconstruction method is proposed to tackle the above problems by expansion

[•] Part of this work has been submitted to Journal of Pattern Recognition, Special Issue on Image Understanding for Digital Photographs.

based methods in an analysis-by-synthesis manner. First, frontal face detection and alignment are utilized to locate a frontal face and the facial feature points within an image, such as the contour points of the face, left and right eyes, mouth and nose. Then, the 3D face shape is reconstructed according to the feature points and a 3D face database. After that, the face model is texture-mapped by projecting the input 2D image onto the 3D face shape. Based on this 3D face model, virtual samples with variant PIE are synthesized to represent the 2D face image space. Finally, face recognition is conducted in this enlarged face subspace after standard normalization of testing sample face images. The only input to this system is a frontal face image with normal illumination and neutral expression. The outputs are images with variant PIE for recognition. Compared with previous works, this framework has the following advantages: 1) only one single frontal face is required for training, which avoids the burdensome enrollment work; 2) the synthesized face samples provide the capability of recognizing faces under complex conditions such as arbitrary PIE; 3) the proposed integrated 2D-to-3D face reconstruction approach is fully automatic and the speed is fast. It takes about four seconds per face image (512 × 512 pixels) on a P4 1.3GHz, 256M RAM computer, which is about fifteen times faster than the 3D face alignment in [9].

The rest of this paper is organized as follows. The 2D-to-3D face reconstruction algorithm and the method of generating realistic virtual face sample images with variant PIE is detailed in section 2. Face recognition experimental results are provided in section 3 to justify the efficiency of the proposed algorithm. Section 4 gives conclusions.

2. Automatic 3D Face Reconstruction for Face Recognition

In this section, an efficient and fully automatic framework is proposed for face recognition by performing 3D face reconstruction and generating virtual faces from a single frontal face with normal illumination and neutral expression. The framework consists of two parts: 1) 2D-to-3D integrated face reconstruction; and 2) face recognition using the virtual faces with different PIEs. The following subsections will describe these two parts in detail.

2.1. 2D-to-3D Integrated Face Reconstruction

The only required input to the system is a frontal face image of a subject with normal illumination and neutral expression. In [10], a novel semi-supervised ranking prior likelihood models for accurate local search and a robust parameter estimation approach for face alignment are presented. Based on this 2D alignment algorithm, 83 key feature points are automatically located. The feature points are accurate enough for face reconstruction in most cases. A general 3D face model is applied for personalized 3D face reconstruction. The 3D shapes have been compressed by the Principal Component Analysis (PCA). After the 2D face alignment, the key feature points are used to compute the 3D shape coefficients of the eigenvectors. Then, the coefficients are used to reconstruct the 3D face shape. Finally, the face texture is extracted from the input image. By mapping the texture onto the 3D face geometry, the 3D face model for the input 2D face image is reconstructed.

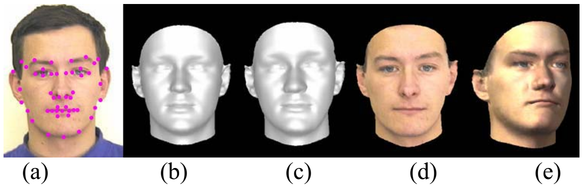

Figure 1. 3D reconstruction.

(a) 2D alignment; (b) 3D shape reconstructed by PCA coefficients of eigenvectors; (c) 3D shape after Kriging interpolation; (d) 3D model with texture; (d) a new view with PIE. (The input image is copied from [16])

2.1.1. Efficient 2D face alignment. Automatic alignment on multi-view face images is still an open problem. But face alignment on frontal face has been well studied [10]. In the proposed method, the input 2D face images are in frontal pose with normal illumination and neutral expression, which is the most common case in face recognition. Under such a constraint, a fast and accurate 2D face alignment algorithm is deployed to locate key facial points such as face contour points, eye centers and nose tip. 83 feature points can be aligned, some of which are selected for face reconstruction as shown in Figure 1(a). The position of these feature points can be modified in case the alignment is not accurate, which rarely occurs.

2.1.2. 3D face geometry reconstruction. To reconstruct a 3D face model, we use the USF Human ID 3-D database which includes 100 laser-scanned heads [8]. Each face model in the database has approximately 70,000 vertices. In this paper, the number of the vertices is reduced to about 8,900 for better performance. Then the vertices on the lip line are duplicated and the triangles around the lip line are reconstructed so that the mouth of the face model is able to be opened for lip motion and expression, which will be described later.

In general, matching a 3D geometry to a 2D image is an ill-posed problem. Fortunately, the differences between the 3D shapes of different face models are not dramatic. In this paper, the geometry of a 3D face model is represented with a shape-vector S=$(X_1, Y_1, Z_1, X_2, ..., Y_n, Z_n)^T \in \Re^{3n}$, which contains the *X, Y, Z* coordinates of its n vertices. Since all facial feature points such as corners of eyes and tips of nose of different faces are fully corresponded by means of their semantic position, PCA is appropriate to be conducted to get a more compact shape representation of face by the primary components. Let $\overline{S}$ be the average shape, $P \in \Re^{3n \times m}$ be the matrix of the first *m* eigenvectors (in descending order according to their eigenvalues). A new face shape S' can be expressed as:

$$S' = \overline{S} + P\vec{\alpha}, \qquad (1)$$

where $\vec{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_m)^T \in \Re^m$ is the coefficients of the shape eigenvectors.

In the alignment step, it is assumed that *t* 2D facial feature points are selected for 3D reconstruction. *t* vertices, corresponding to the feature points, are also chosen on the face geometry. Let S_f=$(X_1, Y_1, X_2, ..., X_t, Y_t)^T \in \Re^{2t}$ be the set of *X, Y* coordinates of the feature vertices on the face. Thus, S_f is the sub shape-vector of *S*. According to equation (1), the *X, Y* coordinates of those feature vertices of a new face shape S_f' can be expressed as:

$$S_f' = \overline{S_f} + P_f\vec{\alpha}, \qquad (2)$$

where $\overline{S}_f \in \Re^{2t}$ and $P_f \in \Re^{2t \times m}$ are the *X, Y* coordinates of the feature vertices on $\overline{S}$ and *P* respectively. To transform face coordinate to image coordinate, let S_f'' be the transformed shape, so:

$$S_f'' = (S_f' + T)c, \qquad (3)$$

where $T \in \Re^{2t}$ is the translation vector and $c \in \Re$ is the scale coefficient. Note that since the 2D face image and 3D face model are both frontal, the rotation matrix is not required. Since P_f is an orthogonal matrix, $\vec{\alpha}$ can be derived from equation (2) as:

$$\vec{\alpha} = (P_f^T P_f)^{-1} P_f^T (S_f' - \overline{S_f}) \cdot \qquad (4)$$

Because the coefficient $\vec{\alpha}$ is computed with part vertices, to avoid getting odd values, eigenvalues are applied as constraints for $\vec{\alpha}$, and the equation (4) is changed to:

$$\vec{\alpha} = (P_f^T P_f + V)^{-1} P_f^T (S_f' - \overline{S_f}) / \vec{v}, \qquad (5)$$

where $$V = \lambda \begin{bmatrix} v_1^2 & 0 & .. & .. & 0 \\ 0 & v_2^2 & & & 0 \\ . & & .. & & . \\ & & & .. & . \\ 0 & 0 & .. & .. & v_m^2 \end{bmatrix},$$

$\vec{v} = (v_1, v_2, ..., v_m)^T \in \Re^m$, λ is a constant, and v_i is the eigenvalue of the ith eigenvector.

In equation (2) and (3), there are five variables ($\vec{\alpha}$, S_f', S_f'', *T*, *c*). To compute the face geometry coefficient $\vec{\alpha}$, an iterative procedure is applied, as outlined below.

Before the first iteration, we let $\overline{S_f}$ be the initial value of S_f'.

Step 1. Let $\overline{\Delta X}$ and $\overline{\Delta Y}$ be the average distance of all t feature points between S_f' and $\overline{S_f}$ along *X, Y* axes respectively, so:

$$\overline{\Delta X} = \frac{1}{t}\sum_{i=1}^{t}(\overline{X_i} - X_i'), \quad \overline{\Delta Y} = \frac{1}{t}\sum_{i=1}^{t}(\overline{Y_i} - Y_i') \cdot$$

Then $T = (\overline{\Delta X}, \overline{\Delta Y}, ..., \overline{\Delta X}, \overline{\Delta Y})^T$,

$$c = \frac{\sum_{i=1}^{t}((\overline{X_i} - \overline{\Delta X})(X_i' - \overline{\Delta X_0}) + (\overline{Y_i} - \overline{\Delta Y})(Y_i' - \overline{\Delta Y_0}))}{\sum_{i=1}^{t}((X_i' - \overline{\Delta X})^2 + (Y_i' - \overline{\Delta Y})^2)},$$

where $\overline{\Delta X_0}$ and $\overline{\Delta Y_0}$ are $\overline{\Delta X}$ and $\overline{\Delta Y}$ in previous iteration respectively. In the first iteration, $\overline{\Delta X_0}$ and $\overline{\Delta Y_0}$ are both set to 0. S_f'' can then be computed from Equation (3).

Step 2. Assign S_f'' to S_f'. The face geometry coefficient $\vec{\alpha}$ can be computed using Equation (5). Then a new S_f' can be obtained by applying $\vec{\alpha}$ to Equation (2).

The geometry coefficient $\vec{\alpha}$ generally converges to a fixed value after repeating step 1 and step 2 for at most 10 iterations. Then we apply $\vec{\alpha}$ to equation (1) to get the whole 3D face geometry S'. The reconstructed face shape is shown as Figure 1 (b). The face geometry looks quite well, but the *X, Y* coordinates of the feature vertices on the face are somewhat different from the *X, Y* coordinates of the feature points on the 2D image. The reason is that the shape space is limited by the 3D face database. To ensure that the feature vertices are exactly correct, the *X, Y* coordinates of the feature vertices on the face are forced to be aligned to the *X, Y* coordinates of the feature points on the 2D image. According to the displacements of the feature vertices, the Kriging interpolation method [11] is used to compute the displacement of non feature vertices. For interpolation purpose, radius base function (RBF) is a good alternative. By using the method we described above, the final 3D face geometry is reconstructed with the accurate feature vertices. The final 3D face shape is shown as Figure 1 (c).

2.1.3. Texture extraction. After the 3D face geometry is reconstructed, the 2D image is projected orthogonally to the 3D geometry to generate the texture. No corresponding color information is available for some vertices because they are occluded in the frontal face image. Therefore, it is possible that there are still some blank areas on the generated texture map, which need to be corrected. In this paper, the thin-plate relaxation [15] is employed to interpolate the blank areas by known colors. The texture mapped face model is shown in Figure 1 (c) and (d).

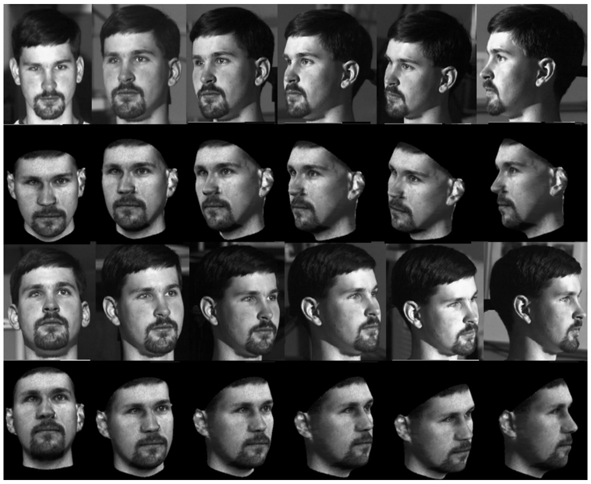

Figure 2. Poses. The first and third lines are poses in CMU-PIE. The second and fourth lines are the corresponding poses generated by rotating the reconstructed model.

2.2. Synthesis with different pose, illumination and expression

In natural environments, PIE remains a critical and challenging issue in face recognition algorithms. To increase the accuracy of face recognition, acquiring sample face images with variant PIE are necessary. However, it's difficult to generate new face images with different PIE from a frontal image using any existing 2D-to-2D methods. The problem could be solved by the proposed approach to reconstructing the 3D face model from the given 2D face image. The reconstructed 3D face model is then rotated to generate images with variant poses. By applying different lights, variant illuminations are created. Finally, a MPEG-4 based facial animation technique is used to generate expressions, which are also an important factor in face recognition but are not considered in most researches.

2.2.1. Pose. Pose variation is the primary source of difficulties for face recognition. The performance of face recognition systems drops dramatically, when large pose variations are presented in the input images, especially when the system's training data have few non-frontal images. A reasonable way to improve multi-view recognition is to use multiple training views. In the proposed work, it is convenient to generate any views by rotating the 3D model to the right pose. The poses in CMU-PIE [12] and those generated by the proposed approach are compared in Figure 2.

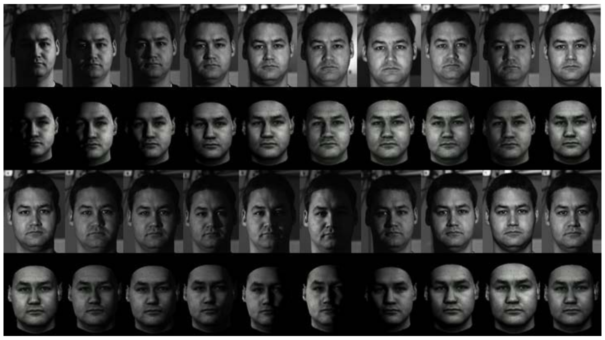

Figure 3. Illuminations. The first and third lines are CMU-PIE images. The second and fourth lines are the corresponding generated images.

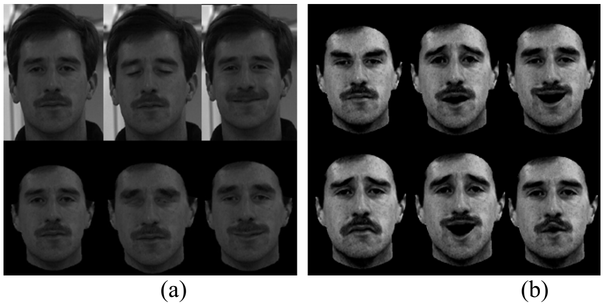

Figure 4. Expressions. (a):The first line is the expressions in CMU-PIE; the second line is the generated expressions corresponding to the first line; (b):Other generated expressions.

For face recognition training purpose, the positions of feature points on the multi-view face images are needed. In general, face alignment on arbitrary multi-view face images is quite difficult, and no technique is able to automatically solve this problem with high accuracy so far. Most multi-view face recognition methods require manually labeling these feature points on large number of training and testing sample images to align them, which is inaccurate and time-consuming.

In the proposed method, since the multi-view images are generated by rotating the 3D face model, the alignment on the new face images is no longer a problem. When a multi-view face image is projected after rotating the 3D model, the positions of facial feature points are obtained by projecting the corresponding feature vertices on the 3D model to 2D image, i.e., no more alignment is required on the generated multi-view images. The acquisition of the feature point positions on the multi-view face images is thus automatic and accurate.

2.2.2. Illumination. Illumination is another important issue for face recognition. The same face appears different due to changes in lighting. The changes induced by illumination are often larger than the differences between individuals.

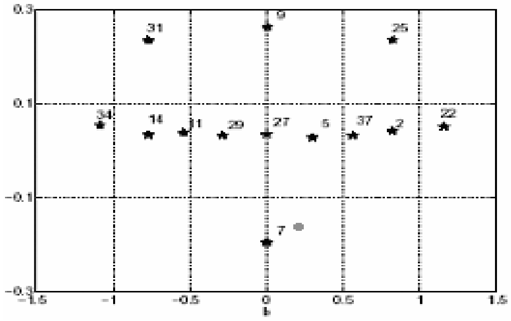

Figure 5. CMU-PIE database camera positions. (The picture is copied from [14].)

In the proposed method, in order to generate variant illumination images of a 3D face model, two lights are applied to the 3D models. One is an environment light; the other is a movable spot light. The whole face model illumination is controlled by the environment light. The specular areas and shadows of the face model are generated by the spot light. Several attributes of the spot light can be controlled, including ambient, diffuse, specular and position. Some illumination images are imitated with CMU-PIE. The Figure 3 shows that the generated illuminations of the images are quite similar to CMU-PIE. If enough lights are applied, most common illumination conditions can be generated.

2.2.3. Expression. In general, expression changes are not as important as pose and illumination changes for face recognition. But, exaggerated expression changes are still a problem to be solved in order to achieve robust face recognition.

In the proposed method, an MEPG-4 based animation framework is used to drive the 3D face model and generate different expressions [13]. As Figure 4 (a) shows, the expressions in the first line are images in CMU-PIE. The expressions in the second line, which corresponds to the first line, are generated by the MPEG-4 based animation algorithm. Figure4 (b) shows some other generated expressions.

3. Experiments

In this section, face recognition performance of the proposed algorithm across variant PIE is systematically evaluated by comparing them with the conventional algorithm that do not use the virtual faces synthesized from the personalized 3D face models. The CMU-PIE database is used in the evaluation since it takes into account all the three factors. The CMU-PIE database contains 68 subjects with 41,368 face images, captured by 13 synchronized cameras and 21 flashes, under varying PIE. The CMU-PIE database camera positions are shown in Figure 5.

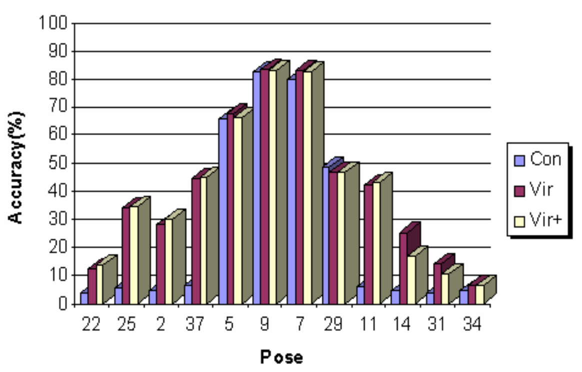

Figure 6. Recognition accuracy comparison between face recognition with/without virtual face using PCA. (Con: Conventional Algorithms; Vir: Using all the virtual faces; Vir+: Using virtual face for special pose)

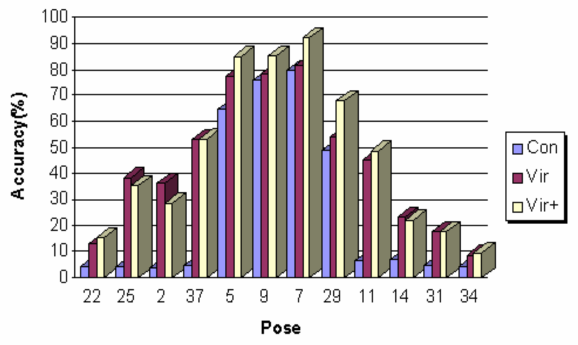

Figure 7. Recognition accuracy comparison between face recognition with/without virtual face using LDA. (Con: Conventional Algorithms; Vir: Using all the virtual faces; Vir+: Using virtual face for special pose)

The frontal face at pose-27 with neutral expression and environment light are used to automatically construct the personalized 3D faces. All the 68 3D faces are constructed and the virtual faces with different PIE combinations are synthesized; the comparisons with the real faces are illustrated in Figure 2-4. Note that all the reconstruction is fully automatic. Only one frontal face of a subject with normal illumination and neutral expression is required to construct the personalized 3D face model of the subject, which can be easily satisfied in real application.

In all the experiments, the conventional method used only the frontal faces at pose-27 for training and the other faces are all used for testing. The comparison experiments have been conducted to evaluate the effectiveness of the virtual faces from constructed 3D face model for face recognition with arbitrary PIE. We used two traditional methods, Principal Component Analysis (PCA) and Linear Discriminant Analysis

(LDA), to perform dimensionality reduction, and we used the Nearest Neighbors (NN) as similarity matching approach for classification. The results using PCA and LDA are illustrated in Figure 6 and Figure 7. From the results, it is observed that:

1) In general, face recognition accuracy using virtual faces from reconstructed 3D faces is higher than conventional algorithms, especially for the experiment using LDA and with the pose information ahead.

2) The proposed algorithm significantly improved the performance in half-profile views, like pose 37 and 11; while for the profile views, the improvements are limited. This is because that the reconstructed texture for the unseen points in frontal view is not accurate enough. We are exploring new methods for realistic missing data reconstruction, like using the 3D texture models.

3) With prior pose information, if separated models for each view are constructed respectively as in the experiments "Vir+", the recognition performance was improved than that using all the virtual faces and one single global model as in the experiments "Vir".

4. Conclusion

Experimental evaluation of face reconstruction for face recognition have illustrated that the proposed fully automatic system is efficient accurate and robust. Compared with other related works, this framework has the following highlights: 1) only one single frontal face is required for face recognition and the outputs are realistic images with variant PIE for the individual of the input image, which avoids the burdensome enrollment work; 2) the synthesized face samples provide the capability to conduct recognition under difficult conditions of complex PIE; and 3) the proposed 2D-to-3D integrated face reconstruction approach is fully automatic and faster than other 3D reconstruction approaches.

5. Acknowledgements

The authors thank the University of South Florida for providing the USF Human ID 3-D Database.

6. References

[1] T. Kanade, "Picture Processing System by Computer Complex and Recognition of Human Faces", *Doctoral Dissertation*, Kyoto University, Nov. 1973.

[2] P. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi, M. Bone, "Face Recognition Vendor Test 2002: Evaluation Report", *FRVT,* 2002.

[3] T. Jebara, "3D Pose Estimation and Normalization for Face Recognition", *Undergraduate Thesis,* Centre for Intelligent Machines, McGill University, 1995.

[4] J. Lai, P. Yuen, G. Feng, "Face Recognition Using Holistic Fourier Invariant Features", *Pattern Recognition*, 34(1), 2001, pp95-109.

[5] K. Lam and H. Yan, "An Analytic-to-Holistic Approach for Face Recognition Based on a Single Frontal View", *IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI)*,.2(7), 1998, pp673-686.

[6] T. Riklin-Raviv, A. ShaShua, "The Quotient Image: Class Based Re-rendering and Recognition With Varying Illuminations", *PAMI*, 23(2), 2001.

[7] R. Zhang, P. Tai, J. E. Cryer, M. Sha, "Shape From Shading: A Survey", *PAMI*, 21(8), 1999, pp690-706.

[8] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D-faces", In *Proc. of ACM SIGGRAPH*, Los angeles, 1999, pp187-194.

[9] S. Romdhani, V. Blanz, T. Vetter, "Face Identification by fitting a 3d Morphable Model Using Linear Shape and Texture Error Functions", In *Proc. of European Conf. on Computer Vision*, V4, 2002,pp3-19.

[10] S. C. Yan, M. J. Li, H. J. Zhang, Q. S. Cheng. "Ranking Prior Likelihood Distributions for Bayesian Shape Localization Framework", *In Proc. of Intl. Conf. on Computer Vision*, France, Nice, 2003.

[11] M. A. Oliver and R. Webster, "Kriging: a method of interpolation for geographical information system", *Int. Jornal on Geographical Information Systems*, 4(3), 1990, pp313-332.

[12] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination, and Expression (PIE) Database", In *Proc. of Intl. Conf. on Automatic Face and Gesture Recognition*, 2002.

[13] D. Jiang, W. Gao, Z. Li, Z. Wang, "Animating Arbitrary Topology 3D Facial Model Using the MPEG-4 FaceDefTables", In *Proc. of IEEE Intl. Conf. on Multi-modal Interface*, USA, 2002, pp 517-522.

[14] R. Gross, J. Shi, and J. Cohn, "Quo vadis face recognition?", In *Third Workshop on Empirical Evaluation Methods in Computer Vision*, 2001.

[15] D.Terzopoulos, "The computation of visible-surface representations", *PAMI*, 10(4), 1988, pp417-438.

[16] A. R. Martinez and R. Benavente, "The AR face database", *Technical Report* 24, Computer Vision Center, Barcelona, Spain, June 1998.

[17] A.Pentland, B. Moghaddam, T. Starner, O. Oliyide, and M. Turk., "View-Based and Modular Eigenspaces for Face Recognition", *Technical Report* 245, MIT Media Lab, 1993.