Reconstructing 3D Face Shapes from Single 2D Images Using an Adaptive Deformation Model

Ashraf Y.A. Maghari[1], Ibrahim Venkat[1], Iman Yi Liao[2], and Bahari Belaton[1]

[1] School of Computer Sciences, Universiti Sains Malaysia, Pinang, Malaysia
myashraf2@gmail.com
[2] School of Computer Science, University of Nottingham Malaysia Campus

Abstract. The Representational Power (RP) of an example-based model is its capability to depict a new 3D face for a given 2D face image. In this contribution, a novel approach is proposed to increase the RP of the 3D reconstruction PCA-based model by deforming a set of examples in the training dataset. By adding these deformed samples together with the original training samples we gain more RP. A 3D PCA-based model is adapted for each new input face image by deforming 3D faces in the training data set. This adapted model is used to reconstruct the 3D face shape for the given input 2D near frontal face image. Our experimental results justify that the proposed adaptive model considerably improves the RP of the conventional PCA-based model.

Keywords: Representational Power, Statistical facial modeling, 3D face reconstruction, PCA, TPS.

1 Introduction

Reconstruction of 3D face images from single 2D images is an open problem in the field of computer vision. The need for 3D face reconstruction has grown in applications such as virtual reality simulations, face recognition [8,9], and plastic surgery [6]. For example, in biometric identification, it has been shown that face recognition rate could be significantly improved by incorporating 3D face shapes with 2D face images [11]. The objective of 3D facial reconstruction systems is to recover the three dimensional shape of individuals from their 2D pictures or video sequences. However, accurate reconstruction of a person's 3D face model from his/her 2D face images still remains as a challenge.

There are many approaches for the reconstruction of 3D faces from single images. One of such early techniques being utilized is Shape-from-Shading (SFS) [1,23], which capitalizes the idea that the depth information is related to the intensity of a face image acquired through a given/chosen reflectance model. It has been shown that SFS suffers from poor global shape control.

Recently, Kemelmacher-Shlizerman and Basri proposed an approach that combines shading information with generic shape information derived from a single reference model by utilizing the global similarity of faces [13]. In this method the

H. Badioze Zaman et al. (Eds.): IVIC 2013, LNCS 8237, pp. 300–310, 2013.

involved fitting process requires boundary conditions and parameters to be adjusted during the reconstruction process. However, a 3D reference model which keeps shape similarities with the input image has not been considered in this method. Owing to this, inaccurate 3D shape estimation might be possible.

There are also conventional learning-based methods, such as neural networks [16] and typical statistical learning-based methods such as 3D Morphable Model (3DMM) [4]. The advancement of 3D scanning technology has led to the creation of more accurate 3D face exemplar models [18]. Example-based modeling allows more realistic face reconstruction than other methods [15]. However, the quality of face reconstruction using such models is affected by the chosen examples. For example, Kemelmacher-Shlizerman and Basri [13] have emphasized that learning a generic 3D face model requires large amounts of 3D faces. Furthermore, analytical results in [20] show that in many cases the representational power of the model may vary if the model is trained with a different sample though the same sample size is retained.

The PCA-based model proposed by Blanz et al. with relatively small sample size (100 faces) which was used for face recognition has obtained reasonable results [3]. Although in some statistical modeling methods both shape and texture are modeled separately using PCA (e.g. 3DMM), it has been suggested that shapes are more amenable than texture, as textures are subject to vast variation when compared to shape based features [12]. Therefore, the model we intend to propose in this contribution is based on modeling of shapes. When shapes are considered, the reconstruction of 3D face from 2D images using shape models is relatively simple. A popular method is a regularization based reconstruction where a few feature points are selected as the observations for reconstruction [12].

This study addresses the problem of increasing the Representational Power (RP) of the PCA-model to improve its capability in depicting a new 3D face from a given input face image. A 3D face shape modeling scheme is proposed to handle the vital model adaptation part of the PCA-based model. There are other methods that intend to create synthetic views in training sets for face recognition. These synthetic views includes different pose and expression [17]. However, this work is different and novel in the context of deforming 3D faces using the given input face for 3D face reconstruction.

The rest of the paper is organized as follows: Section 2 demonstrates the representational power of PCA model. Section 3 describes the methodology of the proposed adaptive 3D face shape modeling approach. Section 4 deals with the experimental evaluation and associated discussions. Section 5 concludes our research.

2 Representational Power of PCA Model

In this study, we define the RP of the PCA-based model as its capability in depicting a new 3D face of a given input face image. The capability of the PCA model can be measured by evaluating the quality of a reconstruction (with respect to its ground-truth). The most obvious and appropriate choice to compare

two 3D surfaces is the Euclidean distance. The sum of the Euclidean distances over all vertices of the shapes weighted by the number of vertices is our measure:

$$Ed_w = \frac{\sum_{i=1}^{n} \sqrt{(s_i - s_{ri})^2}}{n} , \quad (1)$$

where Ed_w is the weighted Euclidean distance, s is the probe face shape, s_r is the reconstructed face shape, n is the number of vertices of the face shape.

As an example, Fig. 1 demonstrates the advantage of RP in terms of evaluating the quality of reconstruction. An original testing face shape is projected to PCA-based models learned from different training set sizes. As one can see in Fig. 1, the projection (representation) gets more realistic and more closer to the ground truth when Ed_w decreases, which means that the PCA model that represent a new 3D face with less Ed_w has more RP. In the next section, we propose a novel method that is able to improve the RP of the PCA model for the same training data set by reducing the Euclidean distance.

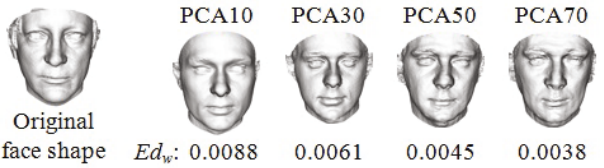

Fig. 1. Projecting the leftmost face shape to PCA models trained with different sample sizes. The RPs that represent the quality of projected face are shown below the shapes. PCA10 means the training set has 10 examples, PCA30 means the training set has 30 examples and so on

3 Adaptive 3D Face Shape Modeling

Fig. 2 shows the bird's eye view of the proposed scheme. It shows how a given input face gets reconstructed via a deformation synthesis mechanism.

Similar to Farkas [10] as referred by Knothe et al. [14] firstly we chose landmarks over facial regions such as eyes, nose, mouth and chin for the purpose of face alignment and deformation.

The input feature points are aligned using a standard algorithm called *Procrustes Analysis*. The concept of the Procrustes Analysis is similar to Iterative Closest Point (ICP) [2].

3.1 Deforming 3D Exemplar Faces

We use TPS to establish the mapping and interpolation for the deformation process. TPS is a commonly used basis function for representing coordinate mappings from rigid to nonrigid deformations. It is used for estimating a deformation function between two surfaces [17]. Let g_0 and g_1 be two 2D/3D shapes, and $P = (p_1, p_2, ..., p_m) \subset g_0$ and $V = (v_1, v_2, ..., v_m) \subset g_1$ be the correspondences (landmarks)

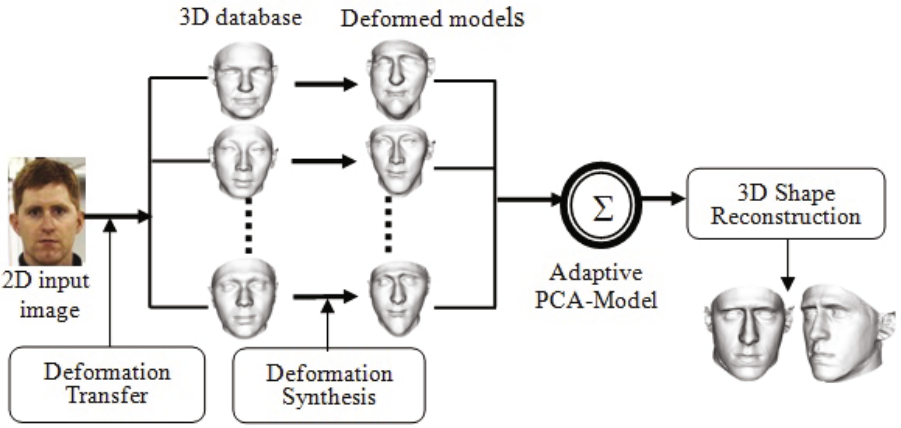

Fig. 2. Proposed scheme of the deformation model for 3D face shape reconstruction

between the two shapes, where m is the number of corresponding points. A warping function F that warps point set P to V is given by the following condition:

$$F(p_j) = v_j \ , \qquad j = 1, 2, ..., m \tag{2}$$

For the two corresponding sets of landmarks P and V, F is unique and has a minimal bending energy [7]. A TPS can minimize the following energy function

$$E_\lambda = \frac{1}{m} \sum_{i=1}^{m} |v_i - F(p_i)| + \lambda J \ , \qquad i = 1, ..., m \tag{3}$$

where v_i represents the i^{th} 2D/3D point (landmark) of the input face (base points) and $p_i = (p_{i1}, ..., p_{id})$ is the i^{th} point given in d-dimensions (in our case d could be 2D or 3D, p_i represent the set of points used to warp an image), m is the total number of corresponding points, J is a smoothness penalty function in d-dimensions, and λ is the smoothness parameter [24]. For the approximating case, minimizing Eq. (3) leads to the following matrix form.

$$PA + (K + m\lambda I)W = V, \quad P^T W = 0, \tag{4}$$

which actually performs the standard QR decomposition. Obviously, a QR decomposition of the matrix P produces an orthogonal matrix Q and an upper triangular matrix R such that $P = QR$ [24].

As an example, for the case $d = 2$ (2-dimension), the interpolation map is form R^2 *to* R^2, where $p_i = (x_i, y_i)$ and $v_i = (x'_i, y'_i)$. Let $r_{ij} = |p_i - p_j|$ be the distance between points i and j. Define matrices

$$K \ = \ \begin{bmatrix} 0 & U(r_{12}) & ... & U(r_{1m}) \\ U(r_{21}) & 0 & ... & U(r_{2m}) \\ ... & ... & ... & ... \\ U(r_{n1}) & U(r_{m2}) & ... & 0 \end{bmatrix}, \ m \times m \tag{5}$$

$$P = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ ... & ... & ... \\ 1 & x_m & y_m \end{bmatrix}, \ m \times 3 \tag{6}$$

and

$$L = \left[\begin{array}{c|c} K & P \\ \hline P^T & O \end{array}\right], \ (m+3) \times (m+3) \tag{7}$$

where O is a 3×3 matrix of zeros. Let $Y = (V|0 \ \ 0 \ \ 0)^T$, a column vector of length $n+3$ and $W = (w_1, w_2, ..., w_m)$. For the interpolating case, TPS provides a linear system of equations [5] which is given by

$$L^{-1}Y = (W|a_1 \ \ a_x \ \ a_y)^T, \tag{8}$$

The element of $L^{-1}Y$ are used to define a function $F(x, y)$ everywhere in the plane:

$$F(x, y) = a_1 + a_x x + a_y y + \sum_{i=1}^{m} w_i U(|p_i - (x, y)|), \tag{9}$$

For simplifying, F can be written in the following matrix form:

$$F(p) = p.A + KW \ , \tag{10}$$

where $p \in g_0, A = (a_1 \ \ a_x \ \ a_y)$ is an affine transformation, W is a fixed m-dimensional column vector of non-affine warping parameters constrained to $P^T W = 0$ and K is m-dimensional row vector with $K_{ij} = U(|p_i - p_j|)$.

Some typical deformed 3D faces registered with reference to three typical 2D images using TPS are shown in Fig. 3.

3.2 Deformable Model Construction

Usually, each synthesized 3D model captures only some characteristics from its corresponding input face. This leads to a 3D morphable model that is a linear combination of 3D face shapes, some of which are obtained as a result of a deformation transfer mechanism from one input face to the neutral 3D example face shapes. The linear combination is controlled by shape parameters α where a new 3D shape can be generated using:

$$s = s_0 + \sum_{i=1}^{m} \alpha_i e_i \ , \tag{11}$$

where s_0 is the mean 3D shape, e_i represent the i^{th} eigenvector of the covariance matrix, α_i is the coefficient of the shape eigenvector e_i and m is the number of significant eigenvectors. The coefficient of a face shape s can be calculated using the following equation

$$\alpha = E^T(s_i - s_0) \ , \tag{12}$$

where $E = [e_1, e_2, ..., e_m]$ are the eigenvectors of the covariance matrix. The projected new face shape can be precisely represented by applying Equation (11).

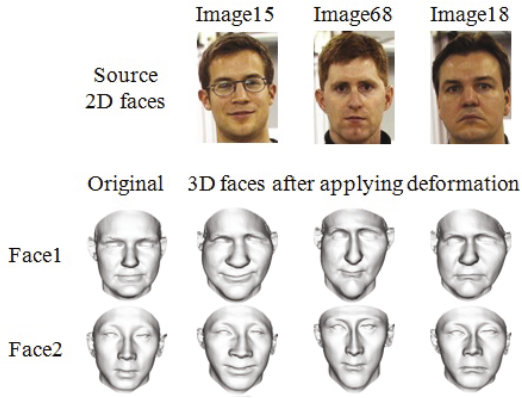

Fig. 3. Typical 3D-2D registration scheme based on the proposed deformation model. The top row shows three 2D face images. Column 1 (left most) shows original two 3D faces. The corresponding deformed faces are shown in columns 2, 3 and 4 (right most).

3.3 3D Face Reconstruction Based on Regularization

After training the 3D face model with the new training set (original training face shapes and deformed 3D face shapes), the well known regularized algorithm has been used for 3D face shape reconstruction. The regularized algorithm has been categorized as one of the existing four core methods for 3D facial reconstruction [15]. The manually selected feature points have been used to compute the 3D shape coefficients of the eigenvectors using equation (13). Then, these coefficients were used to reconstruct the 3D face shape using equation (11). Let t be the number of feature points that can be selected from the input 2D face image, $S_f = (p_1, p_2, ..., p_t) \in R^{2t}$ be the set of selected points on the 2D face image, whereas every point p_i has 2 axes viz., x and y, $S_{f0} \in R^{2t}$ be the t corresponding points on S_0 (the average 3D face shape) and $E_f \in R^{2t \times m}$ be the t corresponding columns on $E \in R^{3n \times m}$ (the matrix of row eigenvectors) where m and n respectively represent the first potential eigenvectors and the number of vertices of the eigenvectors. Then the coefficient α of a new 3D face shape can be derived as

$$\alpha = (E_f^T E_f + \lambda \Lambda^{-1})^{-1} E_f^T (S_f - S_{f0})) \ , \tag{13}$$

where Λ is a diagonal $m \times m$ matrix with diagonal elements being the eigenvalues and λ being the weighting factor. Then α is applied to equation (11) to obtain the whole 3D face shape.

4 Experiments and Discussion

In this section we intend to report the experimental evaluation aspects of the proposed adaptive PCA-based model in comparison to the standard PCA-based

model. We have systematically categorized our experimental study in terms of two phases: In the first phase, the adaptive PCA-based model is evaluated quantitatively in comparison to the standard PCA-based model. In the second phase the model is qualitatively evaluated pertaining to the visualization aspect of the reconstructed faces from their 2D face images. The USF Human ID 3D Face database [4] which contains 100 3D faces has been used. The 100 3D face shapes were divided into two sets; 80 face shapes have been used for training and deforming purposes and the remaining 20 3D face shapes have been used for testing. Further to visually evaluate the accuracy of 3D reconstruction from single 2D images, the CMU-PIE database [22] has been used.

4.1 Representational Power (RP) of the Adaptive Model

As an example of 40 deformed faces, the adaptive PCA model has been trained with 120 3D face shapes which include 80 original training 3D face shapes and the 40 deformed face shapes while the standard PCA model has been trained with 80 original 3D face shapes. Fig. 4 shows the representation errors (Ed_w) found by the standard PCA-based model and the adaptive PCA-based model for 20 probe face shapes. We perceive that Ed_w of the new face shape actually represents the accuracy of representation. It can be seen that the adaptive PCA-based model reduces the representative errors when compared to the standard PCA-based model for all probe faces.

Moreover, the statistical t-Test has been applied to compare the two models. The α-value of the t-Test (level of significance) has been chosen to be 0.05 which means that the two models have been compared at a 95% confidence level. The average results of the 20 probe face shapes shown in Table 1 indicates that the adaptive PCA-based model outperforms the standard PCA-based model with a 95% confidence level, whereas the P-value of the t-Test corresponding to the two models is less than $\alpha = 0.05$ (level of significance). This indicates that there is a statistically significant difference between the representation errors of the two

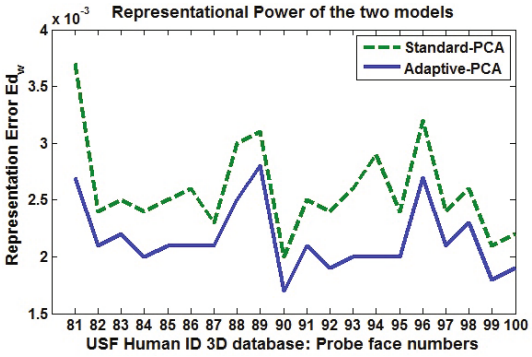

Fig. 4. Comparison between the standard-PCA and the adaptive-PCA in terms of *Representational Power*

Table 1. Comparative results between the representation errors of standard PCA-based model and the adaptive PCA-based model

Model	Mean Error	Std. Dev.	P-value of t-test
Standard Model	2.59×10^{-3}	4.07×10^{-4}	1.07×10^{-8}
Adaptive Model	2.15×10^{-3}	3.02×10^{-4}	

models and justifies that the proposed method yields better RP than that of the standard PCA-based model.

Three typical representations of probe face shapes represented using both models are visualized in Fig. 5. The represented 3D face models for some typical face images shown in Fig. 5 clearly demonstrate that sharp features of the facial components (eg. nose, lips) are well retained by the proposed adaptive PCA-based model when compared to the standard PCA-based model.

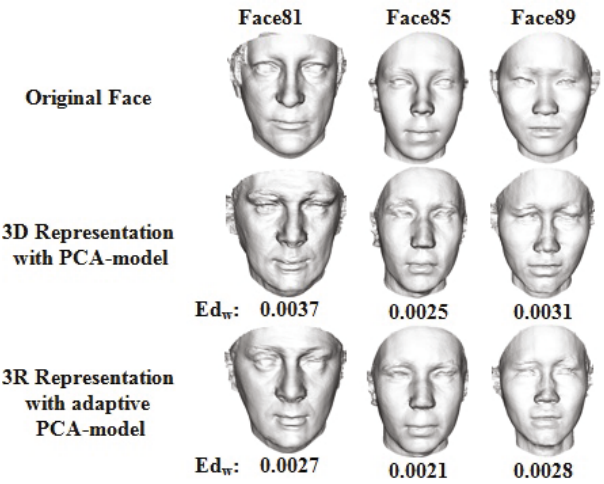

Fig. 5. Visual comparison of represented 3D face shapes using the standard PCA-based model and the adaptive PCA-based model

4.2 3D Face Shape Reconstruction from 2D Image

As an example, the adaptive PCA model has been trained with 90 3D face shapes including 80 original training 3D face shapes and the 10 deformed face shapes while the standard PCA model has been trained with 80 original 3D face shapes. This number of training faces was potentially suitable for building a 3D face reconstruction model where only limited number of feature points was used for 3D face reconstruction. The CMU-PIE database [22] have been used for testing the visual effects of the proposed model. We intend to reconstruct 3D models for the 2D images of CMU-PIE database. The comparisons between some typical 2D face images and their 3D reconstructions using standard PCA model and the adaptive PCA model are illustrated in Fig. 6. From the results

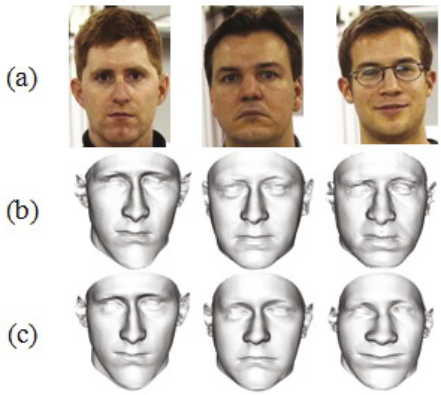

Fig. 6. Visual comparison. (a) Typical input 2D images; (b) 3D reconstruction using normal PCA-based model; (c) 3D reconstruction using adaptive PCA-based model.

in Fig. 6(c), one could notice some visual improvements in the reconstructed 3D face shapes. For example, in the middle face image of Fig. 6(c), the reconstructed 3D face shape has retained some expression of the input image (Fig. 6(a)) such as the facial grimace, chin features, and lips expressions. In the right most face image of the same figure, the 3D face shape has been reconstructed from the 2D image without losing the smile expression. This means that the capability of the model to depict a new 3D face can be improved when 3D exemplar training faces are deformed with the guidance of the input 2D image. However, in addition to the number of feature points used for 3D face shape reconstruction [19], the accuracy of reconstruction can be affected by the number and position of feature points used for deformation modeling, and proportion of deformed faces vs. the original faces in the training data.

Interestingly the proposed model is capable of reconstructing 3D faces from 2D face images by retaining facial expressions though the training samples what we have used contain only neutral expression. By this way we don't impose that the training samples should contain a variety of expressions as imposed by certain recent approaches (eg. Shu-Fan and Shang-Hong [25]).

All experiments were implemented on a workstation with processor Intel(R) Xeon(R) CPU E5620 @ 2.40GH. Assuming that the feature points are available, our MATLAB implementation of the algorithm (including deforming 10 faces, rebuilding the PCA-model and reconstructing the complete face shape vector) takes approximately 72 seconds. Compared with other methods, the proposed method is able to outperform those proposed in [3] and [25] in terms of efficiency. It is also comparably comparative with others such that proposed in [21] and [13].

5 Conclusion and Future Work

In this paper, a novel approach for the problem of 3D face reconstruction from single 2D face images have been proposed. A 3D deformable PCA-based model

has been adapted for a given input 2D face image by deforming 3D faces in the training data set so as to gain significant RP. Then, the adapted PCA-based model is used to reconstruct a 3D face shape from the given input 2D image using a number of feature points. The experimental results demonstrate that the proposed deformation model scheme increases the representational power of the PCA-based model for any given input face image. However, deforming 3D face shapes using TPS tends to increase the computational cost of the proposed scheme compared to the standard PCA-based model. Hence, deformation techniques other than TPS would be considered for the future work to improve the deformation synthesis. Furthermore, we will study the effect of deformed faces number on the reconstruction accuracy.

References

1. Atick, J., Griffin, P., Redlich, A.: Statistical approach to shape from shading: Reconstruction of three-dimensional face surfaces from single two-dimensional images. Neural Computation 8(6), 1321–1340 (1996)
2. Besl, P., McKay, H.: A method for registration of 3D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(2), 239–256 (1992)
3. Blanz, V., Vetter, T.: Face recognition based on fitting a 3D morphable model. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(9), 1063–1074 (2003)
4. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1999, pp. 187–194. ACM Press/Addison-Wesley Publishing Co., New York (1999), http://dx.doi.org/10.1145/311535.311556
5. Bookstein, F.L.: Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence 11(6), 567–585 (1989)
6. Bottino, A., De Simone, M., Laurentini, A., Sforza, C.: A new 3D tool for planning plastic surgery. IEEE Transactions on Bio-medical Engineering (2012)
7. Brown, B., Rusinkiewicz, S.: Non-rigid range-scan alignment using thin-plate splines. In: Proceedings of the 2nd International Symposium on 3D Data Processing, Visualization and Transmission, 3DPVT 2004, pp. 759–765. IEEE (2004)
8. Elyan, E., Ugail, H.: Reconstruction of 3D human facial images using partial differential equations. JCP, 1–8 (2007)
9. Fanany, M.I., Ohno, M., Kumazawa, I.: Face Reconstruction from Shading Using Smooth Projected Polygon Representation NN. In: Proceedings of the 15th International Conference on Vision Interface, Calgary, Canada, pp. 308–313 (May 2002)
10. Farkas, L.G.: Anthropometry of the head and face, 2nd edn. Raven Press (1994)
11. Hu, Y., Jiang, D., Yan, S., Zhang, L.: Automatic 3D reconstruction for face recognition. In: Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 843–848. IEEE (2004)
12. Jiang, D., Hu, Y., Yan, S., Zhang, L., Zhang, H., Gao, W.: Efficient 3D reconstruction for face recognition. Pattern Recogn. 38, 787–798 (2005), http://dx.doi.org/10.1016/j.patcog.2004.11.004

13. Kemelmacher-Shlizerman, I., Basri, R.: 3D face reconstruction from a single image using a single reference face shape. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(2), 394–405 (2011)
14. Knothe, R., Romdhani, S., Vetter, T.: Combining PCA and LFA for surface reconstruction from a sparse set of control points. In: 7th International Conference on Automatic Face and Gesture Recognition, FGR 2006, pp. 637–644. IEEE (2006)
15. Levine, M.D. (Chris) Yu, Y.: State-of-the-art of 3D facial reconstruction methods for face recognition based on a single 2D training image per person. Pattern Recogn. Lett. 30, 908–913 (2009), http://dl.acm.org/citation.cfm?id=1552570.1552692
16. Lin, C., Cheng, W., Liang, S.: Neural-network-based adaptive hybrid-reflectance model for 3D surface reconstruction. IEEE Transactions on Neural Networks 16(6), 1601–1615 (2005)
17. Lu, X., Jain, A.: Deformation modeling for robust 3D face matching. IEEE Transactions on Pattern Analysis and Machine Intelligence 30(8), 1346–1357 (2008)
18. Luximon, Y., Ball, R., Justice, L.: The 3D chinese head and face modeling. Computer-Aided Design 44(1), 40–47 (2012), digital Human Modeling in Product Design, http://www.sciencedirect.com/science/article/pii/S0010448511000133
19. Maghari, A.Y.A., Liao, I.Y., Belaton, B.: Effect of facial feature points selection on 3D face shape reconstruction using regularization. In: Huang, T., Zeng, Z., Li, C., Leung, C.S. (eds.) ICONIP 2012, Part V. LNCS, vol. 7667, pp. 516–524. Springer, Heidelberg (2012)
20. Maghari, A., Liao, I., Belaton, B.: Quantitative analysis on PCA-based statistical 3D face shape modeling. In: Computational Modelling of Objects Represented in Images III: Fundamentals, Methods and Applications, vol. 13 (2012)
21. Romdhani, S., Blanz, V., Vetter, T.: Face identification by fitting a 3D morphable model using linear shape and texture error functions. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part IV. LNCS, vol. 2353, pp. 3–19. Springer, Heidelberg (2002)
22. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination, and expression database. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(12), 1615–1618 (2003)
23. Smith, W., Hancock, E.: Recovering facial shape using a statistical model of surface normal direction. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(12), 1914–1930 (2006)
24. Wahba, G.: Spline models for observational data, vol. 59. Society for Industrial Mathematics (1990)
25. Wang, S.F., Lai, S.H.: Reconstructing 3D face model with associated expression deformation from a single face image via constructing a low-dimensional expression deformation manifold. IEEE Transactions on Pattern Analysis and Machine Intelligence 33(10), 2115–2121 (2011)