Automatic, Effective, and Efficient 3D Face Reconstruction from Arbitrary View Image

Changhu Wang[1,3], Shuicheng Yan[2], Hua Li[2], Hongjiang Zhang[3], and Mingjing Li[3]

[1] Department of EEIS, University of Science and Technology of China
wch@ustc.edu
[2] LMAM, School of Mathematical Sciences, Peking University, P.R. China
{scyan, lihua}@math.pku.edu.cn
[3] Microsoft Research Asia, Beijing, P.R. China
{hjzhang, mjli}@microsoft.com

Abstract. In this paper, we propose a fully automatic, effective and efficient framework for 3D face reconstruction based on a single face image in arbitrary view. First, a multi-view face alignment algorithm localizes the face feature points, and then EM algorithm is applied to derive the optimal 3D shape and position parameters. Moreover, the unit quaternion based pose representation is proposed for efficient 3D pose parameter optimization. Compared with other related works, this framework has the following advantages: 1) it is fully automatic, and only one single face image in arbitrary view is required; 2) EM algorithm and unit quaternion based pose representation are integrated for efficient shape and position parameters estimation; 3) the correspondence between 2D contour points and 3D model vertexes are dynamically determined by normal direction constraints, which facilitates the 3D reconstruction from arbitrary view image; 4) a weighted optimization strategy is applied for more robust parameter estimation. The experimental results show the effectiveness of our framework for 3D face reconstruction.

1 Introduction

Modeling 3D human faces has been a challenging problem in computer graphics and computer vision literatures in the last decades. Since the pioneering work of Parke [8][9], various techniques have been reported for modeling the geometry of faces [5][11]. The 2D-based methods do not consider the specific structure of human faces, thus result in the poor performance on profile face samples. In the work of Lam et al. [4], face samples with out-of-plane rotation are warped into frontal faces based on a cylinder face model, but it requires heavy manual labeling work. Shape from shading [13] has been explored to extract 3D face geometry information and generate virtual samples by rotating the generated 3D models. This algorithm requires that the face images are precisely aligned pixel-wise, which is difficult to be implemented in practice and even impossible for practical face recognition applications.

K. Aizawa, Y. Nakamura, and S. Satoh (Eds.): PCM 2004, LNCS 3332, pp. 553–560, 2004.

The two most popular works on 3D face modeling and analysis are the morphable 3D face model proposed by Vetter et al. [10] and the artificial 3D shape model by Zhang et al. [7]. The former one presented a 3D face reconstruction algorithm to recover the shape and texture parameters based on a face image in arbitrary view, and the latter developed a system to construct textured 3D face model from video sequence. Recently, Hu and Yan et al. [3] presented an automatic 2D-to-3D integrated face reconstruction method to recover the 3D face model based on a frontal face image and it is much faster. However, there are still some shortcomings in these works: 1) both Vetter and Zhang's works require manual initialization and the speed can not satisfy the requirement of a practical face recognition system; 2) Vetter's work needs a large number of samples for a representative texture model, and mostly the small number of texture samples will limit the generalization of the algorithm; 3) Hu and Yan's work assumed fixed pose parameters which limited its extension to side view images.

In this paper, we propose a fully automatic, effective and efficient framework for 3D face reconstruction based on a single face image in arbitrary view. It not only inherits the advantages of the above three works, but also successfully overcomes their shortcomings. First, a recently developed multi-view face alignment algorithm [6] is utilized to localize the feature points in a face; Second, the 3D face shape and pose parameters are estimated synchronously by an EM based algorithm, in which the correspondence between the contour points and their vertex indices in the 3D face models are dynamically determined; moreover, a unit quaternion based pose representation is proposed for efficient position parameter optimization; Finally, the complete 3D face model is obtained by mapping the input 2D image onto the 3D face shape surface with the mirror and smoothing operation.

The rest of this paper is organized as follows. The 2D-to-3D face reconstruction algorithm is described in detail in Section 2. Section 3 provides some experimental results. We draw the conclusions and discuss the future work in Section 4.

2 3D Reconstruction with Single Arbitrary View Image

In this section, we present our fully automatic framework for 3D face reconstruction. In [3], Hu and Yan et al. proposed an automatic algorithm for 3D face reconstruction; however, it can only handle frontal faces. In this framework, we utilize a newly developed multi-view face alignment algorithm [6] to locate the feature points in an arbitrary view face image; then, the 3D shape and position parameters are efficiently estimated with the EM algorithm in term of the unit quaternion [1][2] pose representation and the dynamical correspondence between the contour points and the vertexes on the 3D face model. Moreover, a weighted optimization strategy is applied for robust parameter estimation. This section consists of five parts: 1) the efficient multi-view 2D face alignment; 2) the morphable 3D face model; 3) problem formulation; 4) efficient parameter inference; and 5) robust parameter estimation with the dynamic correspondence strategy and weighted optimization.

2.1 Efficient Multi-view 2D Face Alignment

Automatic multi-view face alignment is still an open problem. In this work, we apply the recently proposed multi-view 2D alignment algorithm [6]. In [6], the texture is redefined as the unwarped grey-level edge in the original image; then, a Bayesian network is designed to describe the intrinsic co-constrains between shape and texture; finally, the EM algorithm is utilized to infer the optimal parameters of the proposed Texture-Driven Shape Model. There are 83 feature points located, part of which are adaptively selected for 3D face reconstruction in different views.

2.2 Morphable 3D Face Model

Similar to Vetter's work [10], the geometry of a 3D face is represented as a shape vector $S = (x_1, y_1, z_1, x_2, ..., y_L, z_L)' \in \mathcal{R}^{3L}$, which contains the x, y and z coordinates of the L vertices. We apply the probabilistic extension of traditional PCA [12] to model the shape variations based on 100 3D faces with about 8900 vertexes.

$$S = U \cdot s + \overline{S} + \varepsilon,\ \varepsilon \sim N(0, \sigma_{3d}^2 I_{3L}),\ \sigma_{_{3d}}^2 = \sum\nolimits_{i=l+1}^{3L} \lambda_i/3L \tag{1}$$

where the columns of U are the most significant eigenvectors and l is the number of eigenvectors, $\overline{S}$ is the average shape of samples and s is the shape parameter to be estimated. ε denotes the isotropic noise in the shape space and σ_{3d} is the standard deviation.

2.3 3D Reconstruction Problem Formulations

The input is the multi-view face alignment result as described in subsection 2.1, denoted as s_{2d}, and the object is to reconstruct the personalized 3D face model. Their relationship can be formulated as:

$$s_{2d} = PfRS + t + \eta, \quad \eta \sim N(0, \sigma_{2d}^2 I_{2L_0}) \tag{2}$$

where η denotes an isotropic observation noise in the image space; σ_{2d} is the standard deviation, which is dynamically decided according to the variation of the shape in each step; $P = P_{2L_0 \times 3L} = (I_{L_0}, 0)_{L_0 \times L} \otimes P_0$ is the projection matrix with $P_0 = \begin{bmatrix} 1\,0\,0 \\ 0\,1\,0 \end{bmatrix}$ and $\otimes$is the *Kronecker product*; f is the scale parameter; $R = R_{3L \times 3L} = I_L \otimes R_0$ is the rotation matrix and $t = 1_{L_0} \otimes t_0 = 1_{L_0} \otimes (t_x, t_y)'$ is the *translation parameter*. Denote c as the pose parameters $\{\alpha, \beta, \gamma, f, t_x, t_y\}$.

2.4 Parameter Estimation

It is difficult to infer the shape parameter s and pose parameter c from the given 2D shape s_{2d} directly. With the hidden data S, the EM algorithm can be applied to conduct parameter optimization. Define the Q-function as:

$$Q(s, c, s^{old}, c^{old}) = E\left[\ln P(s, c|s_{2d}, S)|s_{2d}, s^{old}, c^{old}\right] = \int \ln P(s, c|s_{2d}, S) \cdot P(S|s_{2d}, s^{old}, c^{old}) dS \quad (3)$$

E-Step: With simple computation from Eqn (1) and (2), we have

$$\begin{aligned} -2\ln P(s, c|s_{2d}, S) &= \frac{1}{\sigma_{3d}^2} \parallel S - U \cdot s - \overline{S} \parallel^2 + s' \Lambda^{-1} s \\ &+ \frac{1}{\sigma_{2d}^2} \parallel s_{2d} - PfRS - t \parallel^2 + c_1 \\ -2\ln P(S|s_{2d}, s^{old}, c^{old}) &= \frac{1}{\sigma_{3d}^2} \parallel S - U \cdot s^{old} - \overline{S} \parallel^2 \\ &+ \frac{1}{\sigma_{2d}^2} \parallel s_{2d} - MS - t \parallel^2 + c_2 \end{aligned} \quad (4)$$

where c_1, c_2 are constants and Λ is a diagonal matrix with diagonal elements as leading eigenvalues. The conditional probability $P(S|s_{2d}, s^{old}, c^{old})$ obeys the following Gaussian distribution:

$$P(S|s_{2d}, s^{old}, c^{old}) \sim N(\mu, \Sigma) \quad (5)$$

where $(M = Pf^{old}R^{old})$

$$\mu = \langle S \rangle = (\sigma_{3d}^{-2} I + \sigma_{2d}^{-2} M'M)^{-1} \cdot \left[\sigma_{3d}^{-2}(U \cdot s^{old} + \overline{S}) + \sigma_{2d}^{-2} M'(s_{2d} - t^{old})\right] \quad (6)$$

$$\Sigma = (\sigma_{3d}^{-2} I + \sigma_{2d}^{-2} M'M)^{-1} \quad (7)$$

where $\langle S \rangle$ denotes the conditional expectation $E\left[S|s_{2d}, s^{old}, c^{old}\right]$, then we have:

$$\langle SS' \rangle = \Sigma + \langle S \rangle \langle S' \rangle \quad (8)$$

On the other hand, Σ is the inversion of a very large matrix, which is computed expensively. In fact, M has simple form with $M_0 = P_0 f R_0$ being a 2×3 matrix.

$$M = (I_{L_0}, 0)_{L_0 \times L} \otimes M_0 \quad (9)$$

Then,

$$\Sigma = \begin{pmatrix} I_{L_0} & 0 \\ 0 & 0 \end{pmatrix}_L \otimes (\sigma_{3d}^{-2} I_3 + \sigma_{2d}^{-2} M_0' M_0)^{-1} + \begin{pmatrix} 0 & 0 \\ 0 & I_{L-L_0} \end{pmatrix}_L \otimes \sigma_{3d}^2 I_3 \quad (10)$$

which is much more simple and we only need to compute the inversion of a 3×3 matrix. With the Eqn (5)-(10), the problem is equal to

$$\min_{s,c} \left\langle \frac{1}{\sigma_{3d}^2} \parallel S - U \cdot s - \overline{S} \parallel^2 + s' \Lambda^{-1} s + \frac{1}{\sigma_{2d}^2} \parallel s_{2d} - PfRS - t \parallel^2 \right\rangle \quad (11)$$

***M-Step*:** Notice that pose parameter c is independent to the shape parameter s. Thus they can be optimized separately.

1) ***Optimize shape parameter*** s: shape parameter s can be easily derived by setting the derivative of the Q-function to zero:

$$s = \Lambda(\Lambda + \sigma_{3d}^2 I)^{-1} U'(\langle S \rangle - \overline{S}) \tag{12}$$

2) ***Semi-closed-form solution for pose parameter*** c ***using Quaternion***: From (11),

$$c = arg\,\underset{c}{max}\,Q(s, c, s^{old}, c^{old}) = arg\,\underset{c}{min} \sum_{i=1}^{L_0} \langle \| \, s_{2d}^i - M_0 S^i \, \|^2 \rangle \tag{13}$$

where s_{2d}^i denotes the ith point of s_{2d}, and S^i denotes the correspondent point in S. It's a nonlinear optimization problem and can not be optimized directly. Traditionally, unit quaternion [1][2] based pose representation was applied to solve 3D-to-3D pose parameter variation problem. In the following, we will introduce a semi-closed-from algorithm in terms of unit quaternion for pose estimation.

A quaternion is represented as $\overset{o}{q} = q_0 + q_x i + q_y j + q_z k$, its complex conjugate is defined as $\overset{o}{q}{}^* = q_0 - q_x i - q_y j - q_z k$ and $\mathrm{S}\{\overset{o}{q}\} = (q_x, q_y, q_z)'$. A 3D point p is represented by the purely imaginary quaternion $\overset{o}{p} = 0 + p_x i + p_y j + p_z k$ and a rotation of p is defined as $\overset{o}{q}\,\overset{o}{p}\,\overset{o}{q}{}^*$, then $f = \overset{o}{q} \,.\, \overset{o}{q}{}^*$ and $f\,Rp = \mathrm{S}\{\overset{o}{q} \,.\, \overset{o}{p} \,.\, \overset{o}{q}{}^*\}$. The detailed relation between rotation matrix R_0, scale parameter f and quaternion $\overset{o}{q}$ is referred to [2]. With quaternion representation, the objective function in Eqn (13) can be rewritten as:

$$\min E^2 = < \sum_{i=1}^{n} (\tilde{s}_{2d}^i - \mathcal{S}\{\overset{o}{q}\,\overset{o}{S}{}^i\,\overset{o}{q}{}^*\} - t)' W_i (\tilde{s}_{2d}^i - \mathcal{S}\{\overset{o}{q}\,\overset{o}{S}{}^i\,\overset{o}{q}{}^*\} - t) > \tag{14}$$

where 3D point $\tilde{s}_{2d}^i$ is extended from s_{2d}^i with z-value being zero and W_i represents the directional constraint of the i-th point $W_i = \begin{pmatrix} 1\,0\,0 \\ 0\,1\,0 \\ 0\,0\,0 \end{pmatrix}$ here.

Assume that we have some estimation of $\overset{o}{q}$ available at the r-th iteration as $\overset{o}{q}_r$, a new estimation $\overset{o}{q}_{r+1} = \overset{o}{q}_r + \overset{o}{\delta}$, then

$$\mathcal{S}\{\overset{o}{q}_{r+1}\,\overset{o}{S}{}^i\,\overset{o}{q}{}^*_{r+1}\} = \mathcal{S}\{\overset{o}{q}_r\,\overset{o}{S}{}^i\,\overset{o}{q}{}^*_r + \overset{o}{\delta}\,\overset{o}{S}{}^i\,\overset{o}{q}{}^*_r + \overset{o}{q}_r\,\overset{o}{S}{}^i\,\overset{o}{\delta}{}^* + \overset{o}{\delta}\,\overset{o}{S}{}^i\,\overset{o}{\delta}{}^*\} \tag{15}$$

Assume $\overset{o}{\delta}$ is small with respect to $\overset{o}{q}_r$, then Eqn (15) can be approximated as

$$\mathcal{S}\{\overset{o}{q}_r\,\overset{o}{S}{}^i\,\overset{o}{q}{}^*_r + \overset{o}{\delta}\,\overset{o}{S}{}^i\,\overset{o}{q}{}^*_r + \overset{o}{q}_r\,\overset{o}{S}{}^i\,\overset{o}{\delta}{}^*\} = f_r R_r\,\overset{o}{S}{}^i + G_i\,\overset{o}{\delta} \tag{16}$$

where G_i can be derived from the definition.

Denote $v=(q_0,q_x,q_y,q_z,t_x,t_y)'$, $z_i=\tilde{s}^i_{2d}-f_rR_rS^i$ and $G_{vi}=(G_i\ ,(I_{2\times 2},0)')$, we have:

$$\min E^2 =< \sum_{i=1}^{n}(z_i-G_{vi}v)'W_i(z_i-G_{vi}v) > \tag{17}$$

The optimal solution can be obtained by solving the following traditional function:

$$\sum_{i=1}^{n}\langle g'_{ij}W_i\sum_{k=1}^{6}g_{ik}v_k\rangle=\sum_{i=1}^{n}\langle g'_{ij}W_iz_i\rangle\ (1\le j\le 6) \tag{18}$$

where g_{ij} is the j-th column of matrix G_{vi}. G_{vi} is a linear function of S^i, so are g_{ij} and z_i. Therefore both sides of Eqn (18) which are quadratic functions of S^i at most can be directly computed from Eqn (8).

2.5 Dynamic Correspondence Strategy and Weighted Optimization

Hu's work [3] assumed that the correspondences between the contour points and the 3D face model vertexes are known and fixed, which is inappropriate in the case of out-of-plane rotation. Here we assume that the eyes, mouth, and nose points can be matched accurately from 2D to 3D. For the contour points, the absolute value of z coordinate of the normal direction is small. We utilize the information for the contour points and search for more "proper" points to replace the original contour points after iteration, which results in a more precise correspondence between the contour points of 2D image and 3D face vertexes. The comparison between dynamic correspondence and static correspondence is shown in Fig. 1.

Moreover, there will be part of face occluded in a side-view face image. Thus for the occluded points, the location precision will be degraded. We set the

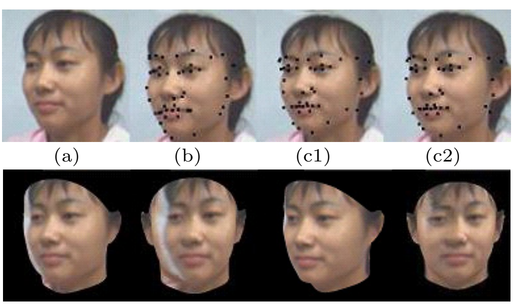

Fig. 1. Comparison between dynamic correspondence and static correspondence. (a) input image; (b) 2D alignment; (c1) 3D geometry reconstruction with static correspondence(black points are the corresponding feature points in 3D model matching the feature points in 2D image in (b).); (c2) 3D geometry reconstruction with dynamic correspondence; (d1)(e1) two views of 3D model with static correspondence; (d2)(e2) two views of 3D model with dynamic correspondence.

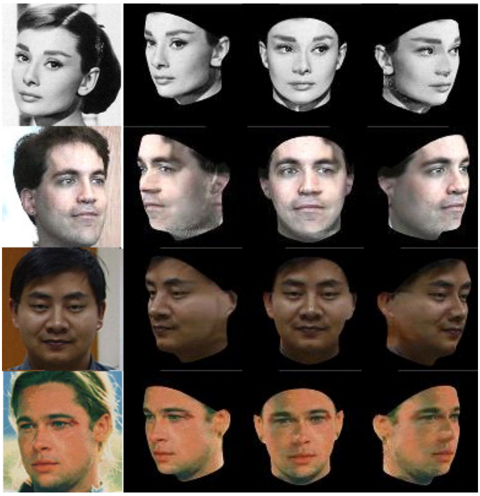

Fig. 2. Comparison of the original images with three different views of reconstructed models of various people.

direction constraint W_i of the contour points dynamically, which improves the final result. After the 3D face geometry is reconstructed, the 2D image is mapped to the 3D geometry to generate the texture. Mostly, there are some vertices occluded in the 3D surface; the “mirror” and “interpolate” strategies are applied to improve the reality.

3 Experiments

We constructed a fully automatic 3D face synthesis system based on the proposed algorithm. Our system is fully automatic. The only input is one face image in arbitrary view and there is no user interaction in the whole process.

In our experiments, we used face images with various poses to automatically construct the personalized 3D faces. Fig. 2 shows some experimental results. It shows that our algorithm can reconstruct the 3D face models for different persons in different views; and the generated virtual faces in different views indicate the realistic of the reconstructed 3D model. The faces in the original images in Fig. 2 are in different illumination conditions and there are different skin colors too. One can see the effective 3D reconstruction results.

The whole process to construct a head model from a face image costs less than 1.6 seconds on a PC with Pentium(R) IV 2.8 GHz processor, which is about eighty times faster than the 3D face reconstruction processing [10], ten times faster than Zhang [7], and 1.25 times faster than Hu [3]. The time cost in 3D face geometry reconstruction process is about 0.6 second and it is much faster than Vetter’s [10] method.

4 Conclusions and Future Work

We have proposed a novel framework to construct 3D face model from a single face image in arbitrary view. The experiments show the efficiency and effectiveness of our proposed algorithm. Compared with other related works, its highlights are two-folds: 1) it is fully automatic and handles face images in arbitrary view; and 2) the efficiency and robustness are guaranteed via the EM algorithm integrated with the unit quaternion based pose representation, dynamic correspondence strategy and weighed optimization method.

The efficient 3D face reconstruction with an arbitrary view face image has many applications including 3D model based multi-view face recognition, face pose estimation and virtual reality in 3D game. Currently, we are exploring to efficiently reconstruct the personalized 3D face model based on multiple face images in different views and conduct the face recognition in variant poses; moreover, we are also applying pose estimation results to detect and locate the attention area.

References

1. A.Hill, T.F.Cootes, C.J.Taylor. “Active Shape Models and the shape approximation problem.” Image and Vision Computing. 14 (8) Aug. 1996 pp 601-608.
2. B.K.P.Horn. Closed-form solution of absolute orientation using unit quaternions. Journal of the Optical Society of America,4(4):629-642,Apr.1987.
3. Y. X. Hu, D. L. Jiang, S. C. Yan, Lei Zhang, H.J. Zhang. “Automatic 3D Reconstruction for Face Recognition”, In FG2004 Proceedings, pages 843-848, 2004.
4. Kin-Man Lam and Hong Yan, An Analytic-to-Holistic Approach for Face Recognition Based on a Single Frontal View, PAMI98, Vol2, No7, page 673-686.
5. J. P. Lewis. Algorithms for solid noise synthesis. In SIGGRAPH ’89 Conference proceedings, pages 263–270. ACM, 1989.
6. H. Li, S.C. Yan, L.Z. Peng. “Robust Multi-view Face Alignment with Edge Based Texture”, submitted to Journal of Computer Science and Technology, 2004.
7. Liu, Z., Zhang, Z., Jacobs, C. and Cohen, M. (2000). Rapid modeling of animated faces from video, Proc. 3rd International Conference on Visual Computing, Mexico City, pp. 58–67. Also in the special issue of The Journal of Visualization and Computer Animation, Vol.12, 2001.
8. F.I. Parke. Computer generated animation of faces. In ACM National Conference. ACM, November 1972.
9. F.I. Parke. A Parametric Model of Human Faces. PhD thesis, University of Utah, Salt Lake City, 1974.
10. S. Romdhani, V. Blanz, and T. Vetter. Face identification by fitting a 3d morphable model using linear shape and texture error functions. In Computer Vision – ECCV’02, volume 4, pages 3-19, 2002.
11. N.Magneneat-Thalmann, H.Minh,M. Angelis, and D. Thalmann. Design, transformation and animation of human faces. Visual Computer, 5:32–39, 1989.
12. M. Tipping and C. Bishop. “Probabilistic principal component analysis” Technical Report Technical Report NCRG/97/010, Neural Computing Research Group, Aston University, Birmingham, UK, September 1997.
13. Ruo Zhang, Ping-Sing Tai, James Edwin Cryer, Mubarak Sha, Shape From Shading: A Survey, IEEE Trans. On PAMI, 21(8). pp690-706. 1999