MVA2005 IAPR Conference on Machine VIsion Applications, May 16-18, 2005 Tsukuba Science City, Japan
13-6

Automatic 3D Face Modeling with Single Image and Perspective Projection Model

Haibing Ren
CASIA-SAIT Joint Lab
95 Zhongguangcun East Road, Beijing, P.R. China
haibing.ren@samsung.com

Kyung-ah Sohn, Seokcheol Kee
Samsung Advance Institute of Technology
Korea
kyungah.sohn@samsung.com
sckee@samsung.com

Abstract

In this paper, a precise and robust algorithm is presented to automatically model the 3D face with only single frontal face image. For shape and projection parameters estimation, an energy function is built to represent the difference between the approximate observed face 3D shape and estimated face 3D shape. To achieve more robust results, shape constraint is added to the energy function. For precision and reality, the perspective projection model is adopted in both 3D shape and 2D texture estimation. The experimental results show that the 3D face modeling is very robust and precise. And it could be applied to the most popular desktop camera systems and mobile-phone camera systems.

1. Introduction

In 2D face recognition, there are two most critical problems limiting the performance of automatic face recognition: variable pose and illumination. It is increasingly clear that the complete solution to these problems is 3D face modeling and recognition. 3D face modeling is the key and pre-determinative factor to 3D face recognition. 3D face modeling is also very important in 3D game and entertainment. Therefore, this paper focuses on 3D face modeling rather than 3D face recognition.

To get a robust 3D face model, some systems relied on specific machines to capture multi-images or some specific images for 3D face modeling. Moghaddam [1] used multi-camera rig to capture 11 face silhouettes of different views to model face 3D shape. Park [2] needed 2 orthogonal-aligned cameras to capture face frontal and profile views. A4 vision [3] made use of inconvenient structured near-infra light and Minolta vivid910 [4] made use of the expensive laser scanner. Though the algorithm in [5] could model human 3D face with single image, manually labeled facial feature points are needed for parameter initialization, such as eye corners, nose tip and mouth corners. And the algorithm is not robust enough for real application. Shape from shading [6] is an algorithm to model 3D human face automatically and with single image. But the result is much far away from satisfaction.

For simplicity, nearly all the current systems [1,2,5,6] utilized the weak perspective projection model to project face 3D vertices on 2D image plane. Though the weak perspective projection model contains less rending parameters and is easy to use, it is just approximation of the perspective projection model on the condition that the face is far away from the optical center. The condition is not satisfied for the desktop camera and mobile-phone camera systems which are most popular application environments now.

In this paper, a precise and robust algorithm is proposed to automatically model the 3D face from the single frontal face image. The 3D face includes 2 parts: 3D shape and 2D texture. Both parts adopt perspective projection models to achieve realistic and precise 3D face.

Firstly the face 3D shape and projection parameters should be estimated at the same time. With automatically detected ASM (Active Shape Model, ASM) feature points [7] and perspective projection model, approximate observed 3D shape can be reconstructed. Given the 3D shape PCA (Principle Component Analysis, PCA) model, the shape parameter can be represented as the PCA coefficients, and the estimated face 3D shape could be synthesized with these coefficients. An energy function is built to represent the difference between the observed and estimated 3D face shapes. To make the algorithm more robust, shape constraint is added to the energy function. Robust precise shape and projection parameters can be quickly estimated via Newton algorithm with step size adaptation.

Once face shape and projection parameters are reconstructed, texture information of the model can be estimated using back projection.

The experimental results show the estimated 3D face is much more robust than the one via the energy function without shape constraint and more precise than the one with the weak perspective projection model.

The rest of the paper is arranged as following: the section 2 gives the face 3D shape representation; the section 3 gives the perspective projection model for 3D face modeling; the shape and projection parameters estimation are presented in section 4. And the last is the experimental result and conclusion.

2. Shape Representation

In geometry, each face 3D shape S has k vertices $\{P_i(X_i, Y_i, Z_i) \mid 1 \le i \le k\}$. Therefore, S can be represented as

$$S = (X_1, Y_1, Z_1, X_2, \cdots, Z_k) \in R^{3k} \quad (1)$$

where X_i, Y_i, Z_i are the X,Y and Z coordinates of the ith vertex in the face coordinate system. Given the 3D shape

PCA model, the face shape S could be represent as:

$$S = \overline{S} + PC * (\alpha \bullet Sigma) \quad (2)$$

Where $\overline{S}$ is the mean shape, $PC = [PC_1, PC_1, \ldots, PC_m]$ is the eigenvector array (PC_i is the ith eigenvector, m is the eigenvector number), $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_m)$ is the shape PCA parameter and $Sigma$ is a weight vector for the shape PCA parameter.

Because both the face image and 3D face model are frontal, there is only a translation T between the face and the camera coordinates systems. In the camera coordinate, the face 3D shape S' is

$$S' = S - T \quad (3)$$

In this paper, only the vertices corresponding to n ASM feature points (as the Figure 1) are considered for parameter estimation. The parts of $S', S, \overline{S}$, and PC that only contains the vertices corresponding to ASM feature points also satisfy the per-mentioned equations(1-3). Only the dimension k in $S', S, \overline{S}$ and PC needs to be changed to n. Therefore, the ASM features could be used to estimate the PCA parameter. In the following part of this paper, $S', S, \overline{S}$ and PC only contain the vertices corresponding to ASM features points.

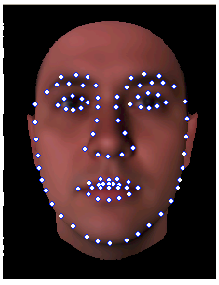


Figure 1. ASM feature points.

3. Projection Model for Shape Estimation

As the given image is generated with the perspective projection model, there exists some error for modeling 3D face with weak perspective projection model. In this paper, perspective projection model is utilized to calculate image projection of 3D vertices. It is much more precise than the weak perspective projection model.

The weak perspective projection model is the approximation of perspective projection model, and the approximation condition is not satisfied in the most popular desk camera and mobile-phone camera systems. Generally, the 3D face model reconstructed with the weak perspective projection model is thinner than the real one. The reason could be explained by Figure 2.

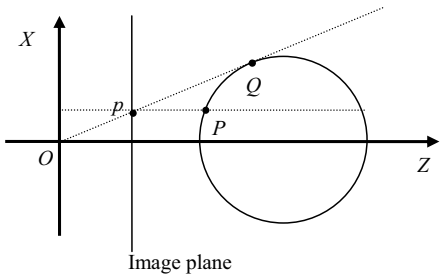


Figure 2. Platform of the face imaging system.

Figure 2 is the platform of the human face imaging system. O is the camera optical center, OX and OZ are the axis of the camera coordinate system, the circle represents the human face that is facing to the camera, p is a ASM feature point. With the weak perspective projection model, p is the projection of P while with the perspective projection, p is the projection of Q.

As the given image is achieved with perspective projection model of the camera system, Q is the real 3D vertex corresponding to p. In 3D face modeling with weak perspective projection model, P is considered to be Q. Therefore, face width would be decreased. The problem doesn't exist in the height of face 3D model because most ASM feature points lie on the outer contour of X direction.

Given any point $P_i(X_i, Y_i, Z_i)$ $(1 \le i \le n)$ in the face coordinate system, its 2D projection (x_i, y_i) via perspective projection model is:

$$\begin{cases} x_i = f\dfrac{X_i - \Delta X}{Z_i - \Delta Z} + \Delta x \\ y_i = f\dfrac{Y_i - \Delta Y}{Z_i - \Delta Z} + \Delta y \end{cases} \quad (4)$$

where f is the camera focus length, $(\Delta X, \Delta Y, \Delta Z)$ are the 3 coefficients of 3D translation T, the $\Delta p = (\Delta x, \Delta y)$ is the image projection offset. Therefore, the unknown parameters include:

- The shape parameter: α
- The 3D translation between the 3D face and camera coordinate systems: $T = (\Delta X, \Delta Y, \Delta Z)$
- Camera focus length: f
- The image projection offset: $\Delta p = (\Delta x, \Delta y)$

The first one is shape parameter and the other three are projection parameters.

4. Shape and Projection Parameters Estimation

Shape and projection parameters should be estimated at the same time. In this section, an energy function is built, including two parts: the observation energy and shape constraint. The observation energy represents the difference between the approximate observed face 3D shape and estimated face 3D shape. The shape constraint represents how much the estimated 3D shape resembles human face and it is very important to get the robust result. If there is no shape constraint, minor changes in ASM feature points may lead to major changes in shape parameters.

After parameter initialization, the Newton algorithm with step size adaptation is used to estimate the precise parameters iteratively.

4.1. Parameters initialization

Shape parameters are initialized to be $\alpha^0 = (0,0,\cdots,0)$. $(\Delta x, \Delta y)$ are the projection coordinates of the optical center and generally it is very close to image center. Therefore,

$(\Delta x, \Delta y)$ is initialized as half of image size. The initialization of 3D translation $T^0 = (\Delta X^0, \Delta Y^0, \Delta Z^0)$ and camera focus length f^0 are very complex. In this paper, ΔX^0 and ΔY^0 are calculated first with the initial face 3D shape, image projection offset and observed ASM feature points. ΔZ^0 and f^0 are calculated later.

✧ ΔX^0 and ΔY^0 calculation

With the initial face 3D shape parameter α^0, the vertices initial 3D coordinate $P_i(X_i^0, Y_i^0, Z_i^0)$ $(1 \le i \le n)$ could be calculated with PCA model. From (2-4), a new equation could be obtained:

$$\begin{cases} \dfrac{x_i - \Delta x^0}{f^0} = \dfrac{X_i^0 - \Delta X^0}{Z_i^0 - \Delta Z^0} \\ \dfrac{y_i - \Delta y^0}{f^0} = \dfrac{Y_i^0 - \Delta Y^0}{Z_i^0 - \Delta Z^0} \end{cases} \tag{5}$$

Via eliminating $Z_i^0 - \Delta Z^0$ and f^0 in (5), (6) could be achieved.

$$\frac{x_i - \Delta x^0}{y_i - \Delta y^0} = \frac{X_i^0 - \Delta X^0}{Y_i^0 - \Delta Y^0} \tag{6}$$

With the integration of *n* ASM feature points, (7) could be got.

$$A\begin{bmatrix} \Delta X^0 \\ \Delta Y^0 \end{bmatrix} = B \tag{7}$$

$$A = \begin{bmatrix} 1 & \dfrac{x_1 - \Delta x^0}{y_1 - \Delta y^0} \\ \vdots & \vdots \\ 1 & \dfrac{x_n - \Delta x^0}{y_n - \Delta y^0} \end{bmatrix} \quad B = \begin{bmatrix} X_1^0 + \dfrac{x_1 - \Delta x^0}{y_1 - \Delta y^0} Y_1^0 \\ \vdots \\ X_n^0 + \dfrac{x_n - \Delta x^0}{y_n - \Delta y^0} Y_n^0 \end{bmatrix}$$

Therefore, ΔX^0 and ΔY^0 can be calculated as:

$$\begin{bmatrix} \Delta X^0 \\ \Delta Y^0 \end{bmatrix} = (A^T A)^{-1} A^T B \tag{8}$$

◆ ΔZ^0 and f^0 calculation

With the integration of (2-4), (9) could be got:

$$C\begin{bmatrix} \Delta Z^0 \\ f^0 \end{bmatrix} = D \tag{9}$$

$$C = \begin{bmatrix} x_1 - \Delta x & X_1^0 - \Delta X \\ y_1 - \Delta y & Y_1^0 - \Delta Y \\ \vdots & \vdots \\ x_n - \Delta x & X_n^0 - \Delta X \\ y_n - \Delta y & Y_n^0 - \Delta Y \end{bmatrix} \quad D = \begin{bmatrix} (x_1 - \Delta x) Z_1^0 \\ (y_1 - \Delta y) Z_1^0 \\ \vdots \\ (x_n - \Delta x) Z_n^0 \\ (y_n - \Delta y) Z_n^0 \end{bmatrix}$$

Therefore, ΔZ^0 and f^0 can be calculated as:

$$\begin{bmatrix} \Delta Z^0 \\ f^0 \end{bmatrix} = (C^T C)^{-1} C^T D \tag{10}$$

4.2. Energy function

An energy function $F(f, T, \alpha, \Delta p)$ includes the observation energy and the shape constraint. The observation energy represents the difference between the approximate observed 3D shape and the estimated one via shape coefficients. The shape constraint makes the estimated shape more robust and more similar to human face. Via Newton algorithm with step size adaptation, better shape coefficients could be obtained. After some times of iteration, the shape coefficients could converge to the optimal value.

The energy function is:

$$F(f, T, \alpha, \Delta p) = (\tilde{S} - S)^T (\tilde{S} - S) + \lambda \alpha^T \alpha \tag{11}$$

where $\tilde{S}$ is the observed 3D face shape, λ is a scale coefficient, the first part is observation energy and the second part is shape constraint.

✧ Approximate observed face 3D shape

The approximate observed face 3D shape is estimated with 2D projection of the ASM features and the parameters of previous iteration. With ASM feature points, the shape and projection parameters the approximate observed 3D shape $\tilde{S} = (\tilde{X}_1, \tilde{Y}_1, \tilde{Z}_1, \tilde{X}_2, \cdots, \tilde{Z}_n)$ could be calculated as following:

$$\begin{cases} \tilde{X}_i = \dfrac{(x_i - \Delta x^{t-1})(Z_i^{t-1} - \Delta Z^{t-1})}{f^{t-1}} + \Delta X^{t-1} \\ \tilde{Y}_i = -\dfrac{(y_i - \Delta y^{t-1})(Z_i^{t-1} - \Delta Z^{t-1})}{f^{t-1}} + \Delta Y^{t-1} \\ \tilde{Z}_i = Z_i^{t-1} \end{cases} \tag{12}$$

where the right superscript *t*-1 means the (*t*-1)th iteration, $(\Delta X^{t-1}, \Delta Y^{t-1}, \Delta Z^{t-1})$, $(\Delta x^{t-1}, \Delta y^{t-1})$, f^{t-1} and Z_i^{t-1} are the 3D translation, image projection offset, camera focus length and the *i*th vertex Z coordinate of previous iteration, respectively.

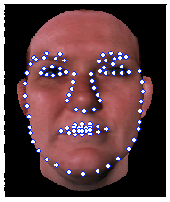
4.3 More Precise Estimation

Sometimes the parameters estimated from a single image are not precise enough. For the same camera, the camera focus length and image projection offset can be considered as const. Via camera calibration, they can be estimated precisely and robustly. Using camera focus length and image projection offset as given, the algorithm would be much more robust and the result will be much more precise.

5. Experimental Results

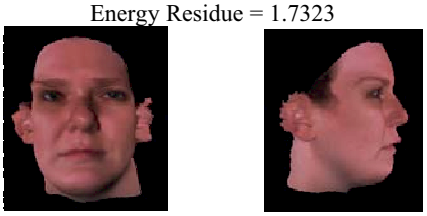
The experimental results show the energy function with shape constraint could achieve much more robust 3D face models than the one without shape constraint. And they also show that the 3D faces with the perspective projection model are more precise than the ones with the weak perspective projection model.

(a) Given image

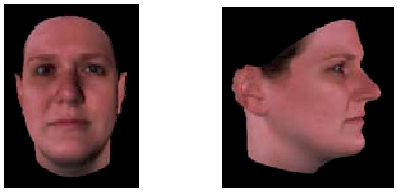


(b) ASM feature points

(c) 3D face model via the algorithm of this paper
Energy Residue = 1.7323



(d) 3D face model via the energy function without shape constraint, Energy Residue = 1.6282



(e) 3D face model via weak perspective projection model, Energy Residue = 3.9132

Figure 3. Comparison of different algorithms.

In Figure 3, (a) is a given frontal face image, (b) shows the ASM feature points, (c) is the 3D face model estimated with the perspective projection model and the energy function with shape constraint, (d) is the 3D face via the energy function without shape constraint, (e) is 3D face via the weak perspective projection model.

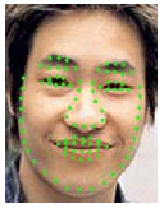
Among these results, the 3D model in (c) is the best. With shape constraint, both the 3D models of (c) and (e) are very good. Without the shape constraint, the 3D model of (d) doesn't look like human face. It shows the shape constraint makes the algorithm very robust. And the 3D face model in (e) is much thinner than the real one. It shows the weak perspective projection model is not precise enough for 3D face modeling in desktop camera system.

When the parameters converge, the energy function with minimal value could be obtained. For comparison of the precision, energy residues are calculated as $\sqrt{\frac{Minimal\quad Energy}{Shape\quad Dimension}}$. As the energy residue of the algorithm with perspective projection model is much less than the algorithm with weak perspective projection model, it shows that the 3D face model with the perspective projection is much more precise in this experiment. The energy residue of the algorithm without shape constraint is a little less than the one with shape constraint because there is no shape constraint item in the former energy function.

The algorithm works very fast and total computation time for Figure 3(c) is 1.156s. The reason lies in the accurate parameter initialization and the step size adaptation in the Newton algorithm.

Figure 4 are another examples achieved via the algorithm of this paper. Though the ASM feature points are not very precise, the 3D face models are very good.

(a) Given image



(b) ASM feature points

(c) 3D face model via the algorithm of this paper
Energy Residue = 1.9572

Figure 4. Another 3D face modeling result.

6. Conclusion

This paper presents an automatic robust and precise algorithm for 3D face modeling. With only frontal face image, the algorithm estimated 3D face shape and texture automatically. In 3D shape estimation, the energy with shape constraint makes the algorithm very robust and the perspective projection model makes the algorithm very precise. The results show that, no matter ASM feature points are precise or not, good 3D face model could be estimated.

As 3D face could be modeled automatically and robustly from single face image, the algorithm could be used in many environments, such as desktop camera systems and mobile-phone camera system.

Reference

[1] Moghaddam, B.; Lee, J.; Pfister, H.; Raghu Machiraju .Model-based 3D face capture with shape-from-silhouettes. IEEE International Workshop on Analysis and Modeling of Faces and Gestures, 17 Oct. 2003, pp.20- 27.

[2] In Kyu Park, Hui Zhang, Vladimir Vezhnevets, etc. Image-based photorealistic 3-D face modeling. The proceeding of sixth international conference on Automatic Face and Gesture Recognition. May 2004, pp.49-54.

[3] A4 vision company, www.a4vision.com/3_accesscontrol.html.

[4] Konica Minolta vivid910, *www.minoltausa.com/vivid/products/vi910-en.asp*

[5] Blanz, V. and Vetter, T. Face recognition based on fitting a 3D morphable model. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol.25 no.9 (2003), pp.1063-1074.

[6] Joseph J. Atick, Paul A. Griffin and A. Norman Redlich. Statistical Approach to Shape from Shading: Reconstruction of Three-Dimensional Face Surfaces from Single Two-Dimensional Images. Neural computing, Vol. 8, Issue 6 - August 15, 1996, pp. 1321-1340.

[7] T. F. Cootes, C. J. Taylor, D. H. Cooper, J. Graham. Active Shape Model – Their Training and Application. Computer Vision and Image Understanding. Vol 61, No 1, 1995, pp. 38-59.