Single Image 3D Face Reconstruction Based on Statistical Model

Yang Zhigang¹, Li Yuguang², Chen Shuangling², Li Bosen², Wu Zhuangzhi¹

¹School of Computer Science and Engineering, Beihang university, Beijing 100191, China
²101 Institute of the Ministry of Civil Affairs, Beijing 100070, China

Abstract: Single image 3D face reconstruction technology has a broad application in the life, also has a high technical difficulties. This paper proposes a single image 3D face reconstruction technology based on statistical model. The method established 3D face statistical model based on 3D face database in advance, and trained 2D facial feature points parameter model based on the 2D face database using the algorithm of SDM. When reconstruct 3D face model from single image, first we use 2D facial feature points parameter model to extract face feature points; Then, according to 3D face model, we use adaptive learning factor gradient descent method to iteratively optimize the energy function, to get statistical model parametric vector, namely got the 3D face model corresponding to the 2D face images. The results show that use the proposed method to rebuild the faces have high similarity.

1. Introduction

3D face reconstruction technique mainly contains 3D face reconstruction technique based on structure light, 3D face reconstruction based on multiple image and 3D reconstruction based on single image. 3D face reconstruction based on single image mainly contains the method based on machine learning and the method based on statistical model. Blanz and Vetter team presents a reconstruction method based on 3D morphable model, first they obtained 3D morphable face model by statistical learning, it including 3D shape model (expressed in about 70000 vertices) and the corresponding 2D texture model, then using Newton iterative method to make the morphable model aimed at specific face image, and then got the 3D model of the human face, this method can obtain accurate 3D shape and texture information, and has been successfully applied to the 3D face recognition [1]. Tran and Hassner using deep neural network build up a system of generating 3D face from single image, this method is more rapid and efficient than Blanz, can be easily applied to real-time system, such as video conference, etc [2].

Statistical model is a kind of powerful graphics research model structure and technology. 2D statistical models have been put forward for a long time, and has been widely used in areas such as object tracking, object recognition, face recognition, face analysis, gesture estimation, etc [3]. 3D statistical model put forward by Blanz and Vetter in 1999, was gradually applied to many research related to image. Compared with the 2D statistical model, the 3D statistical model has many unique advantages. It has depth information, can be easily applied to gesture recognition and illumination simulation, 3D reconstruction, etc.

In order to realize full automatic 3D face reconstruction, we needs automatically calibration facial feature points. Here we use the point calibration technology based on SDM(supervised descent method) algorithm [4]. Newton method is usually used for optimization problems, but it will confronted with...
some problems in solving the questions of graphic images, however SDM algorithm can avoid these problems.

After get the facial feature points, it can based on this set of feature points, adjusting and optimizing statistical model coefficient of each component make up the model close to the set of feature points as far as possible according to certain constraints. In order to get better effect, the process can be repeated iteratively, until satisfied [5].

2. Framework of face reconstruction

Figure 1 depicts the process of 3D face reconstruction based on single image: 3D face statistical model and 2D facial feature points parameter model is established in advance. Then using 2D face feature points parameter model recognize image feature points, and optimize 3D face statistical model to match this set of feature points, eventually get 3D face model corresponding to the face image. When training 2D facial feature points parameter model we used 2000 faces with different age, different gender and differences posture. To accelerate points detect speed and accuracy, we use SDM algorithm instead of traditional Newton method. When build 3D statistical model, 500 laser scanning face models with different age, different gender and differences posture are used. In order to balance the speed of generating face model with accuracy, adaptive learning factor gradient descent method is used instead of traditional gradient descent method.

3 Key technology

3.1 Build statistical model

Accurate facial model is the foundation of building 3D statistical model, a high precision of 3D model is obtained by laser scanning, which is suitable for this application scenario. First assume that these facial model are registered, then the face model can be represented by the serialization of point set \( S_{reg} = (X_1, Y_1, Z_1, X_2, Y_2, Z_2, ..., X_n, Y_n, Z_n)^T \) distributed in the curved surface of the face, \((X_n, Y_n, Z_n)\) represent n-th point. The final 3D statistical model \( S_{mm} \) represent by linear combination of m registered faces \( S_{mm} = \sum_{i=1}^{m} a_i S_{reg, i} \), where \( \sum_{i=1}^{m} a_i = 1 \). Any 3D face model \( S_{mod} \) can be write in
the form of \( S_{nm}(\vec{a}) \), which means facial model set \( \{S_{reg,i}\} \) parameterized by vector \( \vec{a} = (a_1, a_2, ..., a_m)^T \). It means any face model can be generated by adjust the parameterization vector.

To get the registration of 3D face model mentioned above, we use optical flow method here. In fact, 3D model \( I(h, \phi) \) obtained by laser scanning are generally not registered, this original face model cannot be directly used to serialization, so it need to be registered by optical flow method. Optical flow method calculate the flow field \( (\delta h(h, \phi), \delta \phi(h, \phi)) \) of the minimum norm \( ||I_1(h, \phi) - I_2(h + \delta h, \phi + \delta \phi)|| \), to determine the corresponding points of the two 3D model, so as to realize the model registration.

The 3D face model \( S_{reg} = (X_1, Y_1, Z_1, X_2, Y_2, Z_2, ..., X_n, Y_n, Z_n)^T \), which is registered and serialized is a \( 3n \)-dimension vector. When 3D face model has so much vertex, it will consume so much time to calculate the parameterization vector \( \vec{a} = (a_1, a_2, ..., a_n)^T \). To improve efficiency, we adopt PCA method to reduce the dimension of 3D face model, and got PCA 3D face model \( S_{pca} = (X_1, Y_1, Z_1, X_2, Y_2, Z_2, ..., X_{m-1}, Y_{m-1}, Z_{m-1})^T \). After this step, parameterization vector \( \vec{a} = (a_1, a_2, ..., a_{m-1})^T \) can be calculated effectively. Finally generated 3D face model can be represented as \( S_{mod} = \bar{S} + \sum_{i=1}^{m-1} a_iS_{pca,i} \), where \( \bar{S} \) represent average face.

3.2 Detect face feature points
In this paper, using SDM algorithm to detect facial feature points, the process is divided into training and detecting two steps:

(1) Training SDM model
SDM need to get \( R_k \) and \( b_k \) from training set. When training, human face image sets represent as \( \{I_i\} \), \( X_\ast \) represent feature points for manual calibration. Initial feature points is \( X_0 \), so face feature points location becomes a linear regression problem, the goal of regression problems is calculate the iteration step length from \( X_0 \) to \( X_\ast \), the input characteristics of the regression problems is the SIFT features \( \phi_0 \) in place \( X_0 \). So, to paraphrase linear regression problems can get objective function:

\[
\arg\min_{R_0, b_0} \sum_{d_i} \|\Delta x^i - R_0 \phi^i_0 - b_0\|^2
\]

So \( R_0 \) and \( b_0 \) can be get from training set. this objective function described how to get iteration coefficient for the first iteration step. k-th iteration step coefficient \( R_k \) and \( b_k \) can be get similarly.

\[
\arg\min_{R_k, b_k} \sum_{d_i} \|\Delta x^i - R_k \phi^i_k - b_k\|^2
\]

(2) Detect feature points
\( d(x) \in \mathbb{R}^{p+1} \) represent p feature points coordinates in the image, \( h \) is nonlinear feature extraction function on each feature point, which is SIFT feature here. SIFT feature is a 128-dimension vector, so \( h(d(x)) \in \mathbb{R}^{128} \). \( X_\ast \) represent feature points for manual calibration, the SIFT feature of them represent as \( \phi_\ast = h(d(x)) \). So face feature detection can be equal to find \( \Delta x \) to minimize the energy of the following function:

\[
f(x_0 + \Delta x) = \|h(d(x_0 + \Delta x)) - \phi_\ast\|^2
\]

We use SDM algorithm in place of the traditional newton method to iterate calculate from \( X_0 \) to \( X_\ast \).

\[
X_k = X_{k-1} + R_{k-1} \phi_{k-1} + b_{k-1}
\]

Here \( R_k \) and \( b_k \) are getting from SDM model trained in above step. Because \( R_k \) and \( b_k \) is calculated in advance, thus it can avoiding the problem of inefficient by using Newton method, and it is calculated by regression method, so it can avoids the problem of non-positive Hessian matrix and the secondary non-differentiable objective function problem.

3.3 Generate 3D face model
After face feature points detected, real 3D face model can be generated by adjust the parametric vector
of morphable model under the constraints of these feature points. Let $S_{mm}$ represent as 3D morphable model, $\{P_i\}$ represents r face feature points. In order to get a 3D face model in accordance with the 2D face image, we need let the feature points in 3D face model close to the feature points $\{P_i\}$ in 2D face image as far as possible after affine transformation $\bar{\rho}$.

In order to get the feature points in 3D face model affined to 2D plane, affine transformation $\bar{\rho}$ and parametric vector of morphable model is needed. Here we get them from estimation of operator, then the 2D image affined from 3D model can be represent as $I_{mod} = P(S_{mm}, \bar{\rho}, \bar{a})$.

In order to measure the difference between generated face and real face according to the 2D image, we define energy function $E = \|I_{input} - I_{mod}\|^2$ to measure the gap of them. The formula use sum of Euclidean distance squares between 2D face feature point set and rendering image from 3D model feature point set to represent the gap of them, the smaller gap is, the more realistic is. In order to calculate the minimum deviation function, gradient descent method is used to adjust the parameterized vector $a' = a_j - \lambda_j \frac{\partial E}{\partial a_j}$. We assumed that the camera parameters are estimated correctly, so keep $\bar{\rho}$ the same all the time [8].

$\lambda$ is speed factor in above formula, it reflects the speed of optimization process, the greater the value is, the faster the speed is, it is set by experience generally. But if the value is too small, it will lead to optimization process too long, also easy to fall into local extreme value. If the value is too large, it would be impossible to get accurate model, it can also cause energy function not convergence. In order to avoid the above situation, need to dynamically adjust speed factor. Here, we use the following strategies:

1. Initially a larger speed factor is set by experience;
2. When gradient direction changed, let $\lambda' = k \lambda$, $k$ is the proportion of each step, generally take $0.7 \sim 0.8$.

4. Results

In this paper, 500 3D laser scanning face models of different age, different gender and differences posture are used to build the 3D face statistical model. And 2000 2D face image are used to train face feature parametric model. Finally use SDM model to detect the feature points of the 2D face image, and generate realistic 3D face model with constraints of these feature points. As can be seen from the results, the generated 3D face model is very similar to the real face, and it has a lot of detail characteristics. It can be applied to public security forensic, film and television animation, security certification, video conference, etc. Figure 2 is the result of this method.
5. Future work
This paper proposed a method to reconstruct a 3D face model from a single image based on a statistical model, which is established based on a 3D face database in advance. Training of 2D facial feature points parameter model is done using the algorithm of SDM. When reconstructing a 3D face model from a single image, the 2D facial feature points parameter model is used to extract face feature points. Then, according to the 3D face model, an adaptive learning factor gradient descent method is used to iteratively optimize the energy function to get the statistical model parametric vector, obtaining the 3D face model corresponding to the 2D face image. The results of the 3D face model have a high similarity to the real face, so they meet the demands of most real-life applications.

But the information provided by a single image is limited, especially in cases where depth information loss is serious, making the method not suitable for fields requiring high precision, such as medical and cosmetic industries.

The improvement direction in the future: 1) When constructing the statistical model, this article aimed to build the whole face statistical model, which can be difficult to adjust for local facial models. To solve this problem, we should split the face according to the organ, building a statistical model for each part respectively. 2) When constructing the statistical model, only spatial information is used, but if we can incorporate texture information, the generated face effect will be better.

References
[1] Blanz V, Vetter T. A morphable model for the synthesis of 3D faces[J]. In Computer Graphics Proceedings, SIGGRAPH’99, 1999:187-194.
[2] Tran A T, Hassner T, Masi I, et al. Regressing Robust and Discriminative 3D Morphable Models with a very Deep Neural Network[J]. 2016.
[3] Min F, Sang N, Wang Z. Automatic Face Replacement in Video Based on 2D Morphable Model[J]. 2010:2250-2253.
[4] Xiong X, Torre F D L. Supervised Descent Method and Its Applications to Face Alignment[C]// Computer Vision and Pattern Recognition. IEEE, 2013:532-539.
[5] Huber P, Feng Z H, Christmas W, et al. Fitting 3D Morphable Models using Local Features[J].
2015:1195-1199.

[6] Akimoto T, Suenaga Y, Wallace R S. Automatic creation of 3D facial models[J]. IEEE Computer Graphics & Applications, 1993, 13(5): 16-22.

[7] Jiang D, Hu Y, Yan S, et al. Efficient 3D reconstruction for face recognition[J]. Pattern Recognition, 2004, 38(6): 787-798.

[8] Chang K J, Flynn P J. Effects on facial expression in 3D face recognition[J]. Proceedings of SPIE - The International Society for Optical Engineering, 2005, 5779: 132-143.