FACIAL FEATURE EXTRACTION FROM RANGE IMAGES USING A 3D MORPHABLE MODEL

Le Zou[†], *Samuel Cheng*[§1] , *Zixiang Xiong*[†], *Mi Lu*[†] *and Kenneth Castleman*[‡]
[†]Department of Electrical & Computer Engineering, Texas A&M University
[‡]Advanced Digital Imaging Research, LLC., 2450 South Shore Blvd., Suite 305, League City, TX 77573
[§]School of Electrical & Computer Engineering, University of Oklahoma-Tulsa

Abstract — In this paper, a novel scheme is introduced for human facial feature extraction. Unlike previous methods that fit a 3D morphable model to 2D intensity images, our scheme utilizes 3D range images to extract features without requiring manually-defined initial landmark points. A linear transformation is used to achieve the mapping between the 3D model and a 3D range image, which makes the computation simple and fast. Moreover, our scheme is robust to the illumination and pose variations. In addition to features from range images, extra features can be obtained by examining optional 2D texture images. Using our scheme, we can also perform automatic eye/mouth corner localization. Experimental results show the high accuracy and robustness of our scheme.

Index Terms — Face recognition, Computer vision, Geometric modeling, Feature extraction

1. INTRODUCTION

Due to tighten homeland security in the US, face recognition has attracted increasing interests recently. In a face recognition system, facial feature extraction is the most significant component. The goal is to automatically extract features from face images with accuracy and robustness. Currently there are two approaches: one relies on 2D texture images; another utilizes 3D range images.

The first approach is plagued by problems due to viewpoint and lighting variations, which make it difficult to extract facial features accurately without human assistance. For example, although methods based on the 3D morphable model [1, 2] can handle illumination and viewpoint variations, they rely on manually-defined landmark points to fit the 3D model to 2D intensity images.

The second approach utilizes depth information of 3D range images to extract features. Since 3D range images are invariant to illumination changes, the impact of lighting variations is moot in this approach. Furthermore, 3D faces in range images can be rotated and shifted to a standard pose to overcome the problem caused by viewpoint variations. Therefore, this approach becomes increasingly attractive nowadays. Motivated by early research works [3, 4] that began about a decade ago, different feature extraction techniques [5, 6, 7, 8, 9] have been proposed by using wavelet-signature [5], curvature [9] and rigid surface matching [6, 7, 8] techniques. These techniques, however, have the problem of being sensitive to facial expression variations. While this problem was not addressed in [5, 6, 9], a partial solution of the problem was addressed in [7, 8] by using rigid surface matching.

In this paper we propose a method based on 3D morphable model to extract facial features from range images. In our scheme, a range image from a 3D image acquisition system is normalized before being used as an input. Then a synthesized 3D face is generated by minimizing the range difference between the input range image and the 3D morphable model. After this fitting procedure, features can be obtained from the shape coefficients of the newly synthesized 3D face. In addition, an extra optimization step can be performed to extract extra features from an optional texture image, which is generated from the 3D image acquisition system together with the input range image.

Compared to the 3D morphable model methods in [1, 2], which use 2D texture images as inputs, our method has several advantages. First, initialization based on manually-defined landmark points is not needed. In order to fit the 3D morphable model to a 2D texture face image, some key information (e.g., locations of eye and mouth corners) of the human face has to be given as initialization. Because a texture image contains a mixture of shape and color information of a human face, it is difficult to extract the shape information accurately from the texture image without human assistance. In [1, 2], this important information is extracted manually. In our method, since input range images contain only the shape information, this key information can be obtained automatically during the fitting procedure. Therefore, manual initialization is not needed in our method. Second, lighting variations, which typically lower the accuracy of feature extraction, do not present a problem in our method, because range images are robust to illumination changes. Third, a linear transformation is used to map the 3D model to an input 3D range image, which makes the computation simple and fast.

Compared to other range image-based feature extraction techniques [5, 6, 7, 8, 9], our method can overcome the difficulty caused by expression variations. This is due to the fact

[1]The author was with Advanced Digital Imaging Research, LLC when this work was done.

that the 3D morphable model [1, 2] contains example faces with facial expressions. This allows us to closely synthesize faces with facial expressions and extract features accurately. In addition, our method can be used to automatically label eye/mouth corners. Experimental results show the effectiveness of our method.

2. ACQUISITION AND NORMALIZATION OF INPUT IMAGES

In our system, the range images, along with the corresponding texture images, are obtained by using a 3D image acquisition system from 3Q Inc., which can generate high resolution 3D surface images in less than 2 milliseconds.

The obtained 3D meshes from the 3D image acquisition system describe 3D faces with different orientations. These 3D "raw" faces cannot be directly used as inputs to our fitting algorithm. In order to make the 3D faces have the same orientation, we rotate and shift each of them until the mean-squared-difference (MSD) between it and a standard face is minimized. After the normalization, the eyes look straight ahead and lie on a line parallel to the x-axis.

3. 3D MORPHABLE MODEL

The 3D morphable face model [1, 2] is constructed based on a vector space representation of 3D faces. In this model, a synthesized 3D face can be represented by a convex combination of n shape and texture vectors $\mathbf{S}_i$ and $\mathbf{T}_i$ ($i = 1, \ldots, n$) of real human faces (example faces). Let $\overline{\mathbf{s}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{S}_i$ and $\overline{\mathbf{t}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{T}_i$, then we can obtain covariance matrices $A_S = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{S}_i - \overline{\mathbf{s}})(\mathbf{S}_i - \overline{\mathbf{s}})^T$ and $A_T = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{T}_i - \overline{\mathbf{t}})(\mathbf{T}_i - \overline{\mathbf{t}})^T$. We further calculate the ith eigenvalue $\sigma_{R,i}^2$ and its corresponding eigenvector $\mathbf{s}_i$ of A_R. Similarly, let $\sigma_{T,i}^2$ and $\mathbf{t}_i$ be the ith eigenvalue and the corresponding eigenvector of A_T. The shape and color of a synthesized face can be represented by a shape vector $\mathbf{s}$ and a texture vector $\mathbf{t}$, respectively, with

$$\mathbf{s} = \overline{\mathbf{s}} + \sum_{i=1}^{n}\alpha_i\mathbf{s}_i,\ \mathbf{t} = \overline{\mathbf{t}} + \sum_{i=1}^{n}\beta_i\mathbf{t}_i, \tag{1}$$

where the distributions of the coefficients $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^T$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)^T$ are:

$$p(\boldsymbol{\alpha}) \propto \exp(-\frac{1}{2}\sum_{i=1}^{n}\frac{\alpha_i^2}{\sigma_{R,i}^2}),\ p(\boldsymbol{\beta}) \propto \exp(-\frac{1}{2}\sum_{i=1}^{n}\frac{\beta_i^2}{\sigma_{T,i}^2}). \tag{2}$$

Given an index i and coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, a point p of the 3D model can be defined by its 3D position $s(i, \boldsymbol{\alpha}) = (x(i, \boldsymbol{\alpha}), y(i, \boldsymbol{\alpha}), z(i, \boldsymbol{\alpha}))$ and its color $t(i, \boldsymbol{\beta}) = (r(i, \boldsymbol{\beta}), g(i, \boldsymbol{\beta}), b(i, \boldsymbol{\beta}))$.

The 3D model is placed in the 3D space with the eyes on a line parallel to the x-axis and looking straight ahead. So the 3D morphable model and the 3D faces in the normalized input range images have the same head pose. As we will see later, this property simplifies the computation of our algorithm.

4. FEATURE EXTRACTION FROM RANGE IMAGES

The key step of our feature extraction method is to fit the 3D morphable model to an input 3D range image. After the fitting procedure, the shape coefficients of the model are optimized and become range features of the human face in the image.

4.1. Automatic Initialization

In the work presented in [1, 2], several manually-defined landmark points are used to initialize the feature extraction procedure. In our method, as we explained in the Introduction, automatic initialization is achieved by using range images that contain only the shape information.

4.2. 3D Transformation

In the 3D-2D transformation presented in [2], a perspective projection is used to map the 3D morphable model to a 2D image, which makes the transformation nonlinear. In our scheme, the mapping between the 3D morphable model and a 3D range image is a linear transformation because the perspective projection is not needed. Since the 3D model and the normalized input 3D faces have the same head pose, this linear transformation can be further simplified to scaling and shifting. Let the original coordinates of a point p be $(x(i, \boldsymbol{\alpha}), y(i, \boldsymbol{\alpha}), z(i, \boldsymbol{\alpha}))$, which are defined by the index i and the shape coefficients $\boldsymbol{\alpha}$[2]. We use the following transformation to transfer the origin coordinates to the new coordinates $(x'(i, \boldsymbol{\alpha}), y'(i, \boldsymbol{\alpha}), z'(i, \boldsymbol{\alpha}))$:

$$x'(i, \boldsymbol{\alpha}) = scale_{x,y} \cdot x(i, \boldsymbol{\alpha}) - off_x, \tag{3}$$

$$y'(i, \boldsymbol{\alpha}) = scale_{x,y} \cdot y(i, \boldsymbol{\alpha}) - off_y, \tag{4}$$

$$z'(i, \boldsymbol{\alpha}) = scale_z \cdot z(i, \boldsymbol{\alpha}) - off_z. \tag{5}$$

In the above transformation, the values of the scale factors $scale_{x,y}$, $scale_z$ and the offsets off_x, off_y and off_z depend on the input range image format. Note that $scale_{x,y}$ and $scale_z$ may not have the same value.

4.3. Cost Function

Let R represent the input range image and m be the length of the corresponding shape vector $\mathbf{s}$. When an index i ($i = 1, \ldots, m$) of $\mathbf{s}$ and the shape coefficients $\boldsymbol{\alpha}$ are given, the described linear transformation maps the original coordinates $s(i, \boldsymbol{\alpha})$ to the new coordinates $(x'(i, \boldsymbol{\alpha}), y'(i, \boldsymbol{\alpha}), z'(i, \boldsymbol{\alpha}))$. We define the cost function as the sum of the squared difference between the transformed depth $z'(i, \boldsymbol{\alpha})$ and the corresponding range image depth $R(x'(i, \boldsymbol{\alpha}), y'(i, \boldsymbol{\alpha}))$, which can be represented as

$$C(\boldsymbol{\alpha}) = \sum_{i=1}^{m}(z'(i, \boldsymbol{\alpha}) - R(x'(i, \boldsymbol{\alpha}), y'(i, \boldsymbol{\alpha})))^2. \tag{6}$$

Note that this cost function is based on all the indices of the shape vector $\mathbf{s}$. In other words, we use $C(\boldsymbol{\alpha})$ to measure the entire range difference between the 3D model and the 3D face in R. This cost function can then be minimized by using standard optimization methods (e.g., Newton's method).

[2] We do not have texture coefficients $\boldsymbol{\beta}$ when only range images are involved.

4.4. Maximum A Posteriori Estimator

Minimization of cost function $C(\boldsymbol{\alpha})$ with respect to $\boldsymbol{\alpha}$ may produce unrealistic 3D faces. So we use a maximum a posteriori (MAP) estimator to modify $C(\boldsymbol{\alpha})$. Given an input image R, we try to maximize the posterior probability $p(\boldsymbol{\alpha}|R)$ with respect to the shape coefficients $\boldsymbol{\alpha}$. According to the Bayes rule,

$$p(\boldsymbol{\alpha}|R) \propto p(R|\boldsymbol{\alpha})p(\boldsymbol{\alpha}). \tag{7}$$

Given the shape coefficients $\boldsymbol{\alpha}$ of the model, the distribution of an input range image R is assumed to follow a normal distribution, i.e., $p(R|\boldsymbol{\alpha}) \propto \exp(-\frac{1}{2\sigma_R^2}C(\boldsymbol{\alpha}))$. The posterior probability is then maximized by minimizing

$$C'(\boldsymbol{\alpha}) = -2\log p(\boldsymbol{\alpha}|R) = \frac{1}{\sigma_R^2}C(\boldsymbol{\alpha}) + \sum_{i=1}^{n}\frac{\alpha_i^2}{\sigma_{R,i}^2}, \tag{8}$$

where σ_R^2 is a controllable parameter representing the relative weight of $C(\boldsymbol{\alpha})$ in $C'(\boldsymbol{\alpha})$. Thus after the modification of the MAP estimator, $C'(\boldsymbol{\alpha})$ is the cost function that we try to minimize by optimizing the shape coefficients $\boldsymbol{\alpha}$.

4.5. Optimization Procedure

The cost function $C(\boldsymbol{\alpha})$ is based on all the indices of the shape vector $\mathbf{s}$. In order to accelerate the minimization of $C(\boldsymbol{\alpha})$ without falling into local minima, we employ a stochastic version of Newton's method. In each iteration, eighty indices of $\mathbf{s}$ are randomly selected to compose a set $\mathcal{K}$. Based on the set $\mathcal{K}$, a new cost function $C_{\mathcal{K}}$ is defined:

$$C_{\mathcal{K}}(\boldsymbol{\alpha}) = \sum_{i\in\mathcal{K}}(z'(i,\boldsymbol{\alpha}) - R(x'(i,\boldsymbol{\alpha}), y'(i,\boldsymbol{\alpha})))^2. \tag{9}$$

The corresponding cost function after the modification of the MAP estimator becomes:

$$C'_{\mathcal{K}}(\boldsymbol{\alpha}) = \frac{1}{\sigma_R^2}C_{\mathcal{K}}(\boldsymbol{\alpha}) + \sum_{i=1}^{n}\frac{\alpha_i^2}{\sigma_{R,i}^2}. \tag{10}$$

In each iteration, the first and second derivatives of $C_{\mathcal{K}}(\boldsymbol{\alpha})$ with respect to the jth shape coefficient α_j $(j = 1, \ldots, n)$ can be calculated as

$$\begin{aligned}\frac{\partial C_{\mathcal{K}}(\boldsymbol{\alpha})}{\partial \alpha_j} = &\sum_{i\in\mathcal{K}} 2(z'(i,\boldsymbol{\alpha}) - R(x'(i,\boldsymbol{\alpha}), y'(i,\boldsymbol{\alpha})))\cdot \\ &(\frac{\partial z'(i,\boldsymbol{\alpha})}{\partial \alpha_j} - \frac{\partial R(x'(i,\boldsymbol{\alpha}), y'(i,\boldsymbol{\alpha}))}{\partial x'}\cdot\frac{\partial x'(i,\boldsymbol{\alpha})}{\partial \alpha_j} \\ &- \frac{\partial R(x'(i,\boldsymbol{\alpha}), y'(i,\boldsymbol{\alpha}))}{\partial y'}\cdot\frac{\partial y'(i,\boldsymbol{\alpha})}{\partial \alpha_j}),\end{aligned} \tag{11}$$

$$\begin{aligned}\frac{\partial^2 C_{\mathcal{K}}(\boldsymbol{\alpha})}{\partial \alpha_j^2} = &\sum_{i\in\mathcal{K}} 2(\frac{\partial z'(i,\boldsymbol{\alpha})}{\partial \alpha_j} \\ &- \frac{\partial R(x'(i,\boldsymbol{\alpha}), y'(i,\boldsymbol{\alpha}))}{\partial x'}\cdot\frac{\partial x'(i,\boldsymbol{\alpha})}{\partial \alpha_j} \\ &- \frac{\partial R(x'(i,\boldsymbol{\alpha}), y'(i,\boldsymbol{\alpha}))}{\partial y'}\cdot\frac{\partial y'(i,\boldsymbol{\alpha})}{\partial \alpha_j})^2 \\ &+ \sum_{i\in\mathcal{K}} 2(z'(i,\boldsymbol{\alpha}) - R(x'(i,\boldsymbol{\alpha}), y'(i,\boldsymbol{\alpha}))) \\ &\cdot(\frac{\partial^2 z'(i,\boldsymbol{\alpha})}{\partial \alpha_j^2} - \frac{\partial^2 R(x'(i,\boldsymbol{\alpha}), y'(i,\boldsymbol{\alpha}))}{\partial \alpha_j^2}).\end{aligned} \tag{12}$$

With these derivatives, we can further obtain the derivatives $\frac{\partial C'_{\mathcal{K}}(\boldsymbol{\alpha})}{\partial \alpha_j}$ and $\frac{\partial^2 C'_{\mathcal{K}}(\boldsymbol{\alpha})}{\partial \alpha_j^2}$ $(j = 1, \ldots, n)$. The shape coefficients $\boldsymbol{\alpha}$ are updated according to $\boldsymbol{\alpha} = \boldsymbol{\alpha} - \lambda H^{-1}\nabla C'_{\mathcal{K}}$, where $H^{-1} \approx diag(1/\frac{\partial^2 C'_{\mathcal{K}}(\boldsymbol{\alpha})}{\partial \alpha_j^2})$ $(j = 1, \ldots, n)$ is the inverse Hessian matrix and $\lambda \ll 1$ the learning rate.

4.6. Segmentation

In some subregions (e.g., eyes and mouth) of a 3D face, the range differences are subtle. Thus if we use the whole 3D morphable model to minimize the cost function $C(\boldsymbol{\alpha})$, the locations of these subregions on the resulting synthesized 3D face may not match those on the input range image. We therefore segment the 3D model into three separate subregions (e.g., eyes, nose and mouth in Fig. 1). After fitting the whole 3D model to a range image, we independently minimize $C(\boldsymbol{\alpha})$ based on these subregions. This segmentation step improves the detailed description of the subregions on the synthesized 3D face.

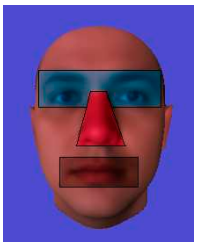

Fig. 1. Three segmented subregions on the 3D morphable model.

5. FEATURE EXTRACTION FROM TEXTURE IMAGES

So far our feature extraction method only utilizes 3D range images for feature extraction. Extra features can be obtained by minimizing the color difference between the 3D model and 2D texture images. The details of the procedure are omitted here because of space limitations. Thus our feature extraction method can be considered as a 3D+2D approach that utilizes both 3D range and 2D texture images for feature extraction.

6. EXPERIMENTAL RESULTS

6.1. Test Data Set

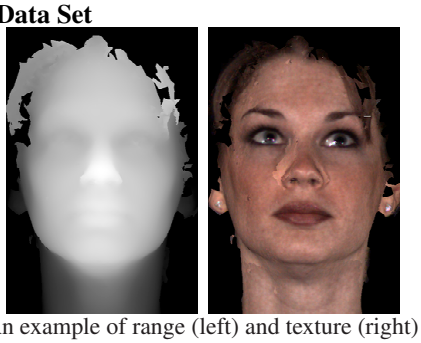

Fig. 2. An example of range (left) and texture (right) images.

The test data set contains both range and texture images. Each human face is represented by a pair of range and texture images as shown in Fig. 2. A range/texture image is 500×750 and each pixel has 8 bit resolution. These images are normalized before being used as inputs of our fitting algorithm. The normalized range and texture images have the following properties: 1) the tip of the nose on a range image is at the center of the image; 2) the depth of the tip of the nose has the gray level of 255; 3) zero gray level represents a plane 82 mm behind the tip of the nose.

6.2. Synthesis of Range and Texture Images

We test our algorithm by reconstructing 3D faces from input images. For display purposes, an extra fitting procedure is used to recover the color information of an input face. We show a fitting example in Fig. 3. The shape coefficients are optimized by fitting the 3D morphable model to the range image in Fig. 3 (a). After that, we fix the shape coefficients and optimize the color coefficients by fitting the 3D model to the texture image in Fig. 3 (b). Finally, we show the frontal view of the synthesized 3D face in Fig. 3 (c).

(a) (b) (c)

Fig. 3. Fitting results. (a) Original range image. (b) Original texture image. (c) Synthesized image.

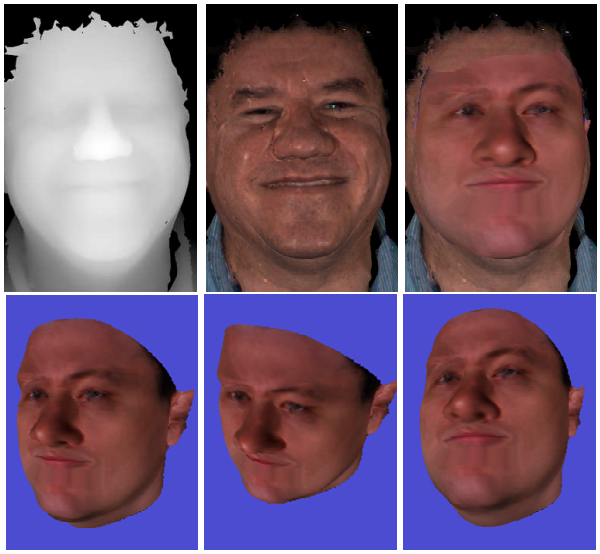

Fig. 4. Fitting results on a smiling face without perfect normalization. Top left: input range image; Top middle: input texture image; Top right: frontal view of the synthesized 3D face; Bottom row: different views of the synthesized 3D face.

Some of the faces in the data set are not well normalized and some have facial expressions, beards and mustaches. Fig. 4 shows a fitting example on one of such faces. An extra fitting based on texture image is performed for display purposes. In this example, the face in the range and texture images has a cheerful facial expression and the right eye is slightly higher than the left eye. After the automatic face fitting procedure, the frontal view (Fig. 4 [top right]) and other views (Fig. 4 [bottom row]) of the synthesized 3D face are shown to demonstrate the effectiveness of our method.

6.3. Feature Localization

To further test the performance of our algorithm, we run our proposed algorithm on 100 range images to label the positions of the tip of the nose, eye corners and mouth corners. Some of the results are shown in Fig. 5. Again for display purposes, we show the labeling results on texture images, although our tests are performed on range images only. It is seen that our proposed algorithm produces accurate localization results.

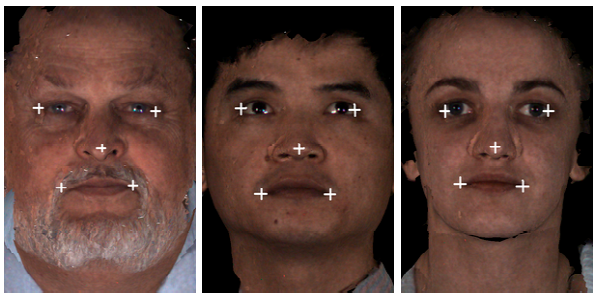

Fig. 5. Localization results.

7. REFERENCES

[1] V. Blanz and T. Vetter, "Morphable model for the synthesis of 3D faces," *The 26th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 187-194, July 1999.

[2] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1063-1074, September 2003.

[3] G. Gordon, "Face recognition based on depth and curvature features," *Computer Vision and Pattern Recognition*, pp. 108-110, June 1992.

[4] T. Nagamine, T. Uemura, and I. Masuda, "3D facial image analysis for human identification," *International Conference on Pattern Recognition*, pp. 324-327, August 1992.

[5] C. Chua, F. Han, and Y. K. Ho, "3D human face recognition using point signature," *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 233-238, March 2000.

[6] Y. Lee and J. Shim, "Curvature-based human face recognition using depth-weighted Hausdorff distance," *International Conference on Image Processing*, pp. 1429-1432, October 2004.

[7] X. Lu, D. Colbry, and A. K. Jain, "Matching 2.5D scans for face recognition," *International Conference on Pattern Recognition*, pp. 362-366, August 2004.

[8] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Three-dimensional face recognition," *International Journal of Computer Vision*, pp. 5-30, 2005.

[9] H. T. Tanaka, M. Ikeda, and H. Chiaki, "Curvature-based face surface recognition using spherical correlation principal directions for curved object recognition," *Third International Conference on Automated Face and Gesture Recognition*, pp. 372-377, April 1998.