A Novel Expression Deformation Model for 3D Face Recognition*

Chuanjun WANG†(a), Li LI†(b), Xuefeng BAI†(c), Nonmembers, and Xiamu NIU††(d), Member

SUMMARY The accuracy of non-rigid 3D face recognition is highly influenced by the capability to model the expression deformations. Given a training set of non-neutral and neutral 3D face scan pairs from the same subject, a set of Fourier series coefficients for each face scan is reconstructed. The residues on each frequency of the Fourier series between the finely aligned pairs contain the expression deformation patterns and PCA is applied to learn these patterns. The proposed expression deformation model is then built by the eigenvectors with top eigenvalues from PCA. Recognition experiments are conducted on a 3D face database that features a rich set of facial expression deformations, and experimental results demonstrate the feasibility and merits of the proposed model.

key words: 3D face recognition, expression deformation modeling, Fourier series, PCA

1. Introduction

With the advent of 3D scanning devices, 3D face recognition becomes efficient and affordable. The captured 3D face scans contain rich 3D shape information that are immune to the effects of illumination and light. Moreover, the captured shapes are not affected by translation or rotation. With these advantages, the 3D face scans are considered potential tools for tackling some of the unsolved challenges in 2D face recognition, and the research work of face recognition has shifted from 2D to 3D. However, in the 3D face recognition system, facial expression deformation is still one of the main challenges [1]. The geometry of human faces can deform drastically when facial expression occurs, which will greatly affect recognition performance.

As stated in [2], approaches to deal with the expression challenges can be classified into two categories. The first type treats human faces as rigid objects and uses the regions that are less affected for recognition [3], while the second treats human faces as non-rigid objects and tries to apply deformations to the 3D face scans to counteract expression deformations [2], [4].

Although many of the current methods focus on the first type, the non-rigid approaches are more promising in handling the expression challenges because it provides potentials to robustly model facial expression deformations and hence be able to extract more information from the face [2], [4]. In the pioneer works of facial expression deformation modeling, Lu et al. [4] first use the hierarchically resampled 94 feature points to represent the surface of 3D face scan. Then, the thin plane spline is used to model the deformation between the non-neutral and neutral faces on these 94 points. In Al-Osaimi’s work [2], they first transform the 3D face scan into range image and use the x and y coordinates as the grid plane. Then the facial scans are equally interpolated as a set of depth map at 1mm distance along the x and y axes. Based on the depth map representations of the 3D face scans, the expression deformation is modeled by the shape residue between the non-neutral and neutral scan pairs. Further researches [5]–[7] are also conducted to exploit the idea of expression deformation modeling [2], [4].

Inspired by the work of Al-Osaimi [2], we find that the surface representation of 3D face is a crucial component in the non-rigid approaches. In this letter, 3D face surfaces are first reconstructed and represented as a set of Fourier series coefficients. Then, the residues between the finely aligned non-neutral and neutral face surfaces that belong to the same subject are computed. These residues contain the expression deformation patterns, and PCA is applied to learn these patterns. The eigenvectors with top eigenvalues in the generated lower dimensional subspace of PCA are then used to build the Fourier Series based Expression Deformation Model (FSEDM). Based on the FSEDM, negative similarity measures for recognition are calculated from the distance between the residues and their projections on the FSEDM subspace. Experimental results show that the FSEDM achieves better performance than that in [2].

2. Fourier Series Representation of 3D Face Surface

2.1 Preprocessing of the 3D Face Data

To derive the uniform Fourier series representation of 3D face scans, we first transform the scanned 3D face into a standard face-based depth map [8]. In this face-based depth map, the nose tip of the 3D face is set as the origin of the coordinates, and the mid-line plane of the 3D face is identified as the YOZ plane. Moreover, the X and Y coordinates...
of the depth map are regarded as the grid plane and the Z coordinate as the shape surface of the 3D face.

Thus we can use an ellipse function of Eq. (1) on the XOY plane to fit the training samples as an average face model and generate the elliptic cross section to crop the facial region. To restrict the effect of outliers like the hair, and the neck, a threshold condition $y_{\text{min}} \leq y \leq y_{\text{max}}$ is also adopted. In our experiment, $y_{\text{min}}$ is set to be 20mm below the average of the two mouth corners and $y_{\text{max}}$ is set to be 30mm above the average of the two inner eye corners. Figure 1 (a) gives an illustration of a cropped 3D face depth map representation within this cross section.

$$\frac{x^2}{a^2} + \frac{(y-y_0)^2}{b^2} = 1$$

(1)

Within the cropped facial region, a resample process is adopted to generate the depth map as a $M \times N$ matrix, where $M$ indicates the number of curves and $N$ is the number of resampled points on each of the curves. Therefore, the cropped facial region of each face scan can be represented as $\{V_{ij}\} = F_d$ shown in Eq. (2) to denote the depth map of the cropped facial region for simplicity.

$$F_d = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_M \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1N} \\ z_{21} & z_{22} & \cdots & z_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ z_{M1} & z_{M2} & \cdots & z_{MN} \end{bmatrix}$$

(2)

in which $z_{ij}$ is the value of depth corresponds at the position of $(x_{ij}, y_{ij})$. A series of the resampled curves are illustrated in Fig. 1 (b).

In order to derive the parametric domain of the Fourier series representation, we extend the depth map representation of the 3D face with a mirror projection in Eq. (3).

$$V_{ij}^{(dm)} = \begin{bmatrix} x_{ij} \\ y_{ij} \\ -z_{ij} \end{bmatrix}, \text{ } 1 \leq i \leq M, \text{ } 1 \leq j \leq N$$

(3)

By this method, the depth map representation of the 3D face scans is extended as $F = \{V_{ij}\} \cup \{V_{ij}^{(dm)}\}$. And an example of $F$ is shown in Fig. 1 (c).

2.2 Reconstruction Using the Fourier Series Coefficients

From the viewpoint of signal processing, each of these resampled curves can be regarded as one dimensional signal and there are different methods for computing the Fourier series coefficients to represent it.

Suppose $f_i = (z_{i1}, z_{i2}, \cdots, z_{iN})$ is resampled from the continuous periodic signal $f$ that last for 1 second on an interval $0 \leq w \leq 2\pi$. Thus we can parameterize each of the extracted curves as Fourier series coefficients in the form of Eq. (4). The parametric domain $0 \leq w \leq \pi$ represents the scanned surface shape of the 3D face while $\pi \leq w \leq 2\pi$ represents the extended reflection of the scanned 3D face.

$$g_i(w) = a_0^i + \sum_{n=1}^{\infty} \left( a_n^i \cos (nw) + b_n^i \sin (nw) \right)$$

(4)

where $n$ indicates the frequency of the wave, and $a_n^i, b_n^i$ is calculated by Eq. (5)

$$a_n^i = \frac{1}{\pi} \int_{0}^{\pi} f_i(z_i) \cos (nw) \, dw,$$

$$b_n^i = \frac{1}{\pi} \int_{0}^{\pi} f_i(z_i) \sin (nw) \, dw$$

(5)

Since simply truncating the first K terms of the Fourier series will introduce numerical instability, we applied a model selection[9] process on Eq. (4) to select the appropriate subspace of the orthogonal basis among $\{1/2, \cos (w), \sin (w), \cos (2w), \sin (2w), \cdots\}$ to approximate each of the extracted surface curves.

Firstly, according to mirror reflection shown in Figure 1 (c), $f_i$ is an odd function due to $f_i(-z_{ij}) = -f_i(z_{ij})$. Therefore, a subset of orthogonal basis $\{\sin (nw)\}_{n=1}^{\infty}$ is selected.

Secondly, noticing that even some low frequencies of the basis $\{\sin (nw)\}_{n=1}^{\infty}$ in Fourier series are not necessarily important, we further take advantage of the importance evaluation of the frequencies to select the necessary frequencies with Eq. (6).

$$\frac{\sqrt{K} (b_n^i - \mu_K)}{\sigma_K} > I^{(K)}_K, \text{ } K < +\infty$$

(6)

where $\mu_K$ and $\sigma_K$ are the mean value and standard deviation of $\{b_n^i\}^{K}_{n=1}$ respectively. The threshold $I^{(K)}_K$ means that we choose the terms that contribute to the Fourier series at $I^{(K)}_K$ level, or in other words that we ignore the terms that contribute to the Fourier series with a ratio less than $I^{(K)}_K$. Based on the model selection procedure, a design matrix $Y_i$ is generated with the selected terms of the Fourier series. And the corresponding Fourier series coefficients $\beta_i$ can be obtained by minimizing the reconstruction error of $e$ in Eq. (7) with the iterative residual fitting algorithm[10].

$$f_i^* = Y_i \beta_i + \varepsilon, \text{ } 1 \leq i \leq M$$

(7)

where the design matrix $Y_i$ and the Fourier series coefficients $\beta_i$ are shown in the following:

$$Y_i = \begin{bmatrix} \Phi_1(z_{i1}) & \Phi_2(z_{i1}) & \cdots & \Phi_K(z_{i1}) \\ \Phi_1(z_{i2}) & \Phi_2(z_{i2}) & \cdots & \Phi_K(z_{i2}) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_1(z_{iN}) & \Phi_2(z_{iN}) & \cdots & \Phi_K(z_{iN}) \end{bmatrix}_{N \times K}$$
\[ \beta^T = \left( b_1^{(0)} b_2^{(0)} \cdots b_K^{(0)} \right). \]

\( \{ \Phi_j \}_{j=1}^K \) are the first \( K \) basis functions of \( \sin(nw) \) \( \alpha_{n=1}^\infty \). The Fourier series coefficient \( b_j \) corresponds to \( \Phi_j \) is equal to 0 if \( \Phi_j \) is not selected. Therefore, a 3D face scan is reconstructed as a set of Fourier series coefficients shown in Eq. (8):

\[ F_{fs} = (\beta_1 \beta_2 \cdots \beta_M)^T = (B_1 \ B_2 \cdots B_K) \quad (8) \]

where \( B_k = \left( b_k^{(1)} b_k^{(2)} \cdots b_k^{(M)} \right)^T, 1 \leq k \leq K. \) It describes the \( M \) coefficients of Fourier series on frequency \( k \) for each of the face scan.

3. Fourier Series based Expression Deformation Model (FSEDM)

3.1 Expression Deformation Modeling

Suppose there are \( N_F \) training face scan pairs (both of each pair belong to the same subject). Let \( F_{fs,x}^{(i)} \) and \( F_{fs,y}^{(i)} \) denote the non-neutral and neutral face scans reconstructed with Eq. (8) for the \( i \)th training pair. The \( i \)th Fourier series residues \( R_{fs}^{(i)} \) can be calculated with Eq. (9).

\[ R_{fs}^{(i)} = F_{fs,x}^{(i)} - F_{fs,y}^{(i)} = \left( R_{1}^{(i)} R_{2}^{(i)} \cdots R_{K}^{(i)} \right), 1 \leq i \leq N_F \]

where \( R_k^{(i)} = \left( r_{ik}^{(1)} r_{ik}^{(2)} \cdots r_{ik}^{(M)} \right)^T \) is the residue corresponds to frequency \( k \) and it contains the patterns respond to the expression deformation. So, on frequency \( k \), we apply PCA on these \( N_F \) training Fourier series residues with Eq. (10).

\[ \Omega_k = \sum_{i=1}^{N_F} \left( R_{k}^{(i)} - \bar{R}_k^{(i)} \right) \left( R_{k}^{(i)} - \bar{R}_k^{(i)} \right)^T, 1 \leq k \leq K \quad (10) \]

where \( \bar{R}_k^{(i)} \) is the average value of all the \( R_k^{(i)} \) in the \( N_F \) training scan pairs.

The eigenvectors of \( \Omega_k \) with higher eigenvalues are largely related to the expression deformation on frequency \( k \). So the deformation of facial expressions on frequency \( k \) can be modeled with the eigenvectors corresponding to the top \( P \) eigenvalues and a subspace residue model \( E_k \) can be generated as Eq. (11).

\[ E_k = [e_{1,k}, e_{2,k}, \ldots, e_{p,k}], 1 \leq P \ll M \quad (11) \]

where \( \{ e_{p,k} \}_{p=1}^P \) are the eigenvectors of \( \Omega_k \) with the top \( P \) eigenvalues. Therefore, the expression deformation model of the whole face can be denoted by the combination of \( E_k \) on all the \( K \) frequencies. In this letter, we call \( E = \{ E_1 \}_{k=1}^K \), Fourier Series based Expression Deformation Model (FSEDM).

3.2 FSEDM Based Face Matching

As the FSEDM \( E \) is built from the \( N_F \) training residues (residues between the non-neutral and neutral face scans of the same subject), it provides the ability to model the expression deformations, while it lacks the capacity to model the inter-personal disparities (residues of expressions plus geometry disparities between different subjects). Thus the idea of the FSEDM based face matching is to test whether a residue is generated from the same subject. Suppose there are \( N_G \) neutral 3D face scans registered as the gallery subjects \( \{ G_i \}_{i=1}^{N_G} \). Given a probe 3D face scan \( p \), the process to calculate the negative similarity distance between \( p \) and \( G_i \) is performed as follow:

1. Residues computation. Let \( F_{fs,p}^{(i)} \) and \( F_{fs,p}^{(i)} \) be the Fourier series coefficients reconstructed from \( 3D \) face \( p \) and \( G_i \) respectively. Similar to Eq. (9), the residues \( R_{fs,p}^{(i)} = \left( R_{1,p}^{(i)} R_{2,p}^{(i)} \cdots R_{K,p}^{(i)} \right) \) between \( F_{fs,p}^{(i)} \) and \( F_{fs,p}^{(i)} \) is first calculated. Obviously, \( R_{k,p}^{(i)} (1 \leq k \leq K) \) describes the expression deformations if \( p \) and \( G_i \) belong to the same subject. Otherwise, \( R_{k,p}^{(i)} \) is expression deformations plus inter-personal disparities between subjects.

2. FSEDM projection. Each \( R_{k,p}^{(i)} (1 \leq k \leq K) \) is projected on the subspace \( E_k \) with Eq. (12).

\[ R_{k,p}^{(i)} = E_k (E_k^T E_k)^{-1} E_k^T R_{k,p}^{(i)}, 1 \leq k \leq K \quad (12) \]

3. Negative similarity distance calculation. \( R_{k,p}^{(i)} \) will be very close to \( R_{k,p}^{(i)} \) if \( p \) and \( G_i \) belong to the same subject. Otherwise, \( R_{k,p}^{(i)} \) will be far away from \( R_{k,p}^{(i)} \). So the negative similarity measure \( S_i \) between \( p \) and \( G_i \) is calculated with Eq. (13).

\[ S_i = \sum_{k=1}^{K} d\left(R_{k,p}^{(i)}, R_{k,p}^{(i)}\right) \quad (13) \]

where, \( d(r_i, r_j) = \left\{ \begin{array}{ll} d_i, & \| r_i - r_j \| \geq d_i \\ \| r_i - r_j \|, & \text{else} \end{array} \right. \)

In this letter, \( d_i \) is empirically set to be 2 in our experiment. As stated above, if \( p \) and \( G_i \) belong to the same subject, \( S_i \) will be very small. Therefore, the gallery 3D face scans with the minimum \( S_i \) in Eq. (13) is considered to be the probe’s match.

4. Evaluations and Experiments

The Bosphorus 3D face database \([11]\) is used in our experiment. This database features a rich set of facial expression deformations. It has been widely used for evaluation of face recognition algorithms under adverse conditions, as well as facial expressions analysis \([12]\). The expression variations in this database contain not only six universal emotional expressions (happiness, surprise, fear, sadness, anger and disgust) but also face action units of the facial action coding system (FACS). Based on the properties of the Bosphorus 3D face database, 2954 3D facial scans of 105 subjects are used in our experiments, where each subject has 16-37 facial scans with different facial expressions.
We configured parameters $M = 256$ and $K = 8$ empirically for each 3D face scan. Therefore, each of the 3D face scan is reconstructed as a $256^*8$ matrix of the Fourier series coefficients. The average root mean squared distance between the Fourier series representation and the original 3D face scan is 0.39mm under the parameter setting in our experiments. It is close to the resolution of the 3D face scans (0.4mm), therefore, the original 3D face scan is well approximated by the reconstructed Fourier series.

As studied in [2], insufficient instance of facial expression residues may lead to noisy eigenvectors of the expression deformation model. To test the effect of the training data size, 3, 6, and 10 non-neutral face scans of each subject are randomly selected as the training set. These face scans are then combined with the corresponding neutral face scans to generate the training pairs for FSEDM. Meanwhile, one neutral face scans per subject is enrolled as the gallery dataset, and all the other face scans are used as probe scans in the experiment.

Figure 2 shows the ROC curves for different cases. The solid line and the dotted line indicate the proposed FSEDM and the EDM proposed in [2] respectively. The different colors indicate the different number of 3, 6, and 10 non-neutral face scans per subject used in the training set. Therefore, a total of 315, 630 and 1050 pairs of residues are used in the training set, respectively. In Fig. 2, we can see that the recognition performance increases with the larger training set size, which is consistent with [2]. Furthermore, the verification rate at the equal error rate (EER) of the FSEDM is 97.2%, 97.6%, and 98.3%, respectively, compared to the corresponding value of 96.2%, 96.9% and 97.8% respectively of the EDM in [2]. We observe that the FSEDM provides better performance than EDM.

5. Conclusions and Future Research

A novel expression deformation model FSEDM for 3D face recognition is proposed. Inspired by the work of [2], we take advantage of the idea that the residues between the non-neutral and neutral 3D face scans contain the patterns of the expression deformation. Based on this assumption, the 3D face scans are first parameterized and reconstructed as a series of Fourier series coefficients. Then the frequency residues between the non-neutral and neutral face scans are calculated. Finally, PCA is applied to learn the expression deformation patterns and used to build the expression deformation model. Experimental results demonstrate the feasibility and merits of FSEDM.

In the future, we will improve the FSEDM against the pose variation and investigate the potential ability of FSEDM in further application, including facial expression analysis and expression synthesis.

Acknowledgment

We would like to thank for the development of the Bosphorus 3D face database [11]. We also thank all anonymous reviewers for their helpful comments and suggestions.

References

[1] K. Bowyer, K. Chang, and P. Flynn, “A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition,” J. Computer Vision and Image Understanding, vol.101, no.1, pp.1–15, 2006.
[2] F. Al-Osaimi, M. Bennamoun, and A. Mian, “An expression deformation approach to non-rigid 3D face recognition,” International Journal of Computer Vision, vol.81, no.3, pp.302–316, 2009.
[3] A. Mian, M. Bennamoun, and R. Owens, “An efficient multimodal 2D-3D hybrid approach to automatic face recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol.29, no.11, pp.1927–1943, 2007.
[4] X. Lu and A.K. Jain, “Deformation modeling for robust 3D face matching,” IEEE Trans. Pattern Anal. Mach. Intell., vol.30, no.8, pp.1346–1357, 2008.
[5] F. Al-Osaimi, M. Bennamoun, and A. Mian, “On decomposing an unseen 3D face into neutral face and expression deformations,” Proc. 3rd International Conference on Advances in Biometrics, pp.22–31, Alghero, 2009.
[6] S. Wang and S. Lai, “Reconstructing 3D face model with associated expression deformation from a single face image via constructing a low-dimensional expression deformation manifold,” IEEE Trans. Pattern Anal. Mach. Intell., vol.33, no.10, pp.2115–2121, 2011.
[7] G. Pan, S. Han, Z. Wu, and Y. Zhang, “Removal of 3D facial expressions: A learning-based approach,” Proc. 23rd IEEE Conference on Computer Vision and Pattern Recognition, pp.2614–2621, San Francisco, 2010.
[8] D. Colbry and G. Stockman, “Real-time identification using a canonical face depth map,” IET Computer Vision, vol.3, no.2, pp.74–92, 2009.
[9] S. Wang, Weighted Fourier image analysis and modeling, Ph.D. dissertation, University of Wisconsin, pp.86–111, 2008.
[10] K. Chung, K. Dalton, L. Shen, A. Eyans, and R. Davidson, “Weighted fourier series representation and its application to quantifying the amount of gray matter,” IEEE Transactions on Medical Imaging, vol.26, no.4, pp.566–581, 2007.
[11] A. Savran, N. Alyuz, H. Dibeklioglu, O. Eliktutan, B. Kherk, and B. Sankur, “Bosphorus database for 3D face analysis,” Proc. 1st Cost 2101 Workshop on Biometrics and Identity Management, pp.47–56, Denmark, 2008.
[12] A. Savran, B. Sankur, and M. Bilge, “Comparative evaluation of 3D versus 2D modality for automatic detection of facial action units,” Pattern Recognit., vol.45, no.2, pp.767–782, 2012.