Multimed Tools Appl (2017) 76:13–31
DOI 10.1007/s11042-015-3012-8

Expression-robust 3D face recognition based on feature-level fusion and feature-region fusion

Xing Deng[1,2] · **Feipeng Da**[1,2] · **Haijian Shao**[1,2]

Received: 4 January 2015 / Revised: 30 August 2015 / Accepted: 19 October 2015 /
Published online: 26 October 2015

Abstract 3D face shape is essentially a non-rigid free-form surface, which will produce non-rigid deformation under expression variations. In terms of that problem, a promising solution named Coherent Point Drift (CPD) non-rigid registration for the non-rigid region is applied to eliminate the influence from the facial expression while guarantees 3D surface topology. In order to take full advantage of the extracted discriminative feature of the whole face under facial expression variations, the novel expression-robust 3D face recognition method using feature-level fusion and feature-region fusion is proposed. Furthermore, the Principal Component Analysis and Linear Discriminant Analysis in combination with Rotated Sparse Regression (PL-RSR) dimensionality reduction method is presented to promote the computational efficiency and provide a solution to the curse of dimensionality problem, which benefit the performance optimization. The experimental evaluation indicates that the proposed strategy has achieved the rank-1 recognition rate of 97.91 % and 96.71 % based on Face Recognition Grand Challenge (FRGC) v2.0 and Bosphorus respectively, which means the proposed approach outperforms *state-of-the-art* approach.

Keywords 3D face recognition · Feature-region fusion · Feature-level fusion · Dimensionality reduction · Non-rigid point set registration

1 Introduction

The two-dimensional (2D) face recognition in past two decades has made significant progress in practical application. However, its recognition rate may be reduced due to some

This research is supported by National Natural Science Foundation of China (No. 51175081, No. 51475092, No. 61405034), Doctoral Fund of Ministry of Education of China (No. 20130092110027).

✉ Feipeng Da
dafp@seu.edu.cn

[1] Department of Automation, Southeast University, Nanjing 210096 Jiangsu, China

[2] Key Laboratory of Measurement and Control for Complex System of Ministry of Education, Southeast University, Nanjing 210096 Jiangsu, China

insurmountable limitations such as pose, illumination, expression and age variations. Comparing to the traditional 2D facial image, 3D facial scan has unique advantages because it contains more important feature information such as topological characteristic and geometric information. So the 3D face recognition has gained widely attention in recent years, and its broad survey has been provided by M. Daoudi [9].

In general, the current 3D face recognition methods according to the processing strategy can be categorized into three aspects: spatial matching-based, global features-based and local features-based. 3D face recognition approaches based on spatial matching don't provide the explicit facial feature extraction in 3D space, but directly match identification. The iterative closest point (ICP) method [3] was usually employed for face recognition, and its improved versions [26, 31] had become extremely effective methods for 3D point cloud matching and used widely for 3D point clouds alignment. At the same time, researchers utilized the Hausdorff distance [17] and its variants [20, 36] to calculate the similarity degree between 3D point clouds as the classification basis for face recognition. The global features-based recognition algorithms could be used to extract the entire 3D face features. The radial curves emanating from the nose tip in facial surface and the elastic shape analysis of these curves were used to develop a Riemannian framework for the entire facial surface shape analysis [10]. Apparently, the expression variations are one of the most difficult factors in 3D face recognition because they typically reduce the recognition rate. Therefore, local features of the 3D face were employed to overcome the influence from the expression variations [21, 25, 42]. The local features-based approach can overcome the influence from the facial expression variations but it loses some feature of human face.

3D face under expression variations results in sharply deterioration recognition accuracy, particularly the 3D face with large expression. Most recent pertinent literature on 3D face identification has directly extracted the global facial features without considering the non-rigid deformation caused by the expression variations [4], or commonly extracted local features from the rigid parts of the face and discarded non-rigid parts caused by the expressions [15, 21, 22, 42]. In addition, it is noting that the point cloud data of processed human face contains about 30, 000 points, in other words, the 3D face recognition approaches of the spatial matching-directly will increase the computational cost and may lower recognition speed. Motivated by the above considerations, the feature-level and feature-region fusion scheme is proposed, which can overcome the influence from the facial expression variations and utilize the global information of the human face fully.

The literature with relevant topics is summarized, which mainly focus on two aspects: face registration and feature extraction. Face registration is guided by automatically detecting landmarks and it greatly influents the subsequent recognition. Thus face registration is potentially one of the most expensive phases in 3D face recognition process, and it can be categorized into rigid and non-rigid registration, where the former aligns facial scans by an affine transformation, and the latter applies deformations to align facial structures more closely [13]. For rigid registration, the standard technique is the iterative closest point (ICP) algorithm [3]. Most of the 3D face recognition approaches in existence are based on rigid registration [27]. A typical example for non-rigid registration techniques is proposed in [5], which extracts non-rigid deformation by finding the mapping between two shapes.

Most 3D face registration approaches (rigid registration & non-rigid registration) are based on the whole face, regardless of whether the part of the face have deformed. If unnecessary areas such as the non-rigid region of the face are also registered, the computational cost will be increased. This is not conducive to the practical application. In addition, the 3D face shape is

essentially a non-rigid free-form surface, so the rigid registration doesn't meet the requirement of the actual situation. In this work, we conduct the non-rigid registration to the non-rigid region (the lower half of face) in order to achieve high recognition accuracy and computational efficiency simultaneously.

Feature extraction is a very critical step toward the high performance because the quality of the extracted characteristics will affect the final recognition result directly. Most 3D face recognition approaches were based on the global face representation or matching, for instance, Wang et al. [40] extracted the Gabor feature, local binary patterns (LBP) feature, and principle component feature on the depth image. Then they used the Corresponding Point Direction Measure (CPDM) to match the 3D face with the gallery and obtained the score as the similarity measurement. Finally, they fused all extracted features to finish the recognition. A typical region-based 3D face recognition approach was that Chang et al. [7] proposed multiple nose regions matching under varying facial expressions. They segmented the 3D face into multiple sub-regions, and then combined the match scores from matching multiple overlapping sub-regions around the nose to determine the final recognition results.

The literature [40, 41] has shown that the application of wavelet transformation and wavelet analysis can yield relatively satisfactory result on the facial range images. Motivated by outlined considerations, we extract the features from the facial geometry images which not only retain the spatial coordinate information of the point set but also contain 3D topology information. The nearest neighborhood classification method is applied to compute the finial identification results based on the feature-level fusion and feature-region fusion technology.

The rest of this paper is organized as follows: Section 2 describes the details information related to our proposed approach. Section 3 presents the identification experimental results, and this paper is concluded in Section 4.

2 Proposed approach

In this paper, a fully automatic 3D face recognition method based on feature-level fusion and feature-region fusion is proposed in order to remove expression influence from the non-rigid facial region as well as utilize the feature of the whole facial region. The proposed strategy consists of five phases, and the corresponding block diagram is shown in Fig. 1. Firstly, face preprocessing step is fundamental to the 3D face recognition, which consists of smoothing

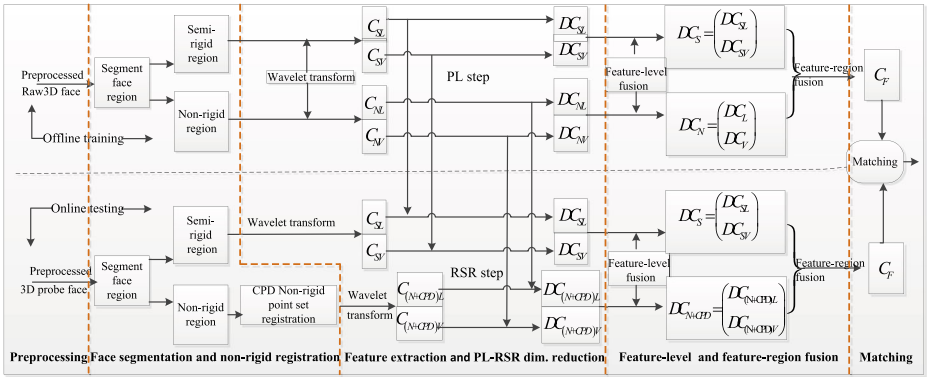

Fig. 1 The block diagram of the proposed approach

outlier, filling hole, face cropping and pose correction. Secondly, the face is automatically segmented into the semi-rigid and non-rigid regions, and CPD technique is utilized to develop a non-rigid point set registration for non-rigid region. Thirdly, the Haar wavelet feature with multi-resolution based on the geometric image is extracted and the PL-RSR dimensionality reduction method is used to reduce the dimensionality of the extracted wavelet feature. Fourthly, the feature-level fusion and feature-region fusion schemes are used to overcome expression variations and fully utilize the information of the whole face. Finally, the cosine similarity is employed to compute the similarity between two faces for face matching.

The experimental evaluation based on the Face Recognition Grand Challenge (FRGC) v2.0 [33] and Bosphorus [37] datasets is given to demonstrate the performance of proposed approach and show the competitive results with respect to 3D face recognition solution in existence. We summarize the main contribution of the paper as follows,

(1) The 3D face is divided into the semi-rigid and non-rigid regions for single-handing the expression-sensitive region, and the Coherent Point Drift (CPD) technique is utilized to develop a non-rigid point set registration for non-rigid region.
(2) The feature-level fusion technique is utilized so that the facial features of the low frequency and the vertical high-frequency are used fully.
(3) The feature-region fusion technique based on the semi-rigid region and non-rigid region after CPD non-rigid point set registration is used to extract the discriminative feature information of the whole face under expression variations.
(4) The proposed PL-RSR dimensionality reduction method is applied to promote the computational efficiency and provide a solution to the curse of dimensionality problem, which benefit the performance optimization.

2.1 Face preprocessing

The raw 3D facial models in FRGC v2.0 [33] and Bosphorus [37] datasets contain a number of holes and spikes, and include some undesired parts such as neck, ears, and shoulder (e.g. Fig. 2a), which will deteriorate the recognition performance. Furthermore, the data collected at different angle and the relative position between the scanner and the human face will lead to the raw model with different size and pose. Face preprocessing is an important issue for the quality guaranteed of the utilized sample, which mainly consists of smoothing outlier and filling hole, nose tip detection, face cropping and pose correction. Our previous work [23] proposes an automatic face preprocessing algorithm for the raw 3D facial model (see Fig. 2), and now a brief description is given below:

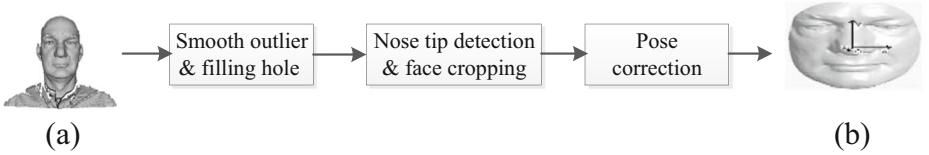

Fig. 2 Block diagram of the face preprocessing. (**a**) Raw 3D face model, (**b**) Preprocessed 3D face

Smoothing outlier and filling hole The Gaussian filter is used to remove the spikes caused by outlier points. Then, small holes in the face are filled by a bi-cubic interpolation from the valid neighbors of all three coordinates (x, y and z).

Nose tip detection and face cropping An automatically preprocessing algorithm proposed in our previous work [23] is used to detect the nose tip. The sphere of radius 90 mm centered at the detected nose tip is used to crop the 3D face.

Pose correction Principal component analysis (PCA) [39] is used to correct the pose changes. The corresponding eigenvectors can be obtained with largest and smallest eigenvalues by performing the PCA transform on the cropped face. The nose tip and the eigenvectors corresponding to the largest and smallest eigenvalue are selected as the coordinate origin, Y axis and Z axis, respectively. Thus, a new right-hand coordinate system called pose coordinate system (PCS) [43] is established. In the new coordinate system, the face region has a positive attitude, and each point is represented by a unique x, y, z coordinates (see Fig. 2b).

2.2 Face segmentation and non-rigid point set registration

2.2.1 Face segmentation

The recognition rate may be reduced if the whole facial features are used directly. As can be seen from the Fig. 3a, the mouth area is mainly located at the lower part of the face, which is very sensitive to the expression variations (such as laugh and angry). So we segment each face into semi-rigid region and non-rigid region. Taking into account the case of FRGC v2.0, the upper part of the face ($y \geq -15$) represents the expression insensitive semi-rigid region (Fig. 3b) and the lower part of the face ($y < -15$) represents the expression sensitive non-rigid region (Fig. 3c).

2.2.2 Non-rigid point set registration

It can be seen from the Fig. 3c that expression variations can make non-rigid region produce a hole in the mouth region, consequently the topology structure of the lower part of the face even

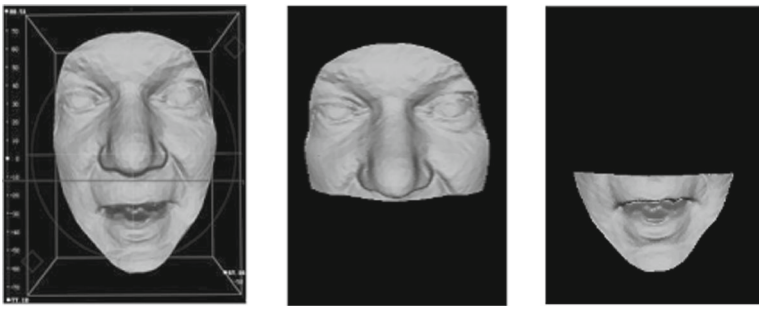

Fig. 3 (**a**) Before face segmentation, (**b**) Semi-rigid face region, (**c**) Non-rigid face region

the whole face is changed. In order to overcome the influence from the different facial expressions, a non-rigid point set registration method is employed. Note that we assume that all faces are neutral in gallery and the upper part of the face is insensitive for expression variations, therefore, we only apply non-rigid point set registration to non-rigid region in probe face.

Reference face We choose a neutral expression face randomly after preprocessing as reference face. The reference face in the FRGC v2.0 is shown in Fig. 4, where Fig. 4a is the corresponding 2D texture image, and Fig. 4b is the preprocessed reference face. As mentioned in Section 2.2.1, the reference face is segmented into the upper part of the reference face in Fig. 4c and the lower part of the reference face in Fig. 4d. The lower part of reference face is used for the non-rigid point set registration in this paper.

Non-rigid point set registration In this section, we introduce Coherent Point Drift (CPD) [32] to estimate complex non-linear and non-rigid transformation in face. Let the non-rigid region of the probe face is represented by the point set $Y_{M\times D}=(y_1,\cdots,y_M)^T$. $Y_{M\times D}$ should be aligned with the reference point $X_{N\times D}=(x_1,\cdots,x_N)^T$ related to the lower facial region of the reference face. Essentially, the points in $Y_{M\times D}$ is selected as the centroids of a Gaussian Mixture Model, and then $Y_{M\times D}$ is fitted to the data points $X_{N\times D}$ by maximizing the likelihood function. The width of Gaussian kernel (smoothness) is about 2, the regularization weight is about 3, the tolerance is 1e-10 and the max number of iterations is 100. All the inputs are normalized to unit variance and zero mean before registering.

The FRGC v2.0 and Bosphorus datasets consist of six different expressions *i.e.* surprise, fear, happy, sadness, disgust and anger. For the six kinds of expressions, we choose a representative image respectively and these texture images are listed in turn in Fig. 5. The Fig. 5 illustrates the results after non-rigid point set registration for non-rigid regions under six different expressions. Six typically different expressions: surprise, fear, happy, sadness, disgust and anger, are shown in the 1st row of Fig. 5, where the first six faces and the rest ones come from the FRGC v2.0 and Bosphorus datasets, respectively. The 2nd row shows the whole face which are used to segment the upper part of the face and the lower part of the face corresponding to semi-rigid and non-rigid regions of the face, respectively. The 1st row represents their corresponding texture images. The 3rd row presents the results of non-rigid point set registration for non-rigid regions and the 4th row gives the semi-rigid face regions by segmenting the corresponding face in the 2nd row. It can be seen from the Fig. 5 that the CPD

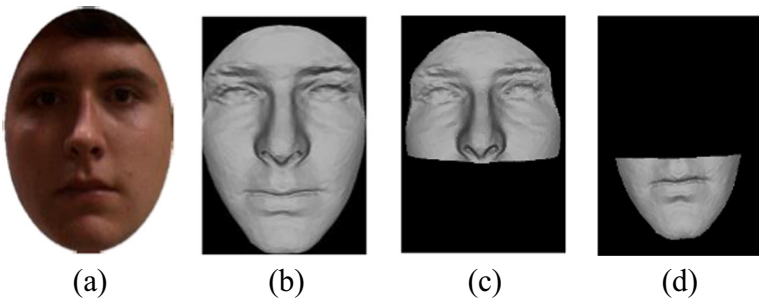

Fig. 4 (**a**) The corresponding texture image of reference face, (**b**) The preprocessed reference face, (**c**) Semi-rigid reference face region, (**d**) Non-rigid reference face region

Fig. 5 The results of non-rigid point set registration under six different expressions

non-rigid point set registration can be utilized to overcome the non-rigid deformation produced by the large expression, such as laugh and surprise, especially the mouth opening will make the face produce a hole.

2.3 Feature extraction and PL-RSR dimensionality reduction

2.3.1 Geometric image and mesh parameterization

Geometry images [14] are regularly sampled 2D images, which have three channels (x, y and z components of a vertex in R^3) encoding geometric information. In space triangle mesh, the boundary points and non-boundary points are mapped into the four sides and the inside of the square planar through the planar mesh parameterization, respectively. The coordinates related to x, y and z are utilized to be the vertices properties of triangular mesh, and the properties of each pixel in the region are calculated by the square shape of the linear interpolation method, thus the obtained 2D image is called as geometric image which contains all the 3D geometry information.

2.3.2 Feature extraction

The wavelet transform is a time-frequency domain analysis methods with the multi-resolution analysis, which is developed based on short-time Fourier transform. The wavelet transform is widely used in image processing because its advantages benefit for the Human visual system [28]. The 2D wavelet transform can be applied to the coarser version at half resolution [35], recursively.

Multiscale Haar wavelet transform can perform multiple iterations transform for the geometric image so that the image can be decomposed into the different frequency component, and then transformed into image in spatial domain. The transform matrix is given as follows,

$$Tg = H^T gH = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}, H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{1}$$

where Tg and g represent the original sub-block and transformed sub-block of geometric image, respectively. g_{12}, g_{21} and g_{22} denote the high-frequency coefficients at horizontal, vertical and diagonal directions of geometric image, respectively. g_{11} is the associated low-frequency coefficients which contain the most information of geometry image, and the corresponding high-frequency coefficients g_{12}, g_{21} and g_{22} represent the edge information

and detailed characteristics at the direction related to horizontal, vertical and diagonal of geometric image.

In this paper, the geometric image is decomposed into six layers, and only the first two layers are shown in Fig. 6 in order to display the decomposed image well. The preprocessed 3D face model is mapped into the parametric grid square plane with the size 512*512 pixels. Figure 6 denotes the Haar wavelet decomposition with multi-resolution for the geometric image. The (a) and (b) in Fig. 6 represent the preprocessed 3D face model and the mesh parametric planar image, respectively. The (c) is the geometric image related to mesh parametric planar image. The (d) and (e) in Fig. 6 are the Haar wavelet decomposition of geometric image at level-1 and level-2, respectively.

2.3.3 The PL-RSR dimension reduction

Essentially, the orders of magnitude related to the extracted wavelet feature of geometric image for each 3D face is almost 10^6, which means the required computational effort cannot satisfy the requirement of the real-time practical application indeed. Moreover, the pixels for geometric image derived by Haar wavelet transform are larger, the corresponding frequency domain information are more. Simultaneously, the associated computational complexity is increased. Therefore the PL-RSR method is proposed to reduce the dimensionality of wavelet features and improve the computational effectively. The whole PL-RSR dimension reduction method includes the training phase and the testing phase. The former is divided into two steps:

In the first step This step is called PL (PCA [18]+LDA [2]) step [16] since PCA is utilized to reduce the dimensionality of feature space, and LDA is used to optimize the features which are finally used to face identification. (1) PCA is utilized to compress the high-dimensional raw feature C. (2) Compute the weights $w_1, w_2, \cdots w_p$ for each training face image, where $p<M$ is the dimension of Eigen subspace on which the training face image is projected, and M is the number of the training images. (3) Store the weights $w_1, w_2, \cdots w_p$ for each training image as its facial features. (4) Perform LDA on the feature subspace to extract the discriminative feature information DC for 3D face recognition so that the dimension is further reduced potentially.

In the second step This step is called RSR (Rotated Sparse Regression [8]) step. The RSR method is applied to learn a sparse linear projection matrix φ through the mapping from C to DC learned in the first step. In particular, Chen et al. [8] proposed a l_1-based regression method to learn the sparse matrix φ with additional rotation matrix R which can be used to improve the sparisity.

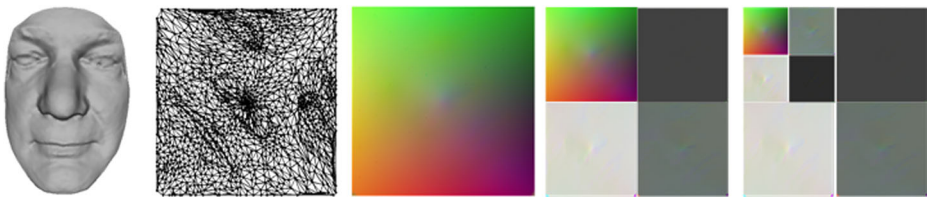

Fig. 6 Multi-resolution Haar wavelet decomposition. (**a**) Preprocessed 3D face model. (**b**) Mesh parametric planar image. (**c**) Geometric image. (**d**) The Haar wavelet decomposition of geometric image at level-1. (**e**) The Haar wavelet decomposition of geometric image at level-2

In the testing phase, the low-dimensional feature DC is calculated through the projection of the high-dimensional feature based on sparse matrix φ.

Rotated sparse regression (RSR). Suppose $C=[c_1,c_2,\cdots,c_n]$ and $DC=[dc_1,dc_2,\cdots,dc_n]$ are the input high-dimensional feature and the corresponding low-dimensional feature set derived by the PL step, respectively, where n is the number of training samples. Then a sparse linear projection φ maps the C to DC with low error, which is derived through the objective function defined by

$$\min_{\varphi} \left\|DC-\varphi^T C\right\|_2^2 + \lambda\|\varphi\|_1 \tag{2}$$

where the first term is the reconstruction error and the second term is the enforced sparse penalty. The scalar λ is used to adjust the balances between two terms. The RSR step introduces additional freedom in rotation to promote sparsity without sacrificing accuracy. With an additional rotation matrix R, a new formulation is established based on (2) and given by

$$\min_{\varphi,R} \left\|R^T(DC)-\varphi^T C\right\|_2^2 + \lambda\|\varphi\|_1 \quad s.t. \quad RR^T = I \tag{3}$$

Essentially, (3) can be called as RSR because it is a linear regression with sparse penalty and additional freedom in rotation.

Optimization. For similarity, an alternative optimization method is adapted, and the iteration is initialized through the assumption that matrix R is equal to the identify matrix.

(1) Solving φ via given R. The objective function is convex if R or φ is given, which can be rewritten as follows if let $\widetilde{DC} = R^T(DC)$,

$$\min_{\varphi} \left\|\widetilde{DC}-\varphi^T C\right\|_2^2 + \lambda\|\varphi\|_1 \tag{4}$$

Each column of φ can be optimized in parallel because it is independent for each other in (4). An effective coordinate descent method [12] is applied to implement that issue, which is initialized through the obtained value in a previous iteration.

(2) Solving R via given φ. The sparse penalty term is a constant if the matrix φ is fixed. The following objective function is derived by removing the constant penalty term

$$\min_{R} \left\|R^T(DC)-\varphi^T C\right\|_2^2 \quad s.t. \quad RR^T = I \tag{5}$$

This issue can be solved by a closed form solution. Suppose the singular value decomposition (SVD) of $(DC)C^T\varphi$ is given by UDV^T, then the closed form solution of matrix R can be computed by $R=UV^T$.

Thus the RSR is derived by optimizing two sub-problems iteratively. The low-dimensional feature DC is computed by $\varphi^T C$ based on the learned linear projection matrix φ. The number of non-zero elements of matrix φ is reduced by the orders of magnitude due to the sparse penalty. Finally, the cost of the linear projection is reduced dramatically because the complexities of linear projection in computation and memory are linear to the number of non-zero elements.

2.4 Feature-level fusion and feature-region fusion

2.4.1 Feature-level fusion

In this work, we choose the low frequency feature C_L and the vertical high-frequency feature C_V for feature-level fusion, and the detailed analysis is given in Section 3.2.1. Score-level fusion in previous work is more commonly used to improve the performance of the algorithm [7, 29]. But the literature e.g. [22] recognized that feature-level fusion could achieve a better performance than the ones using score-level fusion. So we introduce the feature-level fusion scheme for face recognition. The corresponding low frequency and vertical high-frequency features after the PL-RSR dimensionality reduction method are derived respectively by $DC_L=\varphi_L^T(C_L)$ and $DC_V=\varphi_V^T(C_V)$, where the $\varphi_L(\cdot)$ and $\varphi_V(\cdot)$ are calculated by the training phase of the PL-RSR dimensionality reduction method. In order to make full use of the low frequency and the vertical high-frequency information, the feature-level fusion is defined by

$$C_F = \begin{pmatrix} \varphi_L^T(C_L) \\ \varphi_V^T(C_V) \end{pmatrix} = \begin{pmatrix} DC_L \\ DC_V \end{pmatrix} \tag{6}$$

2.4.2 Feature-region fusion

It is a slightly different from the feature-region fusion technique for the training set and the testing set. The CPD non-rigid point set registration has been applied only to the non-rigid region in testing set. The feature-region fusion is defined by

For the training face

$$C_F=\omega_{N+CPD} \times DC_{NF} + \omega_S \times DC_{SF} \tag{7}$$

For the testing face

$$C_F=\omega_{N+CPD} \times DC_{(N+CPD)F} + \omega_S \times DC_{SF} \tag{8}$$

where $\omega_{N+CPD} = \frac{R_{1N+CPD}}{R_{1S}+R_{1N+CPD}}$ and $\omega_S = \frac{R_{1S}}{R_{1S}+R_{1N+CPD}}$ are the weight coefficients of the rigid and semi-rigid regions, respectively. R_{1S} and R_{1N+CPD} represent the rank-1 recognition rate of semi-rigid region and non-rigid facial region after CPD non-rigid registration, respectively. The weight coefficients for (7) and (8) are equivalent to each other due to the same experiment for the testing and training phase.

DC_{NF} and DC_{SF} in (7) are derived by the first step *i.e.* PL dimension reduction and feature-level fusion step. For (8), $DC_{(N+CPD)F} = \begin{pmatrix} \varphi_{(N+CPD)L}^T\left(C_{(N+CPD)L}\right) \\ \varphi_{(N+CPD)V}^T\left(C_{(N+CPD)V}\right) \end{pmatrix} = \begin{pmatrix} DC_{(N+CPD)L} \\ DC_{(N+CPD)V} \end{pmatrix}$ represents the extracted discriminative feature coefficient of non-rigid face region after CPD non-rigid registration and $DC_{SF} = \begin{pmatrix} \varphi_{SL}^T(C_{SL}) \\ \varphi_{SV}^T(C_{SV}) \end{pmatrix} = \begin{pmatrix} DC_{SL} \\ DC_{SV} \end{pmatrix}$ represents the extracted discriminative feature coefficient of semi-rigid face region.

Finally, the cosine similarity technique is utilized to compute the similarity between two faces. The 3D face recognition result will be derived by the nearest neighborhood classification algorithm.

3 Experiment results

In order to test the performance of our proposed approach, we conduct a set of identification experiments on the FRGC v2.0 dataset [33] and Bosphorus dataset [37]. In this section, we evaluate the performance of the proposed approach and present our recognition results.

3.1 Dataset description

FRGC v2.0 is constructed by the University of Notre Dame, and acquired by a Minolta Vivid 900/910 series sensor in the form dense point-clouds. Every scan composes of 80,000~130,000 3D point clouds, and every point is exclusively represented by x, y and z coordinate values. The FRGC v2.0 dataset consists of 4007 recordings of 466 individuals (Fall 2003 and Spring 2004) with neutral or non-neutral expressions. In our work, we perform the "Neutral vs. Others" recognition experiment, in other words, the neutral scan of each individual makes a gallery of 466 different individuals, and the other 3541 scans make up the probe faces.

The Bosphorus dataset consists of 4666 scans from 105 subjects and contains various poses, facial expressions and occlusion variations. In the same way, the neutral scan of each individual make a gallery of 105 different individuals, and the remaining 2797 scans with frontal pose and without occlusions and pose variations make up the probe. Both datasets is automatically preprocessed as described in Section 2.1.

3.2 Performance evaluation

3.2.1 Performance evaluation of feature selection and fusion

For the sake of convenience, the experimental results of the Section 3.2.1 and 3.2.2 are based on the FRGC v2.0. The low frequency feature C_L obtained by the top-level's wavelet transform is usually single-handed used for face recognition because the relevant authors take into account that it plays a significant role for an approximate representation of the original face. In fact, the vertical high-frequency feature also contains much important information by our repeatedly experiments. Although the vertical high-frequency component does not contain as much information as the low frequency component, it also contains very important information for the 3D face recognition. Thus, it is reasonable to fusion the information of the low frequency feature C_L and vertical high-frequency feature C_V for face recognition. In addition, the feature-level fusion $C_{(L+V)F}$ can make full use of the low frequency and the vertical high-frequency information. Table 1 shows the rank-1 recognition rate using the single feature based on the whole face in FRGC v2.0. The rank-1 recognition rate associated to the feature-level fusion is higher than that based on the single feature according to the experimental results of the Table 1. C_L, C_H, C_V and C_D represent the low-frequency, horizontal high-frequency, vertical high-frequency and diagonal high-frequency wavelet features of geometric image,

Table 1 The rank-1 recognition rate (PR) based on the whole face (Neutral vs. Others)

Different frequency	C_L	C_H	C_V	C_D	$C_{(L+V)F}$
The rank-1 PR	93.52 %	84.60 %	92.61 %	71.80 %	96.10 %

respectively. $C_{(L+V)}\ F = \left({C_L}^T \quad {C_V}^T \right)^T$ represents the feature-level fusion of low-frequency and vertical high-frequency features without dimension reduction.

3.2.2 Performance evaluation of the dimension reduction method

In this section, we compare the PL-RSR dimension reduction method based on the Principal Component Analysis and Linear Discriminant Analysis in combination with Rotated Sparse Regression for 3D face recognition in FRGC v2.0. For the PL-RSR method, the dimension of low-frequency and vertical high-frequency features is reduced to 400 and 340 by PL step, respectively. The recognition rate is chosen when the sparsity is 0.98. For the typically PCA dimension reduction method, the number of low frequency feature dimension reduction is obtained by the proportion of the principal component eigenvalues, *i.e.* the energy of principal components. The energy proportion is 0.94 from the repeated experiments. The same is to be the dimension of vertical high-frequency, and the energy proportion is 0.85. The dimension of low frequency and vertical high-frequency feature are reduced to 340 and 300 only utilizing the LDA as well, respectively. The PL-RSR method directly utilizes the original subspace method to compute the low-dimensional feature. Furthermore, the rotation term in PL-RSR method provides additional freedom and further advances the sparsity. The performance of the PL-RSR is better than two other methods based on Table 2.

3.2.3 Performance evaluation of the proposed feature-region fusion

The PL-RSR dimensionality reduction criterion in Section 3.2.2 is utilized in all the experimental evaluation based on FRGC v2.0 and Bosphorus. Figures. 7 and 8 represent the CMC curves for 3D face recognition results based on the whole face, semi-rigid facial region and the feature-region fusion in FRGC v2.0 and Bosphorus, respectively. The rank-1 recognition rate for the whole face, semi-rigid facial region and the feature-region fusion are 94.08, 95.01, 97.91 % in FRGC v2.0, and 93.06, 93.99, 96.17 % in Bosphorus respectively.

Practically, the performance of the proposed feature-region fusion strategy related to the PL-RSR dimensionality reduction is better than the one which directly use the whole face or semi-rigid facial region. 3D face recognition accuracy is easily susceptible to the facial expressions and the non-rigid facial region for instance the mouth is the most sensitive region for the facial expressions variations, therefore, the method based on semi-rigid face region has better performance than the whole face region from the CMC curves of Figs. 7 and 8. Therefore, the proposed feature-region fusion scheme achieves the best recognition performance with 97.91 and 96.17 % in FRGC v2.0 and Bosphorus datasets comparing to two other methods, respectively. This is because we apply non-rigid point set registration to non-rigid region which can eliminate the influence from the facial expressions. Simultaneously, the feature-region fusion approach can guarantee the complete feature information of the whole face.

Table 2 The rank-1 recognition results about three methods on the whole face in FRGC v2.0

Methods	PCA	LDA	PL-RSR
Rank-1 recognition rate	91.27 %	92.13 %	94.08 %

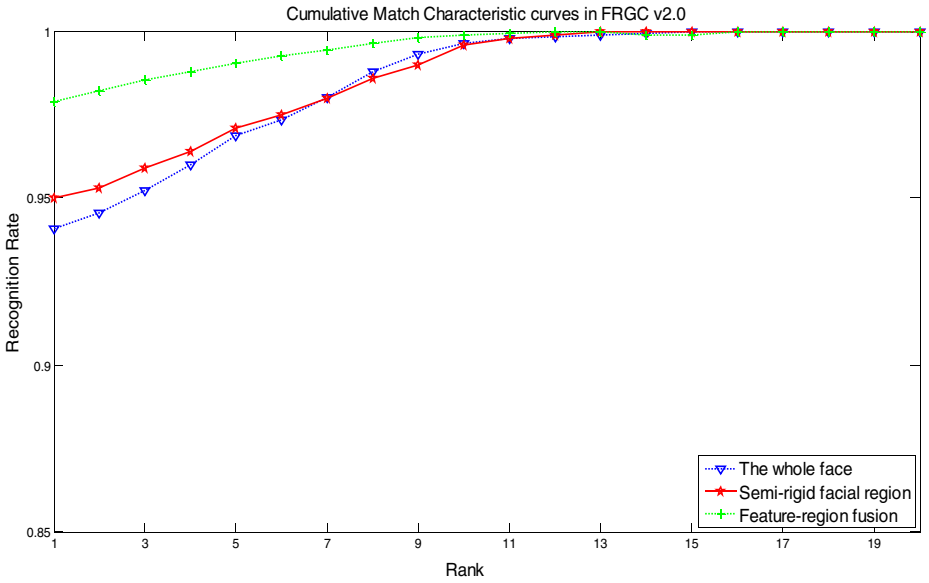

Fig. 7 CMC curves in FRGC v2.0

3.3 Performance comparison and analyses

In this part, we present a comparison result with *state-of-the-art* works in FRGC v2.0 and Bosphorus datasets (Table 3).

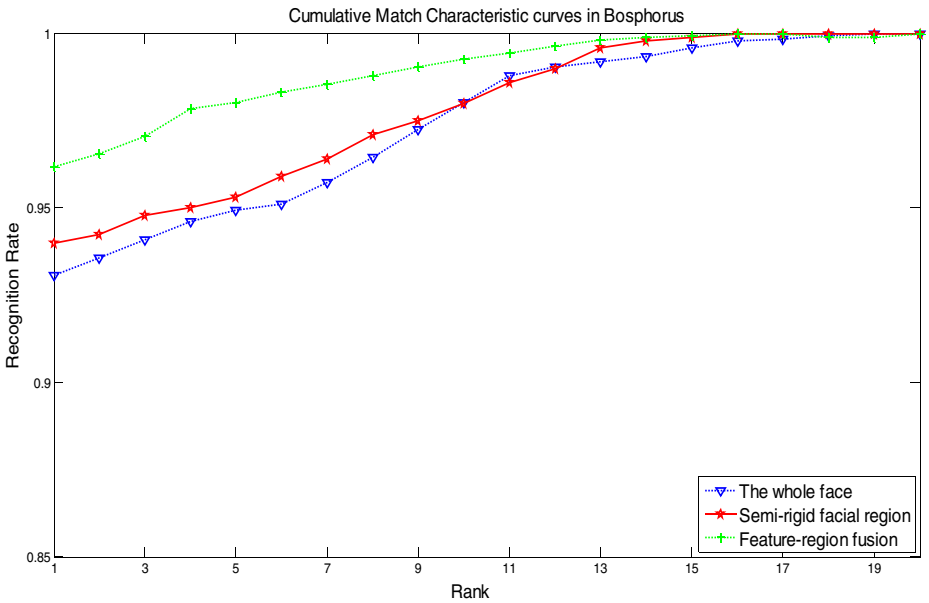

Fig. 8 CMC curves in Bosphorus

Table 3 Comparison of the *state-of-the-arts* methods based on the rank-1 recognition rate

Method	Year	FRGC v2.0		Bosphorus		
		Rank-1	Type	Rank-1	Gallery	Probe
Lei et al. [21]	2014	94.10 %	Nonneutral vs. Neutral	–	–	–
Li et al. [23]	2012	97.80 %	First vs. Others	–	–	–
Kakadiaris et al. [19]	2007	97.00 %	First vs. Others	–	–	–
Smeets et al. [38]	2013	89.63 %	First vs. Others	93.7 %	105	4351
Cai et al. [6]	2012	97.50 %	First vs. Others	–	–	–
Faltemier et al. [11]	2008	97.20 %	First vs. Others	–	–	–
Mian et al. [30]	2008	93.50 %	Neutral vs. Others	–	–	–
Queirolo et al. [34]	2010	98.40 %	First vs. Others	–	–	–
Li et al. [24]	2014	96.30 %	First vs. Others	95.4 %	105	2797
Alyüz et al. [1]	2010	97.50 %	First vs. Others	98.2 %	105	2814
Our method	–	97.91 %	Neutral vs. Others	96.17 %	105	2797

Lei et al. [21] A novel facial signature called Angular Radial Signature (ARS) is extracted from the semi-rigid region of the face and used for 3D face recognition.

Li et al. [23] A rejection classifier is proposed by using facial curves which can produce a facial deformation mapping and then an adaptive region selection scheme is used for handling facial expression and hair occlusion.

Kakadiaris et al. [19] A fully automatic method is developed by using a composite alignment algorithm to register 3D facial scans with an Annotated Deformable Model.

Smeets et al. [38] The meshSIFT algorithm and its application under expression variations and partial data are presented for 3D face recognition.

Cai et al. [6] A multi-scale model by approximating expression changes is proposed, which decomposes a face into three components according to its frequency domains. The Low-frequency components are used to estimate the intra-personal residue for face recognition.

Faltemier et al. [11] A new 3D face recognition method is proposed based on score-level fusion resulting from 28 small individual region matching.

Mian et al. [30] The point cloud geometry of the facial local region is used to extract the keypoints, and the recognition rate is derived by the feature-level fusion and decision-level fusion techniques in combination with extracted 3D and 2D features, respectively.

Queirolo et al. [34] The intelligent algorithm annealing-based approach (SA) is used for image registration. The surface interpenetration measure is utilized as a similarity measure to match two face images. The finally authentication score is obtained by combining the SIM values corresponding to the matching of four different face regions: circular and elliptical areas around the nose, forehead, and the entire face region.

Li et al. [24] A novel highly discriminative multi-scale and multi-component local normal patterns (MSMC-LNP) facial shape descriptor is proposed. And a weighted sparse representation-based classifier (W-SRC) is used to identify the probe.

Alyüz et al. [1] The component-based regional registration methodology with the help of a generic face model and generic region models is utilized.

The highest 3D face recognition rate is derived in [24]. The possible reason is that Queirolo et al. [24] repeatedly uses of multiple regions includes the entire face region. The corresponding feature information is more than the extracted one in our proposed approach. This may benefit the enhancement of the feature representation and improvement of the 3D face recognition rate.

Holistic 3D face recognition algorithms are sensitive to expressions, occlusions and pose variations. Essentially, the statistical model [19] derived by learning various facial expressions is improper for the 3D face recognition because the generalized model cannot describe anyone accurately. The assumption is that 3D face can be considered as a rigid object and its associated face surface deformation is considered as the isometric transformation [6], which is not also improper because the 3D face surface is essentially a non-rigid surface. Partially, rigid region-based algorithms are robust to facial expression variations [11, 21, 23, 30, 38], however, it lost some useful feature information of non-rigid facial region result that the information of the whole face is not utilized completely.

Therefore, a novel expression-robust 3D face recognition method is proposed by using feature-level fusion technology based on the low frequency and the vertical high-frequency Haar wavelet features and feature-region fusion technology associated to semi-rigid region and non-rigid region after CPD non-rigid registration, which can both take full advantage of the feature information of the whole face and extract the discriminative feature information of the whole face even under expression variations. By comparison, our proposed algorithm outperforms other *state-of-the-arts* methods from Table 3.

First vs. Others The first scan per subject in the complete dataset is considered as gallery, the other scans are the probes.

Neutral vs. Others The neutral scan of each individual makes a gallery of 466 different individuals, and the other 3541 scans make up the probe faces.

Non-neutral vs. Neutral One thousand five hundred ninety seven facial scans out of 4007 with non-neutral expressions including surprise, happy, puffy cheeks, and anger are chose as gallery. The other 2410 scans with a neutral expression make up the probe.

4 Conclusions

This paper proposes a novel approach to eliminate the expression influence from the non-rigid facial region and improve the accuracy of the 3D face recognition, and develops a novel automatic 3D face recognition approach based on the feature-level fusion and the feature-region fusion schemes after using the PL-RSR dimension reduction. Firstly, in order to separate handing the expression-sensitive region, 3D

face are divided into the semi-rigid and non-rigid regions, and the CPD technique is utilized to develop a non-rigid point set registration for non-rigid region. Secondly, the feature-level fusion technique ensures the low frequency and the vertical high-frequency features are fully used, which typically avoids the discriminative feature information may be lost. Thirdly, the feature-region fusion technique associated to the semi-rigid region and non-rigid region after CPD non-rigid point set registration are used to extract the discriminative wavelet feature of whole face under expression variations. Fourthly, the PL-RSR dimensionality reduction method is proposed to reduce the dimensionality of wavelet features to satisfy the requirement of the real-time application. Finally, the recognition rate of our method in experimental evaluation based on Matlab 2012b and Visual C++ 2010 indicates that this approach is always higher than those of the *state-of-the-arts* methods. It is noting that the face needed identification has large occlusion, such as great mustache with large facial expression, is hard to be recognized. Under this condition, 3D face is approximately equivalent to be covered. Thus, the facial expression cannot be eliminated effectively due to the occlusion. As a result, the proposed framework performs an effective 3D face recognition approach, so as to define more accurate and efficient 3D face recognition method. How to improve the rank-1 recognition rate using the optimization of the processing framework is one of main purposes of what is attempted to be addressed in our further work.

Acknowledgments The authors would like to thank the editorial office and five anonymous reviewers who gave valuable comments and helpful suggestions which greatly improved the quality of the paper.

References

1. Alyuz N, Gokberk B, Akarun L (2010) Regional registration for expression resistant 3-D face recognition. IEEE Trans Inform Forensics Sec 5(3):425–440
2. Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Trans Patt Anal Mach Intell 19(7):711–720
3. Besl PJ, McKay ND (1992) A method for registration of 3-D shapes. IEEE Trans Patt Anal Mach Intell 14(2):239–256
4. Bronstein AM, Bronstein MM, Kimmel R (2007) Expression-invariant representations of faces. IEEE Trans Image Process 16(1):188–197
5. Cai L, Da FP (2012) Nonrigid-deformation recovery for 3D face recognition using multiscale registration. Comput Graph Applic 32(3):37–45
6. Cai L, Da FP (2012) Estimating inter-personal deformation with multi-scale modelling between expression for three-dimensional face recognition. IET Comput Vis 6:468–479
7. Chang KI, Bowyer KW, Flynn PJ (2006) Multiple nose region matching for 3D face recognition under varying facial expression. IEEE Trans Patt Anal Mach Intell 28(10):1695–1700
8. Chen D, Cao XD, Fen F, and Sun J (2013) Blessing of dimensionality: high-dimensional feature and its efficient compression for face verification, in: Proc Conf Comput Vis Patt Recognit. 3025–3032
9. Daoudi M, Srivastava A, Veltkamp R (2013) 3D face modeling, analysis and recognition. Wiley, Chichester
10. Drira H, Ben Amor B, Srivastava A, Daoudi M, Slama R (2013) 3D face recognition under expressions, occlusions, and pose variations. IEEE Trans Patt Anal Mach Intell 35(9):2270–2283
11. Faltemier TC, Bowyer KW, Flynn PJ (2008) A region ensemble for 3-D face recognition. IEEE Trans Inform Forensics Sec 3(1):62–72
12. Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. J Stat Softw 33:1–22

13. Gökberk B, Salah AA, Alyüz N, Akarun L (2009) 3D face recognition: technology and applications, Handbook of Remote Biometrics. 217–246
14. Gu XF, Gortler SJ, Hoppe H (2002) Geometry images. ACM Trans Graph (TOG) 21(3):355–361
15. Gupta P, Zaroliagis C, Maiti S, Sangwan D, Raheja JL (2014) Expression-invariant 3D face recognition using K-SVD method, applied algorithms. 266–276
16. Hiremath PS, Hiremath M (2014) 3D face recognition based on radon transform, PCA, LDA using KNN and SVM. Int J Comput Network Inform Sec 7:36–43
17. Huttenlocher DP, Klanderman GA, Rucklidge WJ (1993) Comparing images using the Hausdorff distance. IEEE Trans Patt Anal Mach Intell 15(9):850–863
18. Jolliffe IT (1986) Principal component analysis, 1st edn. Springer, New York
19. Kakadiaris IA, Passalis G, Toderici G, Murtuza MN, Lu Y, Karampatziakis N, Theoharis T (2007) Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach. IEEE Trans Patt Anal Mach Intell 29(4):640–649
20. Lee YH, Shim JC (2004) Curvature based human face recognition using depth weighted hausdorff distance, in: Proceed Int Conf Imag Process. 1429–1432
21. Lei YJ, Bennamoun M, Hayat M, Guo YL (2014) An efficient 3D face recognition approach using local geometrical signatures. Pattern Recogn 47(2):509–524
22. Lei YJ, Bennamoun M, El-Sallam AA (2013) An efficient 3D face recognition approach based on the fusion of novel local low-level features. Pattern Recogn 46(1):24–37
23. Li XL, Da FP (2012) Efficient 3D face recognition handling facial expression and hair occlusion. Image Vis Comput 30(9):668–679
24. Li HB et al (2014) Expression-robust 3D face recognition via weighted sparse representation of multi-scale and multi-component local normal patterns. Neurocomputing 133:179–193
25. Lin WY, Chen MY (2014) A novel framework for automatic 3D face recognition using quality assessment. Multimed Tools Applic 68:877–893
26. Liu YH (2004) Improving ICP with easy implementation for free-form surface matching. Pattern Recogn 37(2):211–226
27. Lu XG, Jain AK, Colbry D (2006) Matching 2.5D face scans to 3D models. IEEE Trans Patt Anal Mach Intell 28(1):31–43
28. Mallat SG (1989) A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans Patt Anal Mach Intell 11(7):674–693
29. Mian AS, Bennamoun M, Owens R (2007) An efficient multimodal 2D-3D hybrid approach to automatic face recognition. IEEE Trans Patt Anal Mach Intell 29(11):1927–1943
30. Mian AS, Bennamoun M, Owens R (2008) Keypoint detection and local feature matching for textured 3D Face recognition. Int J Comput Vis 79:1–12
31. Mohammadzade H, Hatzinakos D (2013) Iterative closest normal point for 3D face recognition. IEEE Trans Patt Anal Mach Intell 35(2):381–397
32. Myronenko A, Song X (2010) Point set registration: coherent point drift. IEEE Trans PattAnal Mach Intell 32(12):2262–2275
33. Phillips PJ, Flynn PJ, Scruggs T, Bowyer KW (2006) Overview of the face recognition grand challenge, in: Proc Conf Comput Vis Patt Recognit. 947–954
34. Queirolo CC et al (2010) 3D face recognition using simulated annealing and the surface interpenetration measure. IEEE Trans Patt Anal Mach Intell 32(2):206–219
35. Ritter J (2002) Wavelet based image compression using FPGAs. Martin-Luther-University Halle, Wittenberg
36. Russ TD, Koch MW, Little CQ (2005) A 2D range Hausdorff approach for 3D face recognition, in: Proc IEEE Comput Soc Conf Comput Vis Patt Recognit-Workshops. 169–176
37. Savran A. et al (2008) Bosphorus database for 3D face analysis, biometrics and identity management. 47–56
38. Smeets D, Keustermans J, Vandermeulen D, Suetens P (2013) meshSIFT: Local surface features for 3D face recognition under expression variations and partial data. Comput Vis Image Underst 117(2):158–169
39. Turk MA, Pentland AP (1991) Face recognition using eigenfaces, in: Proc Conf Comput Vis Patt Recognit. 586–591
40. Wang XQ, Ruan QQ, Ming Y (2010) 3D face recognition using corresponding point direction measure and depth local features, in: Proc 10th Int Conf Signal Process. 86–89
41. Yi J, Wang YZ, Ruan QQ, Wang XQ (2011) A new scheme for 3D face recognition based on 2D Gabor Wavelet Transform plus LBP, in: Proc 6th Int Conf Comput Sci Educ. 860–865
42. Yue M (2013) Rigid-area orthogonal spectral regression for efficient 3D face recognition. Neurocomputing 129:445–457
43. Zhang LY, Razdan A, Farin G, Femiani J, Bae M, Lockwood C (2006) 3D face authentication and recognition based on bilateral symmetry analysis. Vis Comput 22(1):43–55

Xing Deng received her M.Sc. degree in applied mathematics from the Jiangsu University, China, in 2011. She is now a PhD candidate in Southeast University's Department of Automation. Her research interests include pattern recognition and digital geometry processing, with applications to biometrics and computer vision. Contact her at xingdeng@seu.edu.cn.

Feipeng Da received his Ph.D. in automatic control engineering from Southeast University, China, in 1998. He is now a professor in Southeast University's Department of Automation. His research interests include 3D information acquisition and recognition, with applications to biometrics, computer vision, and intelligent control. Da has a PhD in control theory and control engineering from Southeast University. Contact him at dafp@seu.edu.cn.

Haijian Shao received her M.Sc. degree in applied mathematics from the Jiangsu University, China in December 2010. Now he is a Ph.D. degree holder from the Key Laboratory of Measurement and Control for Complex Systems of the Ministry of Education, Research Institute of Automation, Southeast University, Nanjing, China. He has been engaged in research in extensive areas of applied mathematics and mathematical engineering, dynamical systems, control theory and control engineering, differential geometry, functional analysis theory, pattern recognition, and intelligent systems. Contact him at haijianshaoseu@gmail.com