SIViP (2017) 11:1305–1312
DOI 10.1007/s11760-017-1087-6

ORIGINAL PAPER

Adaptive feature selection based on reconstruction residual and accurately located landmarks for expression-robust 3D face recognition

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

Received: 25 March 2016 / Revised: 3 December 2016 / Accepted: 23 March 2017 / Published online: 5 April 2017

Abstract A novel adaptive feature selection based on reconstruction residual and accurately located landmarks for expression-robust 3D face recognition is proposed in this paper. Firstly, the novel facial *coarse-to-fine* landmarks localization method based on Active Shape Model and Gabor wavelets transformation is proposed to exactly and automatically locate facial landmarks in range image. Secondly, the multi-scale fusion of the pyramid local binary patterns (F-PLBP) based on the irregular segmentation associated with the located landmarks is proposed to extract the discriminative feature. Thirdly, a sparse representation-based classifier based on the adaptive feature selection (A-SRC) using the distribution of the reconstruction residual is presented to select the expression-robust feature and identify the faces. Finally, the experimental evaluation based on FRGC v2.0 indicates that the adaptive feature selection method using F-PLBP combined with the A-SRC can obtain the high recognition accuracy by performing the higher discriminative power to overcome the influence from the facial expression variations.

Keywords 3D face recognition · Facial landmark localization · Multi-scale fusion · Adaptive feature selection

✉ Feipeng Da
dafp@seu.edu.cn

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

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

1 Introduction

Although 2D face recognition has made great progress in past two decades [1], its recognition rate is still influenced by many factors such as illumination, gestures, expressions and age variations. With the development of 3D measurement technology, people can quickly and accurately acquire the 3D objet data with more geometric information [2]. So, the 3D face recognition technology has gradually developed and attracted the researchers extensive concern in recent decades. In order to automatically and accurately locate facial landmarks, extract the discriminative features and derive the better recognition accuracy, this paper proposes a novel expression-robust 3D face recognition method from the following aspects: landmark localization, feature extraction and classifier.

1.1 Landmark localization

Exploring a method of automatic landmark detection with robustness and high accuracy is a challenging task, which plays an important role in applications such as face registration, face segmentation and face recognition. Motivated by the practical applications and extensive work focusing on thinking of methods for automatic landmark localization, the Active Shape Model (ASM) proposed by Cootes et al. [3] can guarantee the good performance on 2D faces. However, this algorithm suffers from illumination variation and facial expression changes. Compared to landmark detection on 2D facial images, automatic landmark detection on 3D face is a new research topic. The salient points were detected, and the multi-scale local surface descriptors were built based on curvelet transform by Elaiwat et al. [4], which can be used to obtain the highly distinctive rotation invariant local features around the detected keypoints. In their later work [5],

a multimodal Curvelet-based approach for textured 3D face recognition was proposed to extract both texture and 3D local features. Jahanbin et al. [6] proposed a 2D and 3D multimodal algorithm based on Gabor features to achieve a higher accuracy.

1.2 Feature extraction

Generally, the feature is a stable attributes from one object in different state, which can be used to distinguish from the feature information related to other objects. The quality of the extracted features will directly affect the final recognition results, so it is crucial to extract robust feature to overcome the influence from the expression variations. According to the types of feature extraction, the current 3D face recognition algorithms can be broadly divided into two categories: local features and holistic features methods. 3D face recognition method based on the local features usually depends on the partial 3D facial region. The currently popular method is based on the local features extracted from a number of distinctive points [4–7]. Ouamane et al. [8] utilized the (3D+2D) multimodal score level fusion of the extracted feature for face authentication, and Okuwobi et al. [9] proposed two novel fusion schemes where the first one employed a confidence-aided combination approach, and the second one implemented a two-level serial integration method, whereas the other 3D face recognition algorithm based on the holistic features typically focused on the whole [10].

1.3 Classifier

Wright et al. [11] firstly proposed sparse representation-based classification (SRC) based on the compressed sensing theory [12] and sparse representation method to implement the identification with good generality ability. Recently, a number of research laboratories have shown confidence in the application of the computer vision that the SRC and its improved patterns can be used to obtain the higher recognition rate [7,13].

The rest of this paper is constructed as follows. The proposed approach including 3D facial preprocessing, facial landmarks localization, irregular landmark-based patch facial representation, F-PLBP and classifier is given in Sect. 2. The experimental evaluation based on FRGC v2.0 is given in Sect. 3 to demonstrate the performance of the proposed approach, and this paper is concluded in Sect. 4.

2 Proposed approach

The expression variation is one of the main challenges in 3D face recognition because the facial geometry shape changes drastically under expression variations. In order to overcome the effects of the expression variations, an adaptive feature selection method based on reconstruction residual and accurately located landmarks for expression-robust 3D face recognition is proposed. The main contributions of this paper can be summarized as follows: (1) A novel facial *coarse-to-fine* landmark localization method based on ASM and Gabor Wavelets Transformation (GWT) in range image is proposed to automatically and accurately locate facial landmarks. (2) The irregular landmarks-based patch facial representation is proposed to keep the integrality of face local structure and get more detailed feature information. (3) A novel F-PLBP feature is proposed to obtain the discriminative features and achieve accurate facial representation. (4) An adaptive feature selection method using the reconstruction residual and accurate located landmarks is proposed. The block diagram of the proposed approach is given in Fig. 1.

Firstly, the 3D facial preprocessing including the hole-filling, smoothing, nose tip detection, face cropping and pose correction is proposed to improve the data quality for further analysis. Secondly, a novel *coarse-to-fine* facial landmark localization method based on ASM and GWT in preprocessed range image is presented to accurately locate facial landmarks. Thirdly, the irregular segmentation according to the located 8 facial landmarks is proposed to extract more detailed feature information. Fourthly, the F-PLBP for 3D face representation is presented. Fifthly, A-SRC is used for the identification, and its performance comparisons with the SRC as well as the W-SRC are given. Finally, the experimental evaluation based on the FRGC v2.0 is constructed to demonstrate the performance of the proposed approach.

2.1 3D face preprocessing

2.1.1 Face preprocessing

The raw 3D facial scan in FRGC v2.0 [14] dataset is represented by 3D point clouds without any preprocessing. Prior to 3D face recognition, appropriate preprocessing steps are used for further analysis. An automatically preprocessing algorithm proposed in our previous work [15] is used, which includes hole-filling, smoothing, nose tip detection, face cropping and pose correction. In addition, a basic Iterative Closest Point (ICP) algorithm is used in the face registration.

2.1.2 Range image and normalization

In range image of the preprocessed facial data, the pixel value of every point indicates the distance between the actual point and the camera. The range images are computed by interpolating at the integer x and y coordinates along the horizontal and vertical index, respectively, and determining the corresponding z coordinate as a pixel value. The pixels in range image are then resampled at a distance of 1 mm along both the x and y directions. The range map values are linearly

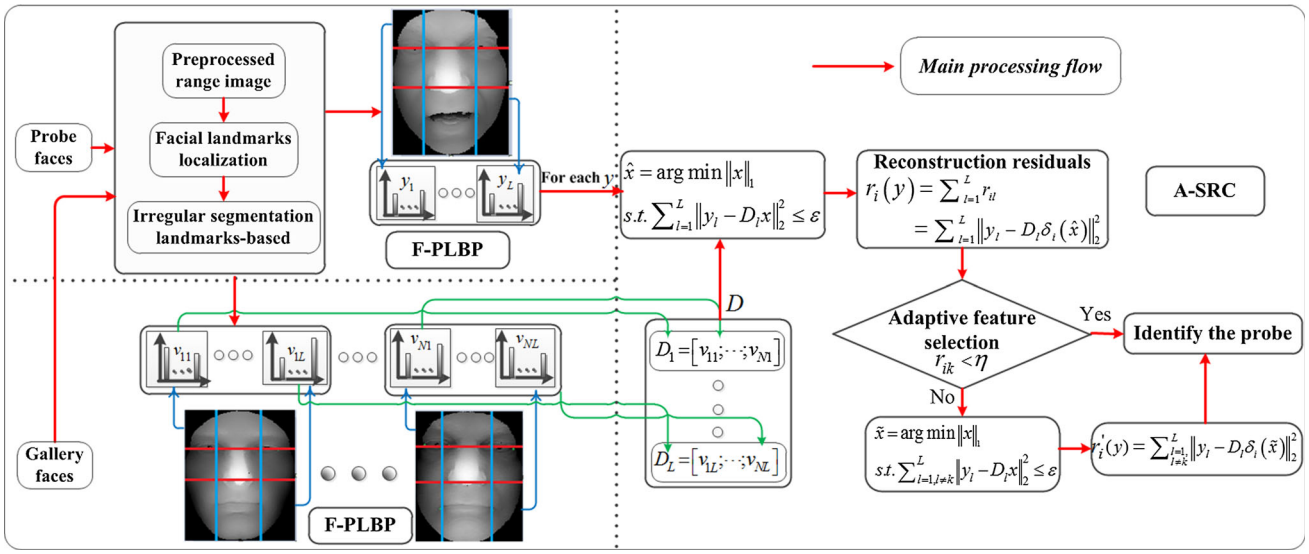

Fig. 1 Block diagram of the proposed approach

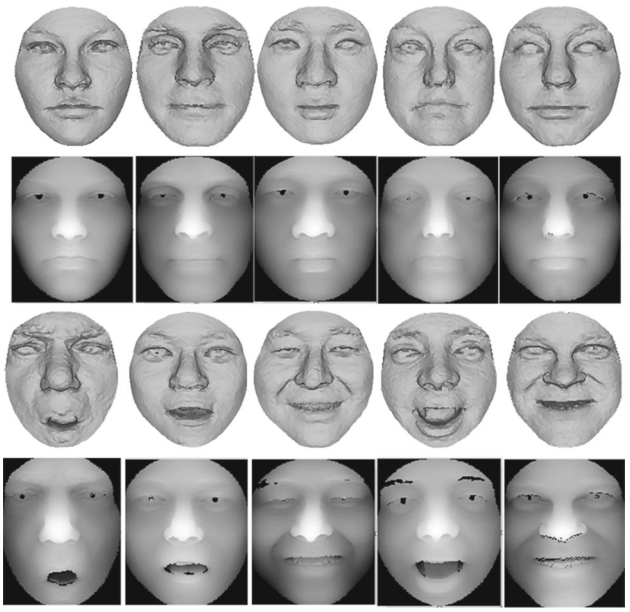

Fig. 2 Several examples of preprocessed range images

scaled to 0–255. All the depth images are normalized into a certain size 400×400 centered at the detected nose tip to build a public coordinate system. Then, each preprocessed range image is resized into 120×150 for further processing. In Fig. 2, the first and third rows are the 3D faces with different expressions, and the second and fourth rows are the corresponding preprocessed range images.

2.2 Facial landmarks localization

In order to automatically and accurately locate facial landmarks, our previous work proposed a novel facial *coarse-to-fine* landmark localization method based on ASM and GWT in range image [16]. The landmarks localization description is briefly described as follows:

(1) Coarse area localization. ASM is implemented to acquire landmarks coarse areas. A 25×25 pixel search area for each landmark is obtained based on the results of ASM. The range image can be described by n landmarks constructed by the shape vector $x = \left[p_{x,1}, \ldots, p_{x,n}, \ p_{y,1}, \ldots, p_{y,n}\right]^T$, where $\left(p_{x,i}, p_{y,i}\right), i = 1, \ldots, n$ is the coordinate in range image. In another way, a shape can be approximated by a shape model: $x \approx \bar{x} + \Phi b$, where $\bar{x}$ means the mean shape of all objects in the ASM training set; the eigenvectors corresponding to the s largest eigenvalues $(\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_s)$ $\lambda_j, j = 1, \ldots, s$ are retained in a matrix $\Phi = \{\Phi_1, \ldots, \Phi_s\}$; b is the model parameter to be solved. In order to get the parameter b, assume the profile of training sample in local texture model is set as $g_k, k = 1, \ldots, t$ and obey the multidimensional Gaussian distribution in ASM, the Mahalanobis distance among the profiles is defined by $f(g_k) = (g_k - \bar{g})^T C_g^{-1} (g_k - \bar{g})$, where $\bar{g}$ and C_g are the mean profile and covariance matrix of the searched landmark. Both shape model and profile model are combined to search each landmark, and points with the minimum values of $f(g_k)$ are the specified landmarks.

(2) Fine landmark detection. Since ASM performs poor in situations such as eye-closing and facial expression variations, we extract a pixel area for each landmark centered at its coarse location as the fine search area. Then GWT is applied to these search area to get the final location of each landmark. The similarity maps are obtained by calculating the similarity between sets of "Gabor jets" at the initial coarse positions and sets of "Gabor bunches" modeled by its corresponding manually marked landmarks in the training set. The point with the maximum value in each

similarity map is extracted as its final facial feature location.GWT is typically the response of the range image $I(x)$ associated with Gabor filters, which is usually defined by the convolution $J_j(x) = \int_j I(x')\phi(x - x')d^2x'$. $J_j(x)$ can also be expressed by $J_j(x) = \alpha_j(x)\exp(i\varphi_j(x))$, where $\alpha_j(x)$ and $\varphi_j(x)$ represent the magnitude and phase, respectively. For any given landmark j, a "Gabor bunch" set is defined as a set constructed by the "Gabor jet" of j and its 8 neighborhood; "Gabor bunch" set is a set constructed by all different images' "Gabor bunch" set of j in the training set. Take LEIC for example, its "Gabor jet" set is $\mathbf{J}_s = \{\mathbf{J}_0; \mathbf{J}_i, i = 1, \ldots, 8\}$, where $\mathbf{J}_0$ is the "Gabor jet" of LEIC, $\mathbf{J}_i$ is the "Gabor jet" of its 8 neighborhoods. $\mathbf{B}' = \{\mathbf{J}_{s,i}, i = 1, \ldots, n\}$ is the "Gabor bunch" set of LEIC, where $\mathbf{J}_{s,i}$ means different images' "Gabor jet" set of LEIC in the training set.

2.3 Irregular landmark-based patch facial representation

In order to get more details of facial expressions, the patch-based face representation method has roused the researchers extensive attention in recent years. The current mainly face segmentation methods can be divided into two categories: The one is based on face model, which is usually segmented into high-level semantic parts, such as the forehead, nose, cheek and chin [17]. This method will sacrifice some useful information which may be useful for face recognition. The other one is based on the regular segmentation [13,18], which only considers the size of the face image and neglect the face geometry structure. So in order to make full use of the information of whole face and keep the integrality of face local structure, this paper proposes the irregular segmentation according to the located 8 facial landmarks in Sect. 2.2. Three different irregular segmentation methods are shown in Fig. 3. LNC: left nose corner; RNC: right nose corner; LMC: left mouth corner; RMC: right mouth corner; LEIC: left eye inner corner; LEOC: left eye outer corner; REIC: right eye inner corner; REOC: right eye outer corner. In order to facilitate subsequent feature extraction and classification, the divided irregular patch facial subregion is numbered according to the principle of from top to bottom and from left to right,

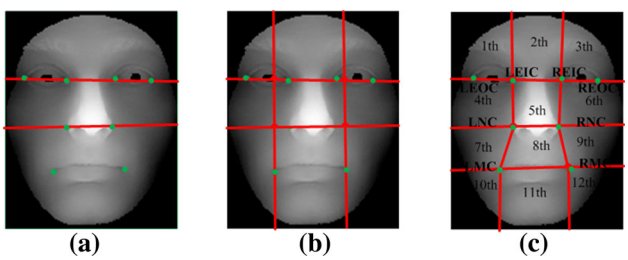

Fig. 3 The irregular segmentation based on landmarks

for example Fig. 3c. The details of the irregular segmentation are described as follows: (1) In horizontal direction, from top to down: The three lines are determined by the LEOC and REOC, LNC and RNC, and LMC and RMC, respectively; (2) In vertical direction, from left to right: Firstly, two rays are determined by the LNC and LEIC, and RNC and REIC, respectively; secondly, two line segments are determined by LNC and LMC, and RNC and RMC, respectively; finally, two vertical rays based on the line LMC and RMC are determined by LMC and RMC, respectively. Two horizontal lines in Fig. 3a, b are same as the position in Fig. 3c. Two vertical direction rays in Fig. 3b are determined by LMC and RMC, respectively.

2.4 PLBP for 3D face representation

Local Binary Patterns (LBP) was proposed by Ojala et al. [19], which had been widely used in the 2D face recognition due to its powerful ability for feature extraction and texture classification. Qian et al. [20] had proved that the pyramid LBP (PLBP) was a powerful approach for local feature representation. Moreover, LBP and its improved patterns have gradually been used for 3D face recognition in recent years [13,21]. So, we propose the F-PLBP for 3D face representation.

The Gaussian and Laplacian are the typically pyramid algorithms, which are mainly about the low pass filter and band pass filter, respectively. Generally, the *cutoff* frequency related to the low pass filter generated by the Gaussian pyramid algorithm is controlled by the Gaussian kernel parameter σ. The purpose of the Laplacian pyramid is continued to derive a set of the band pass filter images related to the two levels of the Gaussian pyramid. In this paper, LBP in Gaussian and Laplacian pyramid domain is used. The Gaussian (Laplacian) pyramid descriptor $G(L)$ is the combination of the extracted feature at all resolution levels, which is expressed as follows:

$$\begin{cases} G = \langle G^1, \ldots, G^n \rangle = G\left(g_c^1, g_1^0, \ldots, g_1^{P-1}; \ldots; g_c^n, g_n^0, \ldots, g_n^{P-1}\right), \\ L = \langle L^1, \ldots, L^m \rangle = L\left(l_c^1, l_1^0, \ldots, l_1^{P-1}; \ldots; l_c^m, l_m^0, \ldots, l_m^{P-1}\right), \\ G^e = G\left(g_c^e, g_e^0, \ldots, g_e^{P-1}\right) \\ \quad \approx G\left(s(g_e^0 - g_c^e), \ldots, s(g_e^{P-1} - g_c^e)\right), \\ L^t = L\left(l_c^t, l_t^0, \ldots, l_t^{P-1}\right) \\ \quad \approx G\left(s(l_t^0 - l_c^t), \ldots, s(l_t^{P-1} - l_c^t)\right), \end{cases} \tag{1}$$

where $G^e(L^t)$ represents the information of the eth(tth) Gaussian (Laplacian) pyramid ($e = 1, \ldots, n$; $t = 1, \ldots, m$), and $g_c^e(l_c^t)$ denotes the corresponding center pixel. $G -$

$PLBP_{P,R,e}\left(L-PLBP_{P,R,t}\right)$ and $F-PLBP$ are, respectively, given by

$$\begin{cases} G-PLBP_{P,R,e}=\sum_{p=0}^{P-1}\left(s\left(g_e^p-g_c^e\right)2^p\right) \\ L-PLBP_{P,R,t}=\sum_{p=0}^{P-1}\left(s\left(l_t^p-l_c^t\right)2^p\right) \end{cases} \tag{2}$$

$$F(PLBP_{P,R}) = \left\langle \cup_e G-PLBP_{P,R,e}; \cup_t L-PLBP_{P,R,t} \right\rangle \tag{3}$$

where $s(x)=\begin{cases}1, x\geq 0\\ 0, x<0\end{cases}$, $G-PLBP_{P,R,e}\left(L-PLBP_{P,R,t}\right)$ denotes the LBP of a pixel at eth(tth) Gaussian (Laplacian) pyramid. F-PLBP is the multi-scale fusion of the G-PLBP and L-PLBP histograms of the all spatial pyramid images. The patterns of binary will decrease from 256 to the same as the 59 uniform LBP.

2.5 Classifier

In this paper, an adaptive feature selection strategy based on reconstruction residual and accurately located landmarks for 3D face recognition is proposed. In order to verify the effectiveness of the proposed approach, both of the SRC and the W-SRC are used to compare the experimental results.

2.5.1 SRC and W-SRC

SRC. Wright et al. [11] introduced the SRC into 2D face recognition which subsequently had been widely used in 3D face recognition. The testing sample is sparse linear represented by all the training samples in SRC, and the linear reconstruction residuals with respect to each class is used to determine the identity of the input face scan. Assume $D_i=\left[v_{i1},\ldots,v_{iN_i}\right]\in R^{M\times N_i}$ is the feature vector set with N_i training samples of the ith class, where $v_{ij}\in R^M$ is the feature vector of jth sample of ith class. $y=Dx\in R^M$ is used to represent the testing sample based on the sparse coefficient vector x and dictionary matrix D developed by concatenating $D_i, i=1,\ldots,k$. The solution of $y=Dx$ is calculated by

$$\hat{x}=\arg\min\|x\|_1\ s.t.\quad \|Dx-y\|_2^2\leq\varepsilon \tag{4}$$

where $\|x\|_1$ denotes the L_1-Norm and $s.t.$ is the abbreviation of 'subject to'. The identity of a testing sample y is determined by the one with the smallest residual obtained by $r_i(y)=\left\|y-D\delta_i\left(\hat{x}\right)\right\|_2^2$.

W-SRC. The facial distortions under expression variations differ in different patches. To make the proposed method robust to the variations of facial expression, Li et al. proposed the W-SRC [13]. The weights w_l of L different patches satisfy $\sum_{l=1}^{L}w_l=1$, which are learned by the recognition performance of single local patch, where $w_l=\frac{R1_l}{\sum_{l=1}^{L}R1_l}$ in this paper, $R1_l$ is the rank-1 recognition rate of single patch $l=1,\ldots,L$. Typically, (4) is transformed into

$$\hat{x}=\arg\min\|x\|_1\ s.t.\ \sum_{l=1}^{L}w_l\left\|y_l-D_lx\right\|_2^2\leq\varepsilon \tag{5}$$

where D is the dictionary composed by $D=[D_1;\ldots;D_L]$, $D_l=[v_{1l},\ldots,v_{il},\ldots,v_{Nl}]$ calculated by the feature vector v_i combing L different patches. $v_{il}\in R^m$ is the lth patch feature vector of ith face. The corresponding weight reconstruction residuals are defined by

$$\begin{aligned} r_i(y) &= \sum_{l=1}^{L}r_{il} \\ &= \sum_{l=1}^{L}w_l\left\|y_l-D_l\delta_i\left(\hat{x}\right)\right\|_2^2,\quad i=1,\ldots,N \end{aligned} \tag{6}$$

Based on (4), (5) and (6), W-SRC [13] is defined by

$$\hat{x}=\arg\min\|x\|_1\ \ s.t.\ \ \|y_w-D_wx\|_2^2\leq\varepsilon \tag{7}$$

where $D_w=[w_1D_1;\ldots;w_LD_L]$, $y_w=[w_1y_1;\ldots;w_Ly_L]$.

2.5.2 A-SRC

Although W-SRC can derive the better recognition rate comparing to SRC, its performance still fail in some faces with large facial expression. Therefore, A-SRC is proposed based on the distribution of the reconstruction residual to select the expression-robust feature and identify the faces in probe, which contains two steps: (1) For a given test face y, the solution of the following equation is calculated by resolving

$$\hat{x}=\arg\min\|x\|_1\ \ s.t.\ \sum_{l=1}^{L}\|y_l-D_lx\|_2^2\leq\varepsilon \tag{8}$$

and the corresponding reconstruction residuals is estimated by

$$\begin{aligned} r_i(y) &= \sum_{l=1}^{L}r_{il} \\ &= \sum_{l=1}^{L}\left\|y_l-D_l\delta_i\left(\hat{x}\right)\right\|_2^2,\quad i=1,\ldots,N. \end{aligned} \tag{9}$$

The construction of the gallery dictionary in A-SRC is same as W-SRC. (2) Once the sparse coefficient in (8) is determined, the minimal reconstruction residuals $r_i(y)$ and its corresponding to patch reconstruction residuals r_{il} in (10) are uniquely determined. If $r_{ik}\geq\eta$, the kth patch of the ith individual will be removed, where η is a given threshold. At this time, we resolve the following equation

$$\tilde{x}=\arg\min\|x\|_1\ \ s.t.\ \sum_{l=1,l\neq k}^{L}\|y_l-D_lx\|_2^2\ \leq\varepsilon \tag{10}$$

y is determined by the one with the smallest residual obtained by $r_i'(y)=\sum_{l=1,l\neq k}^{L}\|y_l-D_l\delta_i(\tilde{x})\|_2^2$.

3 Experiments

FRGC v2.0 database [14] used to conduct the experiments is one of the largest available public human face datasets and consists of 4007 scans of 466 individuals with neutral or non-neutral expressions. Three experiments are considered: (1) Landmark localization experiment. This experiment needs two training sets. ASM training set for coarse landmark detection consists of 1000 preprocessed range facial images with different expressions. The Gabor training set for fine landmark detection includes 50 preprocessed range images. Furthermore, the probe set consists of 1000 range images selected at random from the FRGC v2.0. (2) Threshold estimation experiment. **(I) 50 vs. neu. (II) 50 vs. non-neu. (III) 50 vs. Oth.** 50 individuals with different expressions are randomly selected as the training set, which are used to estimate the threshold η and weights W. The corresponding 50 neutral faces are used as the gallery in three outlined groups. Other faces divided into two categories: neutral and non-neutral faces are, respectively, used as the probe in (I) and (II). The other faces except the gallery are selected as the probe in group (III). (2) Recognition experiment. **(IV) Neu. vs. Oth.** and **(V) Neu. vs. OM.** The neutral scan of each individual image makes up a gallery of 466 different individuals. In (IV) Oth. represents the other 3541 scans are all the probes. In (V), OM. represents the manually selected 816 scans with open mouth expression [22] make up the probe faces.

3.1 Landmark localization experiment

ASM is used to locate landmarks coarsely and the best landmarks can be found by extracting Gabor wavelet feature in the coarse location area. The results of landmark localization about the same individual with three different expressions (neutral(a), laugh(b) and surprise(c)) and the different individual (d) with neutral expression are shown in Fig. 4. The red boxes and green points are the coarse location areas and fine localization landmarks, respectively. Table 1 shows the case of a small error $m_e \leq 0.06$ with average accuracy 95.7%, which indicates that the proposed algorithm can achieve the better performance in 3D landmark localization. The utilized 3D faces in this paper are preprocessed before landmarks localization, which is different from [16]. All preprocessed faces have a positive attitude, and each point is represented by a unique x, y, z coordinates. The precision of the single landmark is almost 2% higher than the results in [16] (94.2%).

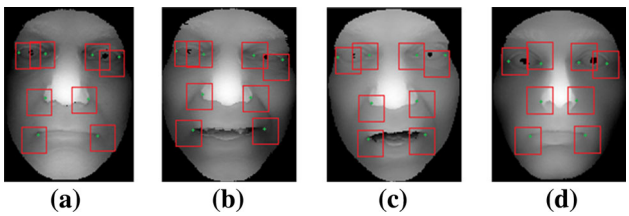

Fig. 4 The results of the landmark localization

3.2 Threshold estimation experiment

The threshold η has a tremendous effect on the performance of the proposed strategy. The inappropriate η may result that the extracted feature based on the selected facial region may contain the expression, which may reduce the recognition rate. So, the threshold estimation is a critical step to improve the performance of A-SRC. Some statistical distributions results of two experiments (I) 50 vs. neu. and (II) 50 vs. non-neu. with 3×3 patches are given in Fig. 5. The x-axis and y-axis in Fig. 5 are the corresponding reconstruction residuals and associated frequency, respectively. The residuals ranges of the expressional and neutral faces are different in Fig. 5. In other words, there is a significant difference in the residuals distribution of the neutral and expressional face. So, 3D face recognition testing based on (III) 50 vs. Oth. is considered to investigate the threshold about images with different patch segment. Based on the statistical results, the

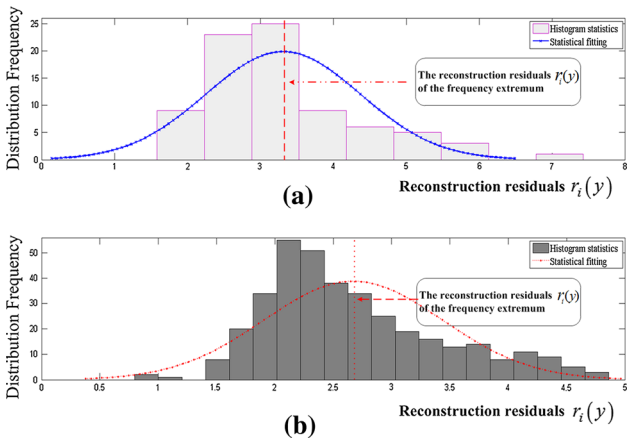

Fig. 5 Statistical fitting of the reconstruction residuals and frequency

Table 1 3D landmark localization results

	LEIC	LEOC	REIC	REOC	LMC	RMC	LNC	RNC	AVE
Mean	1.27	1.65	1.43	1.64	1.92	2.06	1.61	1.69	1.67
Std	1.54	1.50	2.67	2.61	1.95	2.40	2.39	1.35	2.05
m_e (%)	98.3	94.8	98.5	94.1	94.2	96.1	95.0	94.7	95.7

Table 2 Recognition rate using three different path sizes (%)

Method	A-SRC-η_1 (3.6)	A-SRC-η_2 (5.4)	A-SRC-η_3 (4.3)
$Q_{2,8}(1,3)$	91.63	94.57	96.15
$Q_{2,8}(3,3)$	96.61	98.19	98.64
$Q_{2,8}(4,3)$	95.25	97.51	97.29

A-SRC-η_1, A-SRC-η_2 and A-SRC-η_3 are A-SRC using three representative thresholds $\eta_1 = 3.6$, $\eta_2 = 5.4$ and $\eta_3 = 4.3$, respectively

residual distribution range shows the significant difference between the neutral and expressional faces about the patch segment 1×3 (Fig. 3a), 3×3 (Fig. 3b) and 4×3 (Fig. 3c) with the corresponding residual ranges (2.5, 4.5), (3.6, 4.5) and (3.1, 4.5), respectively. The testing results with three different patch segments about the LBP encoding operator $Q_{2,8}$ are listed in Table 2. $Q_{2,8}$ (3, 3) is selected as the feature extraction method due to the achieved highest recognition rate. In fact, the expressional parts still remain in the third parts due to the expression generated by the lower part of face in Fig. 3a. The irregular segments in Fig. 3a, b are not precisely estimated, and in other words, the expressional parts is still not effectively eliminated. The representative thresholds η are used in the recognition experiment although threshold is not unique based on the reconstruction residual distribution.

3.3 Recognition experiment

The performance of the SRC, W-SRC and A-SRC is compared based on the (IV) Neu. vs. Oth. and (V) Neu. vs. OM. in this section. The weights of the W-SRC are estimated based on the testing sample (III) 50 vs. Oth. 3×3 irregular segmentation is used for adaptive feature selection experiment based on the statistical results of Table 2. The experimental results based on (IV) Neu. vs. Oth. and (V) Neu. vs. OM are given in Tables 3 and 4, respectively. In Table 3, the recognition rate of the W-SRC is higher than SRC 4.06%(G-PLBP), 3.95%(L-PLBP) and 4.1%(F-PLBP) because it takes into account the weights related to each patch under expression variations. Although the average processing time is slightly more than the approach using SRC and W-SRC, the A-SRC achieves the highest recognition rate. This indicates the proper feature selection which improves the recognition rate by eliminating the negative influence from expressional parts and proves that the A-SRC is a robust and efficient method. In Table 4, the corresponding rank-1 recognition rate is higher than the SRC method about 7.11%(G-PLBP, A-SRC-η_1), 8.46%(G-PLBP, A-SRC-η_2), 9.2%(G-PLBP, A-SRC-η_3), 6.74%(L-PLBP, A-SRC-η_1), 8.21%(L-PLBP, A-SRC-η_2) and 7.48%(L-PLBP, A-SRC-η_3), 8.58%(F-PLBP, A-SRC-η_1), 10.54% (F-PLBP, A-SRC-η_2) and 10.66%(F-PLBP, A-SRC-η_3), respectively. The performance of the SRC is degraded for the face with large expression. Comparing to the experimental results based on Neu. vs. Oth, the recognition rate is reduced, respectively, about 1.93%(G-PLBP, A-SRC-η_1), 2.11%(G-PLBP, A-SRC-η_2), 2.02%(G-PLBP, A-SRC-η_3), 1.96%(L-PLBP, A-SRC-η_1), 2.05%(L-PLBP, A-SRC-η_2) and 2.13%(L-PLBP, A-SRC-η_3), 2.05%(F-PLBP, A-SRC-η_1), 2.09%(F-PLBP, A-SRC-η_2) and 2.77% (F-PLBP, A-SRC-η_3). Because the negative influence from the faces with large expression is not eliminated, the recognition rate of the W-SRC is still lower than the A-SRC. Based on the outlined discussion, the adaptive feature selection is an effective approach for the 3D face recognition by eliminating the influence from the expressional variations while select the robust feature for face matching.

Table 3 Experimental results based on Neu. versus Oth.

Methods	G-PLBP		L-PLBP		F-PLBP	
	RR (%)	ASSC	RR (%)	ASSC	RR (%)	ASSC
SRC	86.87	0.004620	86.76	0.004683	86.98	0.004736
W-SRC	90.93	0.005823	90.71	0.005792	91.08	0.005896
A-SRC-η_1	93.84	0.006521	93.50	0.006492	95.68	0.006531
A-SRC-η_2	95.37	0.006419	95.06	0.006501	97.68	0.006529
A-SRC-η_3	96.02	0.006503	94.41	0.006498	98.48	0.006542

Table 4 Experimental results based on Neu. versus OM

Methods	G-PLBP		L-PLBP		F-PLBP	
	RR (%)	ASSC	RR (%)	ASSC	RR (%)	ASSC
SRC	84.80	0.003672	84.80	0.003598	85.05	0.003795
W-SRC	88.85	0.004210	88.73	0.004412	88.97	0.004501
A-SRC-η_1	91.91	0.004315	91.54	0.004123	93.63	0.004516
A-SRC-η_2	93.26	0.004306	93.01	0.004092	95.59	0.004523
A-SRC-η_3	94.00	0.004295	92.28	0.004119	95.71	0.004520

RR recognition rate, *ASSC* The computational time in seconds about the average speed of sparse coding

3.4 Comparisons

The highest recognition rate has achieved 98.48% based on Neu. vs. Oth. which is higher than the results based on Neu. vs. OM about 2.77% in this paper. Similarily, the best recognition rate in Berretti et al. [22] and Lei et al. [23] is 95 and 95.2% based on Neu. vs. Oth. This is reduced by 4 and 2.4% comparing to the Neu. vs. OM, respectively. Our proposed approach shows the superiority results comparing to the outlined results based on Neu. vs. Oth. and Neu. vs. OM.. In particular, our results are, respectively, higher than Berretti et al.'s and Lei et al.'s work about 4.71 and 2.91% using the same Neu. vs. OM dataset. Our obtained results is also, respectively, higher than two other methods [24] and [25] about 8.85 and 9.19% under different dataset partition. This demonstrates that the proposed adaptive strategy based on the reconstruction residual and the accurate located landmarks can show comparative performance with *state-of-the-art* methods.

4 Conclusion

This paper proposes an adaptive feature selection based on reconstruction residual and accurate located landmarks for expression-robust 3D face recognition. Firstly, a novel facial *coarse-to-fine* landmark localization method based on ASM and GWT in range image is proposed to automatically and accurately locate facial landmarks. Secondly, the irregular landmarks-based patch facial representation is presented to keep the integrality of face local structure and get more detailed feature information. Thirdly, a novel F-PLBP feature is proposed to obtain the discriminative features and achieve accurate representations of facial surfaces. Fourthly, an adaptive feature selection method using the reconstruction residual and accurate located landmarks is proposed. Finally, the experimental evaluation based on FRGC v2.0 indicates that the proposed approach has the high discriminative power to significantly eliminate the influence from the facial expression variations and effectively improve the recognition accuracy.

Acknowledgements This research is supported by National Natural Science Foundation of China (Nos. 51475092, 61405034), and Fundamental Research Funds for the Central Universities (No. KYLX15_0117). The authors would like to thank A. Enis Cetin, Ph.D. Editor in Chief, associate editor and two anonymous reviewers who gave valuable comments and helpful suggestions for our manuscript.

References

1. Zhao, W., et al.: Face recognition: a literature survey. ACM Comput. Surv. **35**(4), 399–458 (2003)
2. Bowyer, K.W., et al.: A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition. Comput. Vis. Image Underst. **101**(1), 1–15 (2006)
3. Cootes, T.F., et al.: The Use of Active Shape Models for Locating Structures in Medical Images. Springer, Berlin Heidelberg (1993)
4. Elaiwat, S., et al.: 3-D face recognition using curvelet local features. IEEE Signal Process. Lett. **21**(2), 172–175 (2014)
5. Elaiwat, S., et al.: A Curvelet-based approach for textured 3D face recognition. Pattern Recognit. **48**(4), 1235–1246 (2015)
6. Jahanbin, S., Choi, H., Bovik, A.C.: Passive multimodal 2-D+ 3-D face recognition using Gabor features and landmark distances. IEEE Trans. Inf. Forensic Secur. **6**(4), 1287–1304 (2011)
7. Li, H., et al.: Towards 3d face recognition in the real: a registration-free approach using fine-grained matching of 3d keypoint descriptors. Int. J. Comput. Vis. **113**(2), 128–142 (2015)
8. Ouamane, A., Belahcene, M., Benakcha, A., et al.: Robust multimodal 2D and 3D face authentication using local feature fusion. Signal Image Video Process. **10**(1), 129–137 (2016)
9. Okuwobi, I.P., et al.: Three-dimensional (3D) facial recognition and prediction. Signal Image Video Process. **10**, 1–8 (2016)
10. Ming, Y.: Robust regional bounding spherical descriptor for 3D face recognition and emotion analysis. Image Vis. Comput. **35**, 14–22 (2015)
11. Wright, J., et al.: Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. **31**(2), 210–227 (2009)
12. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006)
13. Li, H., et al.: Expression-robust 3D face recognition via weighted sparse representation of multi-scale and multi-component local normal patterns. Neurocomputing **133**, 179–193 (2014)
14. Phillips, P.J., et al.: Overview of the face recognition grand challenge. In: Proc. Conf. Comput Vis Pattern Recognit. San Diego, CA, USA, **1**,947–954 (2005)
15. Li, X., Da, F.: Efficient 3D face recognition handling facial expression and hair occlusion. Image Vis. Comput. **30**(9), 668–679 (2012)
16. Liu, J., Da, F., Deng, X., et al.: An Automatic Landmark Localization Method for 2D and 3D Face. In: Proc. Int. Conf. Image and Graphics. Springer International Publishing, 541–551 (2015)
17. Lei, Y., et al.: An efficient 3D face recognition approach based on the fusion of novel local low-level features. Pattern Recognit. **46**(1), 24C37 (2013)
18. Liang, R., et al.: Bayesian multi-distribution-based discriminative feature extraction for 3D face recognition. Inf. Sci. **320**, 406–417 (2015)
19. Ojala, T., et al.: A comparative study of texture measures with classification based on featured distributions. Pattern Recognit. **29**(1), 51–59 (1996)
20. Qian, X., et al.: PLBP: an effective local binary patterns texture descriptor with pyramid representation. Pattern Recognit. **44**(10), 2502–2515 (2011)
21. Huang, D., et al.: 3D face recognition using elbp-based facial description and local feature hybrid matching. I IEEE Trans. Inf. Forensic Secur. **7**(5), 1551–1565 (2012)
22. Berretti, S., et al.: 3D face recognition using isogeodesic stripes. IEEE Trans. Pattern Anal. Mach. Intell. **32**(12), 2162–2177 (2010)
23. Lei, Y., Bennamoun, M. et al.: A structured template based 3D face recognition approach. In: Proc. 27th Int. Conf. Image Vis. Comput., New Zealand. 2012. Dunedin, New Zealand: ACM. (2012)
24. Smeets, D., et al.: MeshSIFT: local surface features for 3D face recognition under expression variations and partial data. Comput. Vis. Image Underst. **117**(2), 158–169 (2013)
25. Zhang, L., et al.: 3D face recognition based on multiple keypoint descriptors and sparse representation. PLoS ONE **9**(6), e100120 (2014)