Image and Vision Computing 35 (2015) 14–22

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Editor's Choice Article

# Robust regional bounding spherical descriptor for 3D face recognition and emotion analysis☆,☆☆

Yue Ming

*School of Electronic Engineering, Beijing Key Laboratory of Work Safety Intelligent Monitoring, Beijing University of Posts and Telecommunications, Beijing 100876, PR China*

ARTICLE INFO

*Article history:*
Received 23 August 2013
Received in revised form 29 April 2014
Accepted 25 December 2014
Available online 7 January 2015

*Keywords:*
3D face recognition
Emotion analysis
Regional bounding spherical descriptor
Regional and global regression
Kullback–Leiber divergence (KLD)

ABSTRACT

3D face recognition and emotion analysis play important roles in many fields of communication and edutainment. An effective facial descriptor, with higher discriminating capability for face recognition and higher descriptiveness for facial emotion analysis, is a challenging issue. However, in the practical applications, the descriptiveness and discrimination are independent and contradictory to each other. 3D facial data provide a promising way to balance these two aspects. In this paper, a robust regional bounding spherical descriptor (RBSR) is proposed to facilitate 3D face recognition and emotion analysis. In our framework, we first segment a group of regions on each 3D facial point cloud by shape index and spherical bands on the human face. Then the corresponding facial areas are projected to regional bounding spheres to obtain our regional descriptor. Finally, a regional and global regression mapping (RGRM) technique is employed to the weighted regional descriptor for boosting the classification accuracy. Three largest available databases, FRGC v2, CASIA and BU-3DFE, are contributed to the performance comparison and the experimental results show a consistently better performance for 3D face recognition and emotion analysis.



## 1. Introduction

Face recognition and emotion analysis are two important branches in biometric systems in remote communication, medical rescue, intelligent monitoring and so on. A large number of demands of face recognition and expression control are emerged due to the rapid development of 3D movies and entertainment [1–3]. More and more practice requirements are no longer satisfied by facial recognition or emotion analysis separately. The emerging industry needs to find an effective facial representation, which not only achieves a higher-quality discriminative power for the large size individuals but also provides a better expression description for control.

Our previous bounding sphere descriptor demonstrated superior performance in 3D face recognition, especially where there were large pose variations. For practical applications, the proposed feature descriptor should be able to simultaneously perform face recognition and expression analysis. Expression variations imply variations in specific facial regions. Therefore, in this paper, we extend the global bounding sphere descriptor to regional bounding sphere descriptors. Regional weighted selection is used to form a general feature descriptor, which can be used for face recognition and expression analysis simultaneously. The novel low-dimensional features of this approach facilitate both theoretical innovations and practical applications.

Sufficient broad investigations of 3D face processing have been achieved in the literature [4,5]. Although most of them independently analyze the issue of 3D face recognition and 3D facial expression classification, the literature illustrate that facial regional descriptor analysis can provide better accuracy for the two aspects simultaneously. For 3D face recognition, Faltemier et al. [6] divided a face into a group of regions and led to a better recognition performance. A region-based registration was employed to establish correspondence and 3D shape descriptor was used for statistical feature extraction [7]. Passalis et al. [8] used facial symmetry to overcome the challenges of large pose variations and the wavelet-based biometrics signature was used to evaluate the real-world applications. Local shape difference boosting [9] selected optimal local features for assembling three collective strong classifiers and found the most discriminative feature for 3D face recognition. Ioannis Marras [10] introduced novel subspace-based methods for learning the azimuth angle of subspace normal, which were well-suited for all types of 3D facial data for recognition. However, note that these algorithms focus on 3D face recognition. They cannot perform facial expression recognition.

Some scholars are concerned with the descriptions of the different facial expressions, without the discrimination of the different individuals. For example, the Facial Action Coding System (FACS) [11], as a human facial expression representation, has been found over 25 years ago. Prominent regions can describe the variant facial characteristics on the different individuals [12]. However, practical application

☆ Editor's Choice Articles are invited and handled by a select rotating 12 member Editorial Board committee. This paper has been recommended for acceptance by Dr. Stefanos Zafeiriou.

☆☆ The work presented in this paper was supported by the National Natural Science Foundation of China (Grant No. NSFC-61402046), President Funding of Beijing University of Posts and Telecommunications (Grant No. 2013XZ10).

*E-mail address:* myname35875235@126.com.

http://dx.doi.org/10.1016/j.imavis.2014.12.003
0262-8856/

requirements, i.e. high recognition rate, low computational complexity, and easy implementation, are the issues of wide concern.

With the rapid development of science and technology, more and more systems need to distinguish the individual and his/her expression simultaneously, and then provide personalized services, as ours developed the system [2]. One barrier is how to precisely segment the facial regions on the different facial images, which is highly influenced on the accuracy of 3D face recognition and expression classification. In this paper, we first develop a facial segmentation mechanism, and then principally concentrate on how to find an effective facial descriptor with the capability of individuals' identification and expression description. The lower dimensional feature vectors can be used to compensate for computationally expensive.

### 1.1. Outline of our 3D face processing framework

Empirical study shows that facial shape has a significant variance in terms of different regions on the facial surface. In order to better reflect anatomical structure and describe the expression variations, we introduce the segmenting scheme to address the major challenges and present a new recognition framework for 3D processing. Our research consists of three important procedures: facial region segmentation, feature representation and feature extraction as shown in Fig. 1.

1. Facial region segmentation: By combining the facial shape characteristics and shape index information, a committee of facial local regions can be coarsely located. Shape band [13] algorithm is introduced to detect the contours of facial regions and refine the localization to highlight the discriminative and descriptive regions.
2. Feature representation: Exploring our previous research [14], we further develop the robust regional bounding sphere descriptor (RBSR) on the facial surface, which are relatively consistent in the presence of expressions and poses. Combined with facial prior assumption, the region descriptors are weighted and projected into the spherical domain based on the contribution of the descriptiveness and discrimination.
3. Feature extraction: Facial poses, occlusions and corruptions, as the major challenging issues for facial recognition and expression classification, have varying degrees at different facial regions. Regional and global regression mapping (RGRM) is introduced to grasp the intrinsic manifold structure of our feature representation for weighted RBSR descriptors, which better reflect the geometrical properties of the different facial regions with low redundancy.

### 1.2. Contributions of our 3D face recognition and expression classification framework

In our framework, a novel mechanism is proposed to overcome the unsolved challenging issues encountered for facial processing, called Regional Bounding Sphere Representation (RBSR). The main contributions of our scheme can be summarized in the following items:

1. Robustness: Facial segmentation scheme based on shape index and region selection can robustly extract facial shape regions. The regional bounding sphere representation can effectively reduce the impact of non-discriminant areas and provide a promising way to describe facial regional properties. RGRM algorithm can extract the low-dimensional information from RBSR descriptor, which contains the most discriminative face and descriptive expressions.
2. Efficiency: 3D data provides reliable data to support synergetic analysis for face recognition and expression classification. Large-capacity data analysis demands for high-performance and multi-task concurrency. The proposed framework can find a fast and general descriptor to learn different visual tasks simultaneously and compresses redundant data. Efficiency of our proposed framework is more suited to practical applications.
3. Synergy: The current mainstream of visual study focuses on the single visual task. Research is generally conducted on specific tasks, for a specific application, using a specific method to solve certain vision related problems. We propose a generic descriptor for facial information. By means of effective feature learning, the results of face recognition and expression classification can be output simultaneously. The proposed framework lays the foundation for synergetic analysis for multi-visual tasks and face real-time intelligent interactive systems.

The rest of this paper is organized as follows. Facial regional segmentation is described in Section 2. Section 3 presents the regional bounding sphere representation. In Section 4, regional and global regression mapping is used with RBSR for low dimensional feature extraction. The detailed experimental comparison with 3D face recognition and expression classification is analyzed in Section 5. Finally, we conclude the paper in Section 6.

## 2. 3D facial regional segmentation

### 2.1. Data preprocessing

The original 3D facial images in the databases usually contain some non-facial areas, spikes and holes as shown in Fig. 2(a) [15]. Exploiting our previous research [14], facial profile can be extracted by the robust ASM algorithm [12] on the 2D texture image. The main facial area can be coarsely extracted by the "and" logic processing between the 2D profile and its corresponding 3D valid matrix in Fig. 2(b).

In order to further refine the facial area, the position of the nose tip needs to be detected. The frontal face model is chosen as the reference model $F_r$. Then, the mirror face $F_r'$ is obtained by the coordinate plane $\Sigma_{xz}$. The TPS-RPM algorithm [16] is used to register $F_r'$ to $F_r$ and generates the registered image $F_r''$. The facial symmetry plane $\Sigma_r$ is fitted using the point set:

$$p_{ir} = (p_i + p_i'')/2, 1 \le i \le N, \tag{1}$$

where $p_i = (x_i, y_i, z_i)^T$ is the point cloud of 3D reference face model and $p_i'' = (x_i'', y_i'', z_i'')^T$ is the point cloud of 3D registered face model. The cross curve of $F_r$ and $\Sigma_r$ with 1 mm bandwidth is treated as the objective surface $OS_r$ and the highest point on $OS_r$ is determined as nose tip $N_r$ in Fig. 2(d). When there are nose occlusions, according to the proportion of prior knowledge of facial organs, we select the cross point on the fitted surface $OS_r$ between two-thirds on the vertical direction and one-half on the horizontal direction, as the nose tip as illustrated in

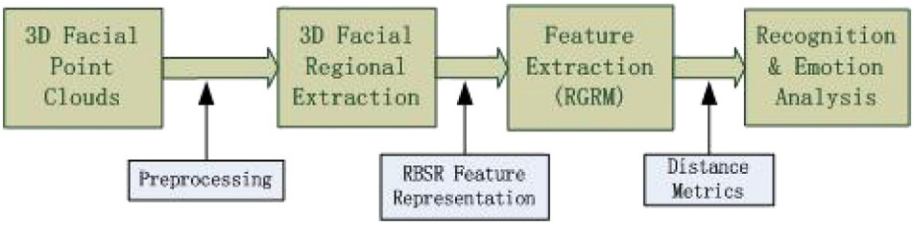


**Fig. 1.** The flow chart of our proposed 3D face recognition and expression classification framework.

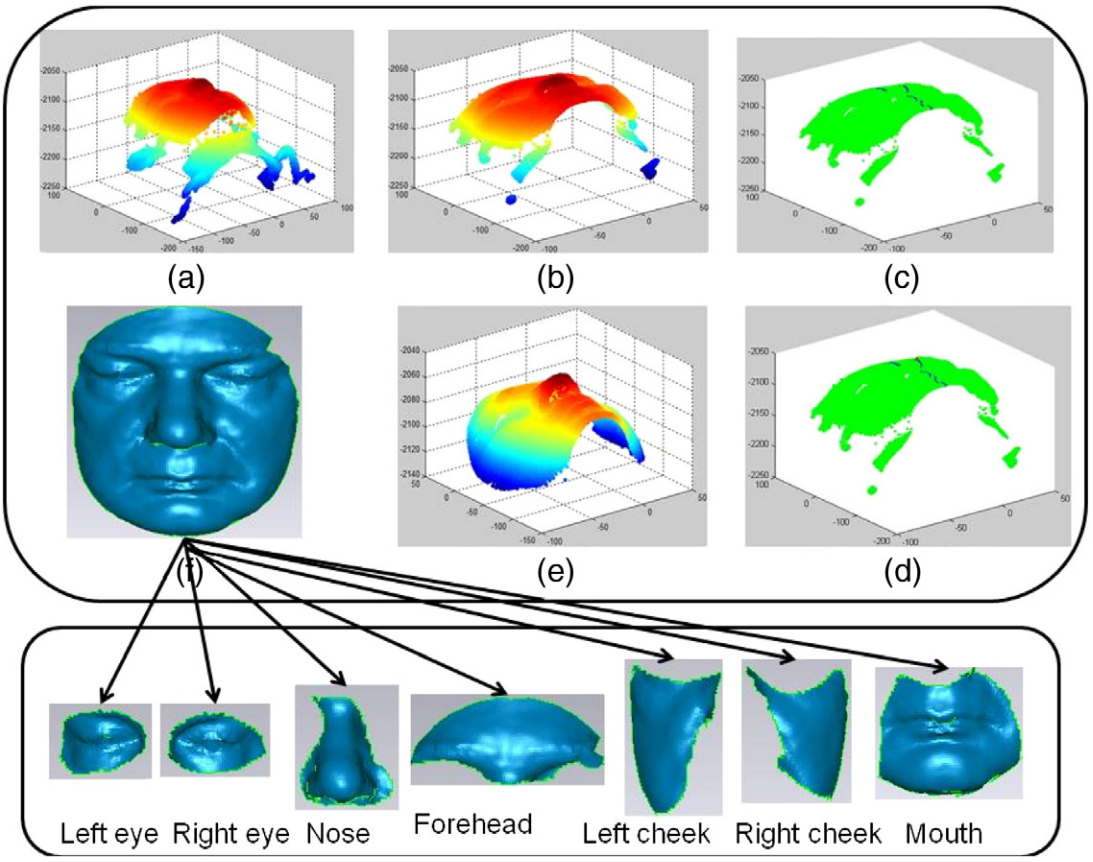


**Fig. 2.** The procedure of facial segmentation. (a) The original 3D point cloud. (b) The facial area extraction. (c) The symmetry curve extraction from the extracted facial area. (d) Nose tip detection. (e) The facial surface alignment. (f) The aligned facial image.

Fig. 3. Nose tip refined detection under occlusion is the next stage of our research focus.

The following procedure is used to normalize the size of the various 3D input images. Firstly, we estimate the horizontal and vertical projection curves for the extracted facial area by calculating row and column sums of the valid point matrix. Next, we select the maximum value of the horizontal and vertical projection curves as the location of the maximum horizontal and vertical distance respectively on the facial area.

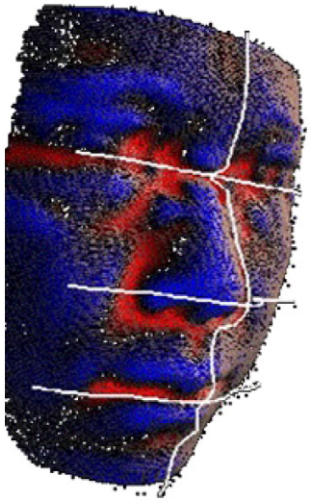

**Fig. 3.** The proportion of prior knowledge of facial organs for nose tip detection.

The ratios of the longest horizontal and longest vertical facial locations in the input and reference images are then used to normalize the width and length of the 3D input images as follows:

$$ratioX = distX/distXone, ratioY = distY/distYone, \tag{2}$$

where *distX*, *distY* are respectively the values of horizontal and vertical lengths on the input 3D face image. *distXone*, *distYone* are the values of horizontal and vertical lengths on the reference face model. Finally, the normalization is calculated using the ratio calculated above.

Then, the distance of two eyes' centers can be computed by locating the eyes' centers on 2D image by ASM [17] and project them to the corresponding 3D images. The ratio of eye center's distances between the input and reference models is introduced to refine the normalization results in Fig. 2(e). Empirical study shows that the results with facial normalization can significantly boost the accuracy for face recognition and expression classification.

Axis-angle representation can align the input with the reference model in Fig. 2(f) as follows:

$$p_{ireg} = p_i \cos\theta + (r \times p_i)\sin\theta + r(r \cdot p_i)(1-\cos\theta), \tag{3}$$

where $p_i$ is the input points of the 3D images and $p_{ireg}$ is the registered face points. The rotation axis $r$ can be obtained via the cross product of the unit normal $n_r$ and $n$ of the reference and input images. The angle $\theta$ between the $n_r$ and $n$ describes the rotation magnitude. Unlike ICP, we just match the point clouds between input and reference models. As a result, our method can save a huge number of computational costs.

### 2.2. Facial region segmentation

A 3D facial image contains more facial discriminative and expression descriptive information than a 2D image and the degree is different from the different facial regions and curvature. The derived shape index serves as an important descriptor of intrinsic surface properties and is used for segmenting the different facial regions.

We can segment that the points like cone shape as the upper nose. To find the inner eye pit locations, shape band [13] of eyes area extracted from the reference model is used as the search window on an input face and Gaussian curvature values, which is close to zero, is used to determine the local minimum inside the search window as the positions of eye pits. Then, we introduce the shape index values for searching nose borders as saddle rut-like shapes [18]. The nose border outline extracted by the shape band of the reference model is developed for searching the rightmost and leftmost points along the band, which can be treated as the left and right nose borders [7]. The shape band search window, based on the sign and values of curvature, is used to coarsely segment the mouth area. Considering facial prior information, the region, above the eyes, can be treated as forehead and the areas, between left/right eyes and mouth, can be determined as left/right cheek as shown in the bottom part of Fig. 2.

## 3. Feature representation

### 3.1. Regional bounding sphere representation

In order to balance the facial discriminative power and expression descriptiveness, extended with our previous research [14], a novel descriptor is proposed, denoted as regional bounding sphere representation (RBSR). For each region of a 3D facial image, the descriptor is formed from the projection of the facial point cloud into regional bounding spheres which are centered at the center points of the regions as shown in Fig. 4. The distance between points in the region and the corresponding center point divided by the radius of regional bounding sphere produces the values of the spherical points. The value of points on the regional bounding sphere can be defined as follows:

$$RBSR\left(c_j\right) = \sqrt{\left(p_{ix}-c_{xj}\right)^2 + \left(p_{iy}-c_{yj}\right)^2 + \left(p_{iz}-c_{zj}\right)^2}/R_{bsj}, \tag{4}$$

where $p_i$ is the coordinate value of each aligned 3D facial region point, $R_{bsj}$ denotes the maximum difference among the three different directions $X$, $Y$, $Z$ on the Cartesian coordinates and $c_j$ is the centroid point of the $j$-$th$ region. Then, the RBSR descriptor of each region $j$ is denoted as $RBSR^j = \{RBSR^j_1, \cdots, RBSR^j_N\}$, $1 \le i \le N$, where $N$ is the number of points on the each facial region. The RBSR has successfully reduced computation complexity based on the discussion in our previous work.

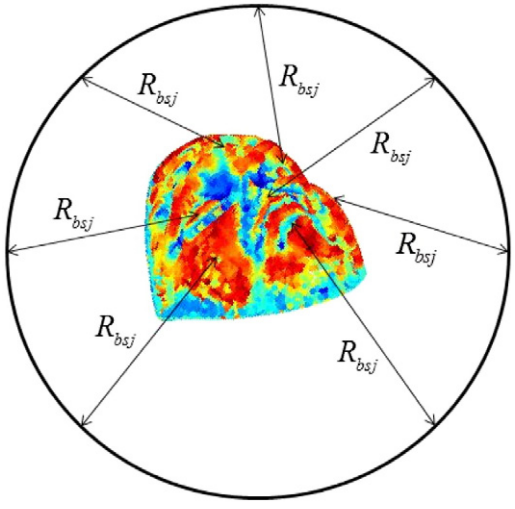


**Fig. 4.** Regional bounding sphere representation of facial surface, where $R_{bsj}$ is the radius of the regional bounding sphere.

### 3.2. Robust region selection based on RBSR

In the following, we estimate a weight for each regional representation. Mutual information is introduced to describe the relativity and redundancy of the different facial regions [19]. We define the average value of all mutual information between the regional descriptor $rbsr_i$ and the region $r$ as the relevance of the whole descriptor $RBSR$:

$$D(RBSR, r) = \frac{1}{|RBSR|}\sum_{rbsr_i \in RBSR} I(rbsr_i; r). \tag{5}$$

The redundancy of the descriptor $RBSR$ is summed by the average value of all mutual information between the regional descriptor $rbsr_i$ and the regional descriptor $rbsr_j$:

$$R(RBSR) = \frac{1}{|RBSR|^2}\sum_{rbsr_i, rbsr_j \in RBSR} I\left(rbsr_i; rbsr_j\right). \tag{6}$$

Suppose region pairs are conditionally independent, we can approximately calculate the probability of a descriptor set $RBSR$ given a region $r$ and select the $r$ that maximizes this probability. We can treat the maximum probability as the regional weight for specified application in the following equation. To build the connection between a specific individual and the corresponding facial regions, the object label information is provided for each facial region:

$$P(RBSR|r) = \prod_{rbsr_i \in RBSR} P(l_i|r) \prod_{rbsr_j \in RBSR} P\left(l_j|l_i, r\right)P\left(D, R|r, l_i, l_j\right) \tag{7}$$

where $l_i, l_j \in L$ is the discrete quantized labels associated with descriptor $rbsr_i \in RBSR$, and $D$, $R$ are the values of relevance and redundancy computed by Eqs. (5) and (6), respectively. The detailed numerical solution method can be seen in Appendix A and the regional weights of RBSR descriptor can be calculated as follows:

$$\omega_r = \log(P(RBSR|r)) = \sum_{l \in L} \log(P(b_l|r)) \tag{8}$$

where $\log(P(b_l|r)) = \sum_{rbsr_i \in RBSR} \sum_{rbsr_j \in RBSR} \log\left(P\left(D, R|r, l, l_j\right)\right)$ is the bin probability and $\omega_r$ is the weight of $RBSR$ descriptor for the region $r$.

Thus, the descriptor of the whole 3D facial images can be expressed for the specified applications:

$$x_k = \omega_1(rbsr_1) + \cdots + \omega_i(rbsr_i) + \cdots + \omega_7(rbsr_7) \tag{9}$$

where $k$ is used for the specified 3D facial processing, which can be referred to face recognition and verification, expression analysis, and pose estimation. $i$ ranges from 1 to 7 for the different facial regions, including nose, left/right eyes, left/right cheeks, forehead and mouth. The processing can further improve 3D facial region segmentation and selection, which not only details local properties but also enhances the shape information of 3D facial images.

## 4. Robust regional and global regression mapping for feature extraction

In this section, we have two learning tasks for face recognition and expression classification when the same input feature descriptor is used. Here, we assume $X \subset R^D$ is the feature descriptor computed by Section 3. Our proposal is to learn two regression functions $a_1$ and $a_2$

**Table 1**
Average Euclidean distances (mm) between manual and automatic found landmarks.

| Facial keypoints | Inner eye | Nose tip | Nasal basis points |
|---|---|---|---|
| FRGC v2 | 4.90 | 3.26 | 4.51 |

and obtain their corresponding lower dimensional feature vectors $z_1$ and $z_2$. Robust regional and global regression mapping algorithm (RGRM) is proposed to dimension reduction and overcome some remaining artifacts left from the region segmentation, such as some stretched or misaligned images, hair occlusions, large data noises and corruptions.

In order to better cope with the unseen data and reflect the facial global properties, we introduce a nonparametric approach [20] for out-of-sample extrapolation. Since the regional structure of manifold in 3D face data is linear, the regression expression can be defined as a linear regression model, i.e., $Y = W^T X + E$, where $W$ is the local projection matrix, and $E$ is the noise matrix. We map the weighted facial RBSR descriptor into a Hilbert space $H$ and $R^d$, i.e., $y_i = \phi(W)^T\phi(x_i) + e_i$, where $\phi(W)$ is the regional regression matrix from $H$ to $R^d$ and $e_i \in R^d$ is noise term.

The objective function can be rewritten to simultaneously learn the low dimensional embedding $Y$ and the mapping matrix as

$$\begin{aligned}&\min_{\omega_i,Y}\sum_{i=1}^{n}\left(\left\|X_i^T W_i + 1_j e_i^T - Y_i\right\|_F^2 + \gamma\|\omega_i\|_F^2\right) + \mu\left(\left\|\Phi(X)^T\Phi(W) + 1_n E^T - Y\right\|_F^2\right)\\ &\quad + \gamma\|\Phi(W)\|_F^2,\\ &s.t. Y^T Y = I\end{aligned} \tag{10}$$

where $1_j \in R^j$ and $1_n \in R^n$ are two vectors of all ones. According to the literature [20], $y$ can be computed by

$$y = Y^T(HKH + \gamma I)^{-1}HK_x + \frac{1}{n}Y^T 1_n - \frac{1}{n}Y^T(HKH + \gamma I)^{-1}HK1_n \tag{11}$$

where $H = I - \frac{1}{n}1_n 1_n^T$ denotes as the global centering matrix, $K_x \in R^n$ denotes as a vector with its $i$-$th$ element $K_{xi} = \phi(x)^T\phi(x_i) = \exp(-\|x - x_i\|^2/\sigma^2)$, and $x_i \in \chi$ is the $i$-$th$ *RBSR* feature descriptor in the training set. In order to avoid overfitting, we perform local PCA to reduce the dimension of each facial RBSR descriptor as preprocessing.

In multi-task learning, given an extended training set of the RBSR descriptors $\{(x_i^k, y_i^k)\}_{i=1}^{m_k}$ computed by Eq. (11), $x_i^k \in R^n$ denotes the *RBSR* descriptor of the $i$-$th$ training sample for the $k$-$th$ application, $y_i^k$ denotes the corresponding output, $m_k$ is the number of facial points. Let $X^k = \left[x_1^k, \cdots, x_{m_k}^k\right] \in R^{m_k \times n}$ denote the data matrix for the $k$-$th$ application, and $Y^k = \left[y_1^k, \cdots, y_{m_k}^k\right] \in R^{m_k}$ denote the lower dimensional embedding as the corresponding label of the whole image. The regression vector for all regions from the regression matrix $A = [a_1, \cdots, a_k] \in R^{n \times k}$ needs to be estimated by the discussion below . Then, we can obtain the maximum posterior estimation for face recognition and expression classification by the following equation:

$$\left(A^k\right)^* = \arg\min_{A^k} P\left(y|A^k, \mu_1\right)P(y|\mu_2)\prod_{j=1}^{N} P\left(a_j^k|\delta_j\right)\prod_{k=1}^{2}\prod_{i=1}^{m_i} P\left(y_i^k|A^k, x_i^k, \sigma^k\right) \tag{12}$$

where $\mu_1$ and $\mu_2$ are the parameters for specified application. We assume the corresponding target $y^k \in R$ for the $k$-$th$ application has a Gaussian distribution with mean $y^k \in R$ and precision $\sigma^k > 0$:

$$p\left(y_i^k|A^k, x_i^k, \sigma^k\right) = \sqrt{\frac{\sigma^k}{2\pi}}\exp\left(\frac{-\sigma^k\left(y^k - \left(A^k\right)^T x^k\right)}{2}\right). \tag{13}$$

Assume the prior probability $p(A^k|\delta^k)$ is generated according to the exponential prior: $p(A^k|\delta^k) \propto \exp(-\|A^k\|\delta^k)$. Here, pairwise data $x$, $y$ is drawn independently from the distribution (Eq. (12)) and $A^1, \cdots, A^n$ are drawn independently from the prior in Eq. (13). Then, the likelihood function and prior can be expressed as follows and the classification result is obtained based on the maximum probability:

$$p(y|A, X, \delta) = \prod_{k=1}^{2}\prod_{i=1}^{m^k} p\left(y_i^k|A^k, x_i^k, \sigma^k\right), p(A|\delta) = \prod_{i=1}^{n} p\left(A^k|\delta^k\right). \tag{14}$$

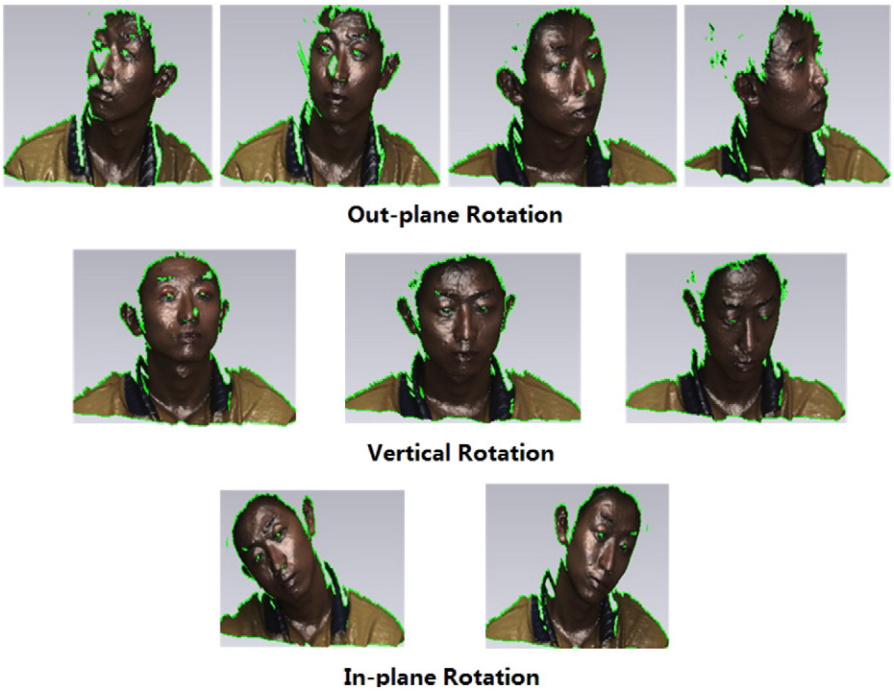


**Fig. 5.** The examples of large pose variations in 3D face database.

**Table 2**
The accuracy of 3D face alignment.

| Methods | TBV | IPRSPV | OPRSPV | IPRLPV | OPRLPV |
|---|---|---|---|---|---|
| Ours | 92.12% | 96.74% | 97.56% | 85.37% | 87.80% |
| Traditional method | 93.49% | 92.68% | 96.34% | 74.79% | 75.38% |

Then, we can obtain the optimal estimation of the regression matrix $A^*$ as follows:

$$A^* = \arg\min_A p(y|A,\mu_1)p(y|\mu_2)p(y|A,X,\delta)p(A|\delta). \quad (15)$$

Finally, the high-dimensional input RBSR descriptor $X$ can be embedded into an intrinsic low-dimension feature vector by

$$Z = AX. \quad (16)$$

Our method can effectively balance the discriminant power of face recognition and the descriptiveness of expression analysis and the extracted features can better reflect facial surface structure, which is immune to the distorted artifacts and noise corruptions.

## 5. Experiments

In this section, we provide a performance comparison of our proposed framework for 3D face recognition and emotion analysis with state-of-the-art solutions. We validate our method using the FRGC v2 database [21], BU-3DFE database [22], and CASIA 3D face database [23]. FRGC is an international general database for 3D face recognition and verification, which is comprised of 4007 images from 466 different individuals. BU-3DFE is a 3D facial expression database, which consists of 2500 3D facial expression models from 100 different individuals. CASIA contains large pose variations for testing the effectiveness of our algorithm for pose variations.

Firstly, in Section 5.1, we show the efforts of 3D face preprocessing using our method described in Section 2. Then, we perform a performance evaluation of our proposed feature description for 3D face recognition and verification in Section 5.2. In Section 5.3, we report the results of 3D facial expression analysis on BU-3DFE database. Finally, we provide a comparative study of our proposed framework with large facial pose variations in Section 5.4.

### 5.1. 3D facial preprocessing

The original 3D facial images in the database usually contain some non-facial areas, such as ears, necks and shoulders as shown in Fig. 2(a). There are also some spikes and holes captured by the 3D acquisition devices. Removing the non-facial areas and facial segmentation is critical for recognition accuracy. We first evaluate the performance of 3D facial preprocessing, including 3D facial regional segmentation and 3D facial alignment.

#### 5.1.1. 3D facial regional segmentation

We first use the FRGC database for performance evaluation. We segment facial regions by manual and use our proposed method to segment automatically. We calculate Euclidean distance between the manual facial regions and automatic facial regions. If the distance exceeds a threshold, we treat there is a segmentation error. The threshold setting is associated with subsequent recognition results.

Table 1 shows the average Euclidean distances (mm) between manually and automatically found landmarks. Our proposed method satisfies the error constraints that can be used for the following feature extraction and classification.

**Table 3**
Test configurations with different test sets.

| Test databases | $i$ | Number of subjects | Training set | Test set |
|---|---|---|---|---|
| 1 | 1 | 410 | 410 | 3541 |
| 2 | 2 | 384 | 768 | 3183 |
| 3 | 3 | 316 | 948 | 3003 |
| 4 | 4 | 285 | 1140 | 2755 |

#### 5.1.2. 3D face alignment

In these experiments, we use FRGC and CASIA 3D face databases to evaluate the performance of 3D face alignment, including top/bottom view (TBV), in-plane rotation with small pose variations (IPRSPV), in-plane rotation with large pose variations (IPRLPV), in-plane rotation with small pose variations (IPRSPV), out-plane rotation with large pose variations (OPRLPV) and out-plane rotation with small pose variations (OPRSPV) as shown in Fig. 5. Table 2 shows 3D face alignment results of our proposed method and traditional method. Our method can be aligned by gestures matched from different object models with the same attitude, which is more robust than the traditional method under the different pose variations and the computational efficiency.

### 5.2. Performance evaluation of 3D face identification and verification

#### 5.2.1. 3D face identification based on the different representations

Our proposed RBSR descriptor is used to evaluate the discriminating power with several alternatives for 3D facial representations. We consider five different representations for identification purposes: surface normal (SN), surface curvature (SC), depth images (DI), shape index descriptor (SID), and our proposed RBSR descriptor. The experimental set is the same as Section 6.1 of our previous work [14]. Table 3 summarizes the configurations of the test and training sets.

Table 4 shows the recognition accuracy based on the different descriptors. Our RBSR descriptor provides a consistently higher level of accuracy on all of the test configurations, especially for the test databases 3 and 4. The shape index achieves the second higher accuracy over the other representations. The surface curvature has a slightly higher performance compared with the surface normal and depth images. Surface normal is computed by differential geometry surface, and they can actually encode the rate of change of the surface [24]. But it is sensitive to the external conditions, especially for large pose and illumination variations. Surface curvature can treat the human face surface as a free-form object and indicate the local descriptor to represent facial shape information. However, with large noises and corruptions, the

**Table 4**
Recognition results with different descriptors.

| Test | RBSR | DI | SC | SN | SID |
|---|---|---|---|---|---|
| 1 | 56.48% | 45.38% | 47.47% | 42.36% | 49.59% |
| 2 | 62.68% | 47.76% | 53.13% | 46.97% | 59.75% |
| 3 | 70.1% | 53.08% | 56.51% | 51.32% | 63.07% |
| 4 | 74.23% | 57.71% | 61.6% | 57.75% | 70.13% |

**Table 5**
Verification results based on the standard protocol with different 3D face recognition algorithms.

| Methods | ROC I | ROC II | ROC III |
|---|---|---|---|
| Faltemier [6] | – | – | 94.9% |
| Husken [26] | – | – | 86.9% |
| Cook [27] | 93.71% | 92.91% | 92.01% |
| Kakadiaris [28] | 97.3% | 97.2% | 97% |
| Ocegueda [12] | 96.2% | 95.7% | 95.2% |
| Queirolo [25] | – | – | 96.6% |
| Alyuz [7] | 85.39% | 85.63% | 85.64% |
| Mian [29] | – | – | 86.6% |
| Ours | 95.67% | 95.28% | 95.03% |

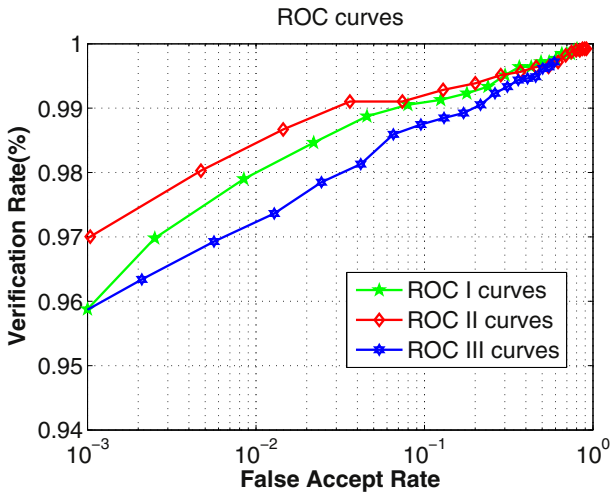


**Fig. 6.** ROC curves on FRGC 3D face database in the different test sets.

performance of surface curvature are significantly degraded by the non-rigid transformation. The depth images can actually reflect the distance between the individuals and the sensor for effectively discriminant feature extraction. The shape index descriptor combined with shape and distance information makes it have more discriminative power than the other three representations. Our RBSR descriptor can effectively overcome the challenging facial variations and illustrate a distinct improvement compared with the other representations.

#### 5.2.2. 3D face verification based on the standard protocol

In the subsection, some existing algorithms are used to evaluate the performance of 3D face verification with the standard protocols, ROC I, ROC II and ROC III of FRGC v2 [21] for the different acquisition times. Table 5 demonstrates the verification results with the FAR of 0.1% and Fig. 6 illustrates the ROC curves. All evaluation results come from the records of the corresponding authors in their paper.

Table 5 illustrates that our proposed scheme is better than the others for the standard protocol, except by Kakadiaris et al. [28] and Queirolo et al. [25]. Kakadiaris et al. [28] proposed a complex multistage method for 3D face verification, including multistage alignment, AFM fitting, feature signature and fusing. Just fitting the AFM had already spent more than 15 s from the input data. For Queirolo's method [25], they pointed out that Simulated Annealing needs approximately 12.19 s computation cost for the global convergence only with slightly better results. As shown in Table 6, our method only spends 5.958 s for the whole data processing, which effectively reduces the running time.

In addition, our proposed methods, compared to other methods with approximate performance, also take into account the accuracy for expression analysis, not only for face verification. Based on the above discussion, our method effectively balances the simple implementation and high accuracy with computational efficiency.

**Table 6**
Identification time for 3D facial data processing in each step.

| | |
|---|---|
| Average point number | 100,474 |
| Facial region extraction | 2.3 s |
| Nose tip detection | 0.04 s |
| Facial region refinement | 1.46 s |
| Facial region segmentation | 1.05 s |
| Posture alignment | 0.26 s |
| RBSR | 0.694 s |
| RGSRM | 0.154 s |

**Table 7**
Details for each of the comparative features.

| Name | Details of facial expression features |
|---|---|
| Feature 1 | Choose first 68 feature points defined in each facial model |
| Feature 2 | Choose 15 feature points defined in face contour |
| Feature 3 | Choose all the 83 feature points defined in each facial model |
| Feature 4 | Choose the distance vector $D_i$ (formed by 6 distance features) |
| Feature 5 | Choose the distance vector $D_i'$ (formed by 24 distance features) |
| Feature 6 | Choose the slope vector $S_i$ (formed by 10 slope features) |
| Feature 7 | Choose the angle vector $A_i$ (formed by 12 angle features) |
| Feature 8 | Choose both the distance vector $D_i$ and the angle vector $A_i$ together |
| Feature 9 | Choose both the distance vector $D_i$ and the slope vector $S_i$ together |
| Feature 10 | Our proposed features |

### 5.3. 3D facial expression analysis

#### 5.3.1. DP comparisons of different 3D facial expression features

We conduct experiments on BU-3DFE to justify the descriptiveness of 3D facial expression analysis. Based on our previous research [30], we first introduce Kullback–Leibler divergence (KLD) [31] to evaluate superiority of our proposed feature compared with the other emotional features. From the view of Bayesian theory, the expression descriptive power relies on the class-conditional distribution [32]. And thus the facial expression descriptiveness of the feature set $y$ can be calculated by KLD:

$$DP(y) = \sum_{i=1}^{c} \sum_{j=i+1}^{c} Ds\left(f_i(y) \| f_j(y)\right) \tag{17}$$

where $c$ is the number of the facial expression classes (six prototypic expressions), $f_i(y)$ and $f_j(y)$ are the *i-th* and *j-th* class-conditional probability of feature vector $y$ respectively.

Referring to the feature sets designed in recent publications [30], a series of experiments detailed in Table 7 are implemented to validate expression classification. All of them are repeated for 100 times to lower the randomness and randomly selected 50 face models from each expression class each time. To better reflect the expression descriptive power, flow-matrix and geometry-matrix are introduced to 3D facial emotion analysis respectively and the results are illustrated in Fig. 7.

Among all the features, our proposed feature provides the highest facial expression descriptiveness consistently. For all features, the descriptive power in flow-matrix form is much higher than that in

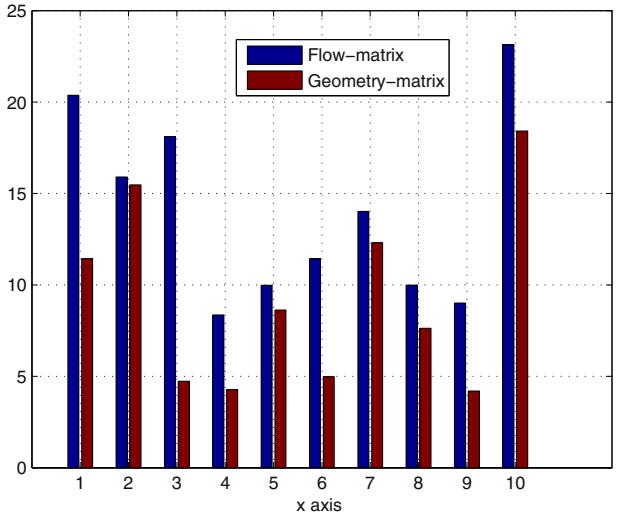


**Fig. 7.** Comparison of descriptive power for Feature 1–Feature 10.

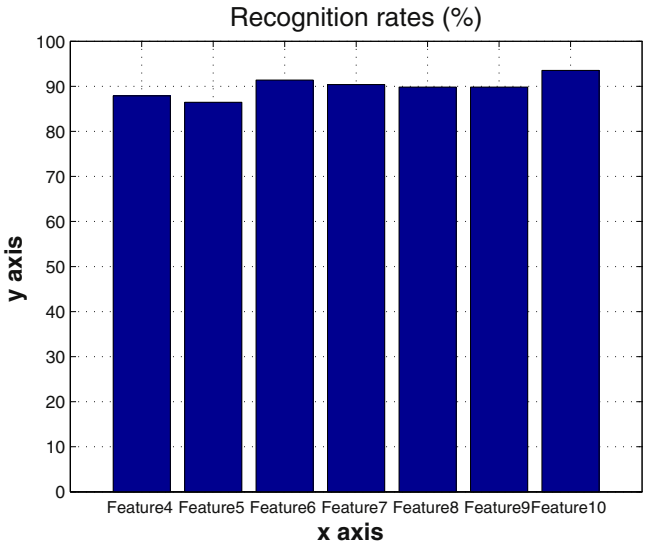


**Fig. 8.** Comparison of the recognition rates for different features.

geometry-matrix form. In both forms, Feature 4 performs the worst descriptiveness. Although Features 1 and 3 can obtain the higher descriptive power in flow-matrix form, the performances degrade sharply in geometry-matrix form. Features 2 and 7 can be treated as the median results of all features. As illustrated in Fig. 7, our features utilize more distinguishing descriptiveness than do the other algorithms and make it possible to encode more emotion descriptive information of 3D facial expression.

*5.3.2. 3D facial expression recognition*

To verify the recognition results of expression classification, we provide the results based on BU-3DFE database. 50 subjects are randomly selected as a training set, and the rest as a testing set; each experiment is repeated 100 times to the average as shown in Fig. 8.

Our proposed feature achieves the best recognition rate among all the features. The performance of Feature 6 and Feature 7 are slightly higher than the others. For all basic expressions, the recognition rates of our features are all higher than that of the second higher Feature 6. Especially in fear (FE) and disgust (DI) expressions, the discrepancy is

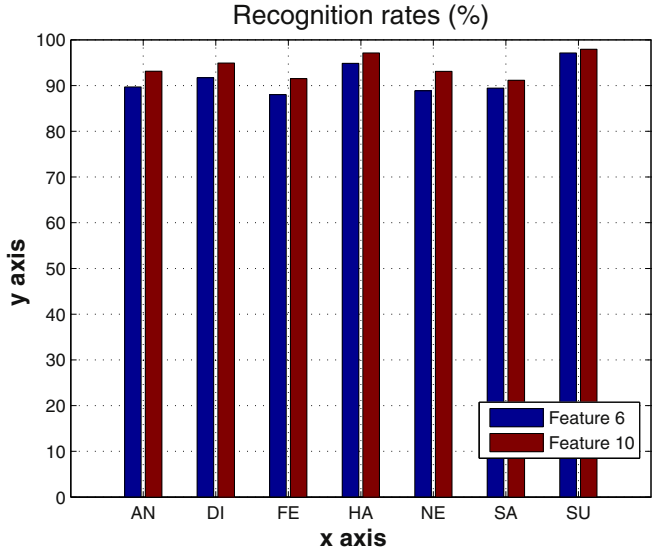


**Fig. 9.** Comparison of the recognition rates for different expressions in Feature 6 and Feature 10.

**Table 8**
Rank-1 recognition accuracy in the CASIA 3D face database (100 subjects).

| Test databases | SPV | LPV | SPVS | LPVS |
|---|---|---|---|---|
| Depth | 90.7% | 50.0% | 81.0% | 46.5% |
| Intensity | 69.9% | 49.5% | 68.1% | 48.5% |
| Depth gabor | 91.4% | 51.5% | 82.4% | 49.0% |
| Intensity gabor | 75.3% | 65.5% | 77.6% | 61.5% |
| Decision fusion | 89.0% | 70.5% | 85.6% | 64.5% |
| Feature fusion | 91.0% | 91.0% | 87.9% | 79.0% |
| Ours | 93.42% | 92.5% | 90.28% | 82% |

3.53% and 3.2% respectively. From Figs. 7 and 8, feature discrimination power and its corresponding expression descriptiveness are consistent. Our feature and slope features (Feature 6) possess much more emotion analysis power and higher recognition rates than the angle and distance features (Feature 7 and Feature 4). From Fig. 9, among all prototypic expressions, our feature and slope features (Feature 6) have more emotion descriptiveness than the distance features (Feature 5), especially for HA (Happy) and SU (Surprise) expressions. Thus, our features achieve the best performance in the expression classification.

*5.4. Robustness to large facial pose variations*

Large pose variations have a huge impact on 3D face recognition and emotion analysis. In this subsection, we evaluate the performance of the different facial descriptions with large pose variations. We follow our previous experimental design based on CASIA database [23], including small pose variations (SPV), large pose variations (LPV), small pose variations with smiling (SPVS) and large pose variations with smiling (LPVS).

From Table 8, we can conclude that our method obtains the best results among all the testing sets. Xu et al. [33] considered that the fusion schemes with different characteristics to describe a subject. The "feature" level has a superior performance than the "decision" level fusion. However, pose variations can result in self-occlusions due to missing data, which significantly reduce the performance of the fusion scheme. Our method can project the 3D facial data into the regional bounding spheres to overcome the patched occlusions caused by large pose variations and improve the recognition accuracy.

## 6. Conclusions

In this paper, we have presented a novel framework for 3D face recognition and emotion analysis at the same time. We propose a multistage preprocessing method to segment a group of facial regions. Then, the RBSR descriptor is proposed to describe the different facial regions with different discrimination and descriptiveness. The RGRM model developed by local and global regression scheme is used to extract facial intrinsic manifold and realize the ability of facial discrimination and emotional descriptiveness. We present experimental results on the three largest standard 3D face databases, FRGC v2, BU-3DFE and CASIA. The performance of our proposed framework is with the properties of effectiveness, robustness and generality for 3D face recognition and emotion analysis.

## Appendix A. Appendix

In practice, log probability is employed to avoid extremely small values with numerical precision [34] and the above equation can be expressed as follows:

$$\log(P(RBSR|r)) = \sum_{rbsr_i \in RBSR} \log(P(l_i|r)) + \sum_{rbsr_j \in RBSR} \log\left(P\left(l_j|l_i,r\right)\right) + \log\left(P\left(D,R|r,l_i,l_j\right)\right). \quad (A.1)$$

To decrease the computational complexity, we assume $P(l_i|r)$ and $P(l_j|l_i, r)$ as the uniform probabilities. Nonuniform label probability can

be computed by concatenating the individual label histogram to the pairwise regional descriptor vector in the probability expression:

$$\log(P(RBSR|r)) = \sum_{rbsr_i \in RBSR} \sum_{rbsr_j \in RBSR} \log\left(P\left(D, R|r, l_i, l_j\right)\right) + C. \quad (A.2)$$

Since the uniform probabilities for labels approximately estimated by a constant $C$ are not depended on the region $r$, we can omit for clarify. The above equation can be rewritten as to bin probability based on the individual descriptor label:

$$\omega_r = \log(P(RBSR|r)) = \sum_{l \in L} \log(P(b_l|r)) \quad (A.3)$$

where $\log(P(b_l|r)) = \sum_{rbsr_i \in RBSR} \sum_{rbsr_j \in RBSR} \log\left(P\left(D, R|r, l, l_j\right)\right)$ is the bin probability and $\omega_r$ is the weight of *RBSR* descriptor for the region $r$.

## References


[1] S.H. Lee, S. Sharma, Real-time disparity estimation algorithm for stereo camera systems, IEEE Trans. Consum. Electron. (3) (2011) 1018–1026.
[2] Y. Ming, Q. Ruan, A mandarin edutainment system integrated virtual learning environments, Speech Comm. (1) (2013) 71–83.
[3] Y. Gao, Q. Dai, N. Zhang, 3D model comparison using spatial structure circular descriptor, Pattern Recogn. (3) (2010) 1142–1151.
[4] K.W. Bowyer, K. Chang, P. Flynn, A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition, Comput. Vis. Image Underst. (1) (2006) 1–15.
[5] Y. Gao, M. Wang, Z. Zha, Q. Tian, Q. Dai, N. Zhang, Less is more: efficient 3D object retrieval with query view selection, IEEE Trans. Multimedia (5) (2011) 1007–1018.
[6] T. Faltemier, K. Bowyer, P. Flynn, A region ensemble for 3D face recognition, IEEE Trans. Inf. Forensics Secur. (1) (2008) 62–73.
[7] N. Alyuz, B. Gokberk, L. Akarun, Regional registration for expression resistant 3D face recognition, IEEE Trans. Inf. Forensics Secur. (3) (2010) 425–440.
[8] G. Passalis, P. Perakis, T. Theoharis, I.A. Kakadiaris, Using facial symmetry to handle pose variations in real-world 3D face recognition, IEEE Trans. Pattern Anal. Mach. Intell. (10) (2011) 1938–1951.
[9] J.L.Y. Wang, X. Tang, Robust 3D face recognition by local shape difference boosting, IEEE Trans. Pattern Anal. Mach. Intell. (10) (2010) 1858–1870.
[10] G.T.I. Marras, S. Zafeiriou, Robust learning from normals for 3D face recognition, Computer Vision-European Conference on Computer Vision, 2012, pp. 230–239.
[11] J. Wang, L. Yin, Static topographic for facial expression recognition and analysis, Comput. Vis. Image Underst. (1) (2007) 19–34.
[12] O. Ocegueda, S.K. Shah, I.A. Kakadiaris, Which parts of the face give out your identity? Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2011, pp. 641–648.
[13] X. Bai, Q. Li, L.J. Latecki, W. Liu, Z. Tu, Shape band: a deformable object detection approach, Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2009, pp. 1335–1342.
[14] Y. Ming, Q. Ruan, Robust sparse bounding sphere for 3D face recognition, Image Vis. Comput. (8) (2012) 524–534.
[15] Y. Gao, J. Tang, R. Hong, S. Yan, Q. Hai, N. Zhang, Camera constraint-free view-based 3D object retrieval, IEEE Transactions on Image Processing (21) (2012) 2269–2281.
[16] A.R.H. Chui, A new algorithm for non-rigid point matching, Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2000, pp. 44–51.
[17] K. Seshadri, M. Savvides, Robust modified active shape model for automatic facial landmark annotation of frontal faces, Proc. of IEEE 3rd International Conference on Biometrics: Theory, Applications and Systems (2009) 1–8.
[18] C. Dorai, A. Jain, Cosmos—a representation scheme for 3D free-form objects, IEEE Trans. Pattern Anal. Mach. Intell. (10) (1997) 1115–1130.
[19] H. Peng, F. Long, C. Ding, Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. (8) (2005) 1226–1238.
[20] Y. Yang, H.T. Shen, Z. Ma, Z. Huang, X. Zhou, L21-norm regularized discriminative feature selection for unsupervised learning, Proc. of International Joint Conferences on Artificial Intelligence (2011) 1589–1594.
[21] P.J. Phillips, P.J. Flynn, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2005) 947–954.
[22] L. Yin, X. Wei, Y. Sun, J. Wang, M. Rosato, A 3D facial expression database for facial behavior research, Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06), 2006, pp. 211–221.
[23] Casia-3D facev1, http://biometrics.idealtest.org/.
[24] B. Gokberk, H. Dutagaci, A. Ulas, L. Akarun, B. Sankur, Representation plurality and fusion for 3D face recognition, IEEE Trans. Syst. Man Cybern. B (1) (2008) 155–173.
[25] C. Queirolo, L. Silva, O. Bellon, M. Segundo, 3D face recognition using simulated annealing and the surface interpenetration measure, IEEE Trans. Pattern Anal. Mach. Intell. (2) (2010) 206–221.
[26] M. Husken, M. Brauckmann, S. Gehlen, C. Von der Malsburg, Strategies and benefits of fusion of 2D and 3D face recognition, Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, p. 174.
[27] J. Cook, C. McCool, V. Chandran, S. Sridharan, Combined 2D/3D face recognition using log-gabor templates, Proc. IEEE Intl Conf. Video and Signal Based Surveillance, 2006, p. 83.
[28] I.A. Kakadiaris, G. Passalis, G. Toderici, N. Murtuza, Y. Lu, N. Karampatziakis, T. Theoharis, 3D face recognition in the presence of facial expressions: an annotated deformable model approach, IEEE Trans. Pattern Anal. Mach. Intell. (4) (2007) 640–664.
[29] A.S. Mian, M. Bennamoun, R. Owens, An efficient multimodal 2D–3D hybrid approach to automatic face recognition, IEEE Trans. Pattern Anal. Mach. Intell. (11) (2007) 1927–1943.
[30] X. Li, Q. Ruan, Y. Ming, A remarkable standard for estimating the performance of 3D facial expression features, Nerocomputing (1) (2012) 99–108.
[31] S. Kullback, R.A. Leibler, On information and sufficiency, Ann. Math. Stat. (1) (1951) 79–86.
[32] B. Levy, R. Nikoukhah, Robust least-squares estimation with a relative entropy constraint, IEEE Trans. Inf. Theory (1) (2004) 89–104.
[33] C. Xu, S. Li, T. Tan, L. Quan, Automatic 3D face recognition from depth and intensity gabor features, 2009.
[34] P. Matikainen, M. Hebert, R. Sukthankar, Representing pairwise spatial and temporal relations for action recognition, Proc. of European Conference on Computer Vision (2010) 1–14.