Pattern Recognition 47 (2014) 509–524

Contents lists available at ScienceDirect

Pattern Recognition

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

An efficient 3D face recognition approach using local geometrical signatures

Yinjie Lei [a,b,*], Mohammed Bennamoun [b], Munawar Hayat [b], Yulan Guo [c,b]

[a] College of Electronics and Information Engineering, Sichuan University, Chengdu, Sichuan 610064, China
[b] School of Computer Science and Software Engineering, University of Western Australia, Crawley, WA 6009, Australia
[c] College of Electronic Science and Engineering, National University of Defense Technology, Changsha, Hunan 410073, China

ARTICLE INFO

Article history:
Received 26 March 2013
Received in revised form
19 June 2013
Accepted 26 July 2013
Available online 7 August 2013

Keywords:
3D biometrics
3D face recognition
3D representation
KPCA
SVM

ABSTRACT

This paper presents a computationally efficient 3D face recognition system based on a novel facial signature called Angular Radial Signature (ARS) which is extracted from the semi-rigid region of the face. Kernel Principal Component Analysis (KPCA) is then used to extract the mid-level features from the extracted ARSs to improve the discriminative power. The mid-level features are then concatenated into a single feature vector and fed into a Support Vector Machine (SVM) to perform face recognition. The proposed approach addresses the expression variation problem by using facial scans with various expressions of different individuals for training. We conducted a number of experiments on the Face Recognition Grand Challenge (FRGC v2.0) and the 3D track of Shape Retrieval Contest (SHREC 2008) datasets, and a superior recognition performance has been achieved. Our experimental results show that the proposed system achieves very high Verification Rates (VRs) of 97.8% and 88.5% at a 0.1% False Acceptance Rate (FAR) for the "neutral vs. nonneutral" experiments on the FRGC v2.0 and the SHREC 2008 datasets respectively, and 96.7% for the ROC III experiment of the FRGC v2.0 dataset. Our experiments also demonstrate the computational efficiency of the proposed approach.

1. Introduction

Face recognition has drawn considerable attention in the last few decades due to its non-intrusiveness and wide number of applications, e.g. surveillance and access control. Although considerable achievements have been attained with 2D face recognition, its accuracy is still challenged by pose and illumination variations [1,2]. With the rapid development of 3D imaging sensors, many researchers have now turned to 3D face recognition due to its potential capabilities to overcome the inherent limitations and drawbacks of 2D face recognition. Moreover, the geometrical information provided by 3D data has the potential to achieve a greater recognition accuracy compared to 2D data [3,4].

Based upon the type of strategy used to measure the similarity between two facial surfaces, the existing 3D face recognition approaches can be categorized into surface registration-based and feature matching-based approaches. The surface registration-based approaches commonly use the Iterative Closest Point (ICP) algorithm [5–7] or one of its variants [8,9]. The ICP error is used as a similarity measure between two matching facial surfaces for recognition. The ICP itself does not require any feature extraction or projection into another space. It is an iterative process and uses the whole surface, which makes it computationally expensive. The second type of approaches (i.e. feature matching-based) match two 3D facial surfaces defined in different coordinate systems based on object-centric shape features [10–13]. Some of them have achieved very promising performance in terms of recognition accuracy. However, despite achieving very high recognition accuracy, many of the existing approaches use complex shape features [14–16,13]. They require complicated mathematical transformations and are therefore not suitable for real-world applications due to the high computational costs associated with them. In this work, we develop an efficient Angular Radial Signature (ARS) to represent a 3D face. It encodes the complex 3D facial surface with a set of 1D feature vectors, which attain sufficient discriminating power while simultaneously achieving significant computational efficiency.

Similar to 2D face recognition, expression variation is also considered to be the main challenge of 3D face recognition since it introduces significant changes in the geometry of the facial surface. Two approaches are usually adopted in the literature of 3D face recognition to handle facial expressions. First, an expression deformable model is learned from facial scans under both neutral and nonneutral expressions. The learned model is then used to morph out the expression deformations from a nonneutral face

* Corresponding author. Postal Address: No. 24 South Section 1, 1st Ring Road, Chengdu, Sichuan 610064, P. R. China. Tel.: +86 028 85505099.
E-mail addresses: leolyj@gmail.com (Y. Lei), bennamou@csse.uwa.edu.au (M. Bennamoun), munawar@csse.uwa.edu.au (M. Hayat), yulan.guo@nudt.edu.cn (Y. Guo).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.07.018

[17–19]. The second approach is simply based on the fact that only few facial regions are affected by the distortions caused by expressions and most of the other regions remain invariant. It was proven that by only taking into account the least affected regions, a more accurate and robust face recognition can be achieved compared to the case of using the whole face [20,9,12]. Our expression invariant face recognition system combines the strengths of the above two approaches. First, when using the facial scans with various expressions of each individual for training, the learned model can accommodate expression variations across different individuals. Second, the proposed ARS features are only extracted from the semi-rigid facial region, which can remain relatively robust under the distortions caused by expressions. Third, unlike many existing 3D face recognition approaches which rely on surface registration or complex feature matching between a probe face and all gallery faces, our proposed ARS based approach is computationally efficient.

The rest of this paper is organized as follows. Section 2 describes the related work in the area of 3D face recognition and provides an overview of our proposed approach. Section 3 describes the 3D face normalization algorithm. Section 4 provides a detailed explanation of our proposed ARS facial representation. Section 5 presents the proposed two-stage mapping-based training and testing strategy for 3D face recognition. An experimental performance evaluation and a comparison with the state-of-the-arts are provided in Section 6. Finally, Section 7 comments on and concludes this work.

2. Related work and overview

2.1. Related work

The work in [1,4] provides a comprehensive survey of the 3D face recognition approaches. In the following, we will restrain our review to the approaches which are closely related to this work. More specifically, we cover the basic 3D face recognition approaches which were tested on the Face Recognition Grand Challenge (FRGC v2.0) dataset (Section 2.1.1), approaches developed to handle expression variations (Section 2.1.2) and the relevant 3D facial analysis approaches which were based on machine learning techniques (Section 2.1.3).

2.1.1. Three dimensional face recognition

3D face recognition approaches can be coarsely classified into two categories: surface registration-based and feature matching-based approaches. Chang et al. [6] segmented the 3D face into multiple regions. The regions around the nose of the gallery and probe faces were matched (using ICP) and their matching scores were combined to determine the final recognition results. In [21], Lu et al. constructed a database of 3D mesh models from several 2.5D images and proposed a recognition approach based on the ICP algorithm. To build their 3D meshes, they detected feature points in the 2.5D images, and extracted the maximum and minimum local curvatures. Next, the ICP was applied around these points to align all the 2.5D overlapping regions. Then, the local feature-based ICP registration error was used as a metric to perform matching between faces. In [7], Faltermier et al. presented a 3D face recognition approach based on the detection and registration of multiple small facial regions (some of them overlapping), which were extracted based on specified distances with respect to the location of the nosetip. These regions were then matched independently using the ICP algorithm. They analyzed their face recognition performance under different combinations of these facial regions. They concluded that one combination, which consists of 28 regions (around the nose) yielded the highest recognition accuracy. This indicates that the nose and the region around it are the least affected by expression deformations.

Al-Olsaimi et al. [10] proposed a 3D face recognition approach which integrated global and local geometric cues. They represented the 3D face with multiple rank-0 tensor fields. Then, the local and global fields were integrated into a 2D histogram. The Principal Component Analysis (PCA) coefficients were extracted from these histograms and combined into a single feature vector. The Euclidean distance between a pair of feature vectors was used as a similarity score to match two faces. Berretti et al. [15] proposed a 3D face recognition system that partitioned a 3D face into a set of isogeodesic stripes. Then, a descriptor named 3D Weighted Walkthroughs (3DWWs) was used to represent these stripes, and a graph-based matching algorithm was used to match a pair of faces. Gupta et al. [22] proposed a 3D face recognition approach that could automatically detect 10 anthropometric fiducial points on a 3D face. They used a stochastic pairwise method to calculate the Euclidean and geodesic distances between these points and combined them into a single feature vector. Then, a linear discriminant classifier was used for face recognition.

Most of the existing 3D face recognition approaches rely on a surface registration or on complex feature (surface descriptor) extraction and matching techniques. They are, therefore, computationally expensive and not suitable for practical applications. In this work, we effectively and simply represent a 3D face by a set of 1D feature vectors. Compared with the existing approaches, it can simultaneously achieve high recognition accuracy and computational efficiency.

2.1.2. Expression invariant approaches

The existing 3D face recognition approaches can be categorized, with respect to handling facial expressions, into two groups: (1) those which learn a model for the faces under different expression deformations and (2) those which only extract and process the rigid facial parts. In [17], Kakadiaris et al. proposed a 3D face recognition approach based on an annotated deformable model. They mapped the 3D geometric information onto a 2D regular grid to generate a so called geometric image. Two types of wavelet transforms were then applied on the geometric image for the extraction of features. In [18], Lu et al. developed a hierarchical geodesic-based resampling method to extract landmarks for the modeling of the deformations of a face under expression. Then, a deformable model was constructed for each individual in the gallery. The model is able to handle expressions and pose changes simultaneously. Their experimental results showed that their proposed model is robust under different facial expressions. Al-Olsaimi et al. [19] presented a so called expression model which was learned from pairs of neutral and nonneutral faces of each individual in the gallery. Next, a PCA subspace was constructed using the shape residues of the finely registered pairs of faces. The learned model was then used to morph out any expression deformation from a nonneutral face.

Amongst the approaches which belong to the second category, Zhong et al. [20] divided the 3D face into upper and lower (including the mouth) regions, but only used the upper region. Gabor filters were applied to extract Gabor features, and the centers of the filter response vectors were then learned by K-means clustering. Finally, the recognition results were obtained using a nearest neighbor classifier that was based on a learned visual codebook representation. Mian et al. [9] proposed an approach which automatically extracted two semi-rigid regions from a 3D face. Then, the extracted regions were matched individually using a modified ICP algorithm. The final matching score was obtained by combining the minimum Mean Squared Error (MSE) from each region. Lei et al. [12] developed a binary

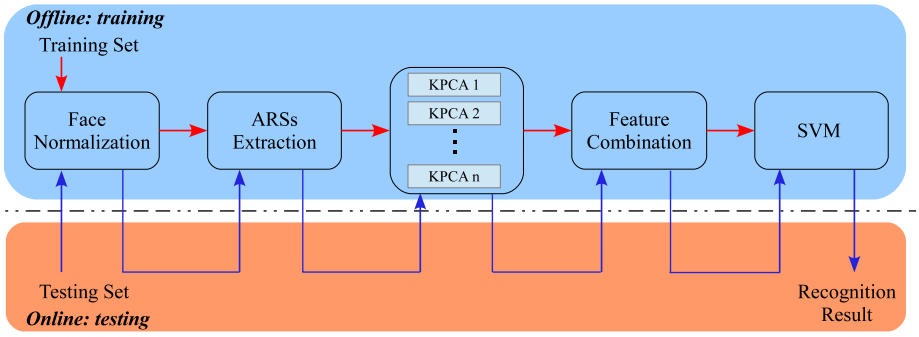

Fig. 1. Block diagram of the proposed 3D face recognition approach.

mask to extract the rigid and semi-rigid facial regions. Four kinds of local low-level geometrical features were extracted from each region. These features were concatenated into one single feature vector and fed into a Support Vector Machine (SVM) to perform face recognition. Their experimental results showed that the 3D face recognition accuracy is highly influenced by the non-rigid facial regions.

These two categories of approaches overlook the fact that facial expressions are specific to each individual, i.e. the facial deformation patterns vary from one individual to another. Furthermore, except for the parts which get distorted the most (e.g. the mouth), the rigid parts of the face are also different and specific to each individual. In this work, we first exclude the large distortion part from the face. We then train a model using faces with different expressions for each individual. This results in a learned model which can accommodate expressions specific to each individual.

2.1.3. Machine learning based approaches

In this section we review some of the machine learning based approaches for 3D facial analysis (face and facial expression recognition). Wang et al. [11] extracted Haar, Gabor and Local Binary Patterns (LBP) features from a Signed Shape Difference Map (SSDM) computed between two aligned faces. They then applied a boosting algorithm to select the most discriminating features and weak classifiers were assembled into three collective strong classifiers. Their approach achieved both high accuracy and computational efficiency when applied to 3D face recognition. Sun et al. [23] investigated a dynamic 3D facial expression recognition approach which computed vertex correspondences across frames. They proposed a spatio-temporal Hidden Markov Model (HMM) to learn the spatial and temporal information of a face. They showed that their dynamic model outperforms static models when tested on a 3D dynamic facial expression dataset. Xu et al. [24] extracted Gabor features with multiple scales and orientations from both facial range and intensity images. They then proposed a hierarchical feature selection algorithm based on Linear Discriminant Analysis (LDA) and AdaBoost learning to select the most discriminating and robust features. Finally, the selected features were used to construct a strong classifier. Ballihi et al. [16] represented a 3D face by a set of circular and radial curves defined by a Euclidean distance function. Next, a large set of geometric features were extracted from those curves. Then, the AdaBoost algorithm was used to select the optimal features and a composite classifier was devised. A high recognition performance was achieved with only a minimal set of features.

In the existing approaches, machine learning techniques have been usually used to select and extract the most optimal features to improve the recognition performance. Unlike those approaches, in this paper, we employ machine learning techniques to effectively address the individual-specific expression deformation problem. Specifically, machine learning techniques are used to train a model using faces of each individual with different expressions.

2.2. Overview of the proposed approach

2.2.1. Approach overview

We propose an efficient 3D face recognition approach based on a novel ARS feature and a two-stage mapping-based classification strategy. Fig. 1 shows the block diagram of the proposed approach. Each block is briefly described below:

- An effective 3D face normalization algorithm is employed to perform nosetip detection, face cropping and pose correction for 3D faces. This face normalization algorithm is automatic and accurate (Section 3).
- A 3D face is divided into the non-rigid and semi-rigid regions. It is reported that the mouth is the most affected area while the nose and eyes-forehead areas are the least affected by expression deformations [12]. Therefore, we only take into account the upper part of the face (with reference to the nosetip) and discard the lower part (Section 4).
- A set of distinctive features is needed to represent a 3D facial surface in order to achieve accurate 3D face recognition. In this work, we propose a novel ARS, which can effectively decompose a 3D face into a set of 1D feature vectors (Section 4).
- Due to the similarity in appearance of the faces of different individuals, the extracted ARS features revealed to be very close in the feature space. This may lead to a linearly inseparable classification problem. To solve this problem, the Kernel Principal Component Analysis (KPCA) is used to nonlinearly map the ARS features into a high-dimensional mid-level feature space (Section 5.1).
- The resulting mid-level features are combined and fed into an SVM for training. The learned SVM model is then used to classify any input face during the testing phase (Section 5.2).

2.2.2. Contributions of this work

The main contributions of the proposed 3D face recognition approach can be summarized as follows:

- A novel ARS based facial representation is proposed that can reduce the large amount of 3D face data into a set of discriminating 1D feature vectors. The proposed representation is distinctive and requires very low computational effort.
- Multiple KPCAs are employed to nonlinearly map the ARS features into a high-dimensional space and convert them to mid-level features. It was shown that the linearly inseparable classification problem in the input feature space can be resolved through this mapping.

- A machine learning framework is proposed to train a model that is able to accommodate different facial expressions (to address the expression variation problem). The model is trained using faces under various expressions of different individuals.
- A set of experiments are conducted on two large publicly available datasets, i.e. FRGC v2.0 and 3D track of Shape Retrieval Contest (SHREC 2008), demonstrating a high recognition accuracy and computational efficiency of the proposed approach.

3. 3D face normalization

We use two large publicly available datasets for 3D face recognition research, i.e. the FRGC v2.0 [25] and the 3D face track of SHREC 2008 [26]. The facial scans of the two datasets were captured by a 3D sensor (typically a Minolta Vivid) in the form of dense pointclouds. Both datasets are comprised of facial scans acquired from the shoulder level up. Most of these scans suffer from missing data, noise spikes as well as minor pose variations. Therefore, these scans need to be preprocessed. We employ an automatic 3D face normalization algorithm which includes face detection and preprocessing, nosetip detection, face cropping and pose correction. A detailed description of our 3D face normalization algorithm is given below.

First, noise spikes are simply reduced by eliminating the outlier points (vertices). An outlier point is defined based on the statistical information of its 8-connected neighbors [9]. A bicubic interpolation is applied along all the three coordinates (x, y and z) to fill in the missing data, which often appear in the eyes and mouth areas. Second, since the facial scans were acquired at different distances from the 3D sensor, they have different depth resolutions. We therefore resample the 3D facial images on a uniform square grid of 1 mm, to ensure that all 3D faces have the same resolution.

We then introduce a coarse-to-fine algorithm based on our previous work [27] to detect the nosetip from the uniformly resampled range images. Compared to the single detection in [9], the coarse-to-fine strategy can significantly improve the accuracy of nosetip detection (the effect of the nosetip detection errors on the system performance is discussed in Section 6.3.1). Our nosetip detection algorithm proposed in [27] has three steps. First, 61 profile projection on the xy-plane are generated for an input facial pointcloud by rotating the face around the y-axis by a set of angles. These angles range from -90° to $+90^\circ$ with an increment of 3°. Second, a set of possible nosetip candidates are determined by moving a circle along all the points on the face profile. The criteria for a point on the face profile to be selected as a nosetip candidate are "the difference in area of the circle lying inside and outside of the face profile must be greater than an experimentally determined threshold" [27]. Finally, the nose shape of the face profiles around all candidates is fitted with a triangle. Based upon a measure of fitness of the triangle, the final nosetip is selected amongst all the candidate nosetips. Note that, our nosetip detection algorithm does not require any training and is robust to pose variations.

Next, we define a sphere with a radius of 80 mm, centered at the nosetip, to crop the 3D face (i.e. the points that are located more than 80 mm from the nosetip are discarded). We then use our pose correction algorithm in [9] to correct the pose of the 3D facial scan to a frontal view. The pose is corrected using the principal directions of the cropped face calculated using the PCA algorithm. A bicubic interpolation is used again to fill in the missing data which appear after pose correction due to self-occlusion. Finally, the facial pointclouds are converted to range images by interpolating at the integer x and y coordinates (1 mm on both directions) used as the horizontal and vertical indices, respectively. These indices are used to determine the corresponding z coordinate which corresponds to the range image pixel value. A Gaussian filter is used to further smooth the resulting facial range image.

Note that, the 3D facial scans from the SHREC 2008 dataset are provided with their detected nosetips. Therefore, in this case, the nosetip detection is not necessary prior to normalization. Fig. 2 shows some examples of the face normalization results.

4. Angular radial signature (ARS) based facial representation

A 3D facial scan is represented by a set of points in $\mathbb{R}^3$. The number of these points is usually quite large, suggesting that this raw 3D facial representation will suffer from the "curse of dimensionality" [28]. Therefore, 3D features (descriptors) are commonly extracted to reduce the dimensionality. The accuracy of the face recognition system depends mainly on the descriptiveness and distinctiveness of these features.

A simple yet effective 3D face representation is required to achieve accurate 3D face recognition. Two types of facial representations are usually used in the literature. The first and simplest one is the pointcloud 3D facial representation. A pointcloud is actually a set of 3D coordinates of all the facial surface (points/vertices) from which different kinds of geometrical features can be extracted [29,12]. The second category of facial representation methods uses facial range images computed from the projection of the pointclouds on the xy-plane along the z-axis. Consequently, standard 2D feature extraction techniques can be applied to these range image based methods [24,11,30].

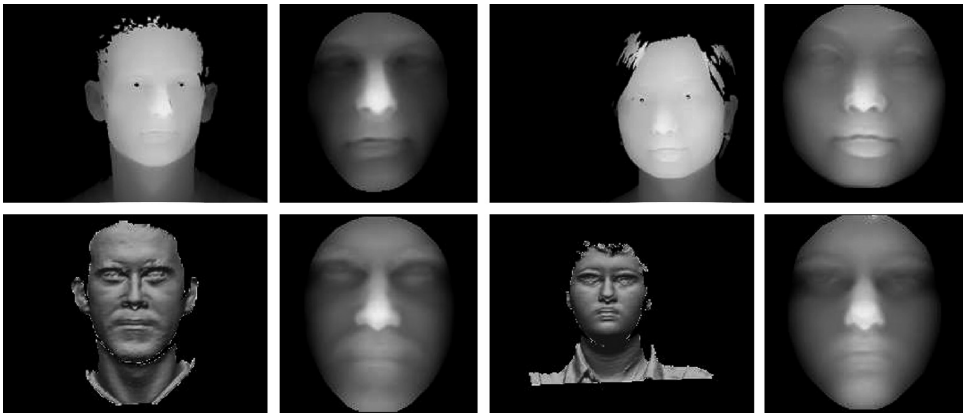

Fig. 2. Examples of normalized faces. The 1st row are two scans before and after normalization from the FRGC v2.0 dataset while the two scans in the 2nd row are from the SHREC 2008 dataset. The faces before normalization exhibit some obvious missing data, spikes and outlier points.

In this work, we propose a novel ARS based facial representation of the normalized facial range images. The ARSs are defined as a set of curves, emanating from the nosetip (chosen as the origin of the facial range images) at an interval of θ radians (where $\theta \in [0, \pi]$, and the total number of ARSs $N = \pi/\theta + 1$). In order to extract the ARSs, we define a binary mask on the xy-plane (see Fig. 3) which can be viewed as the projection of the ARSs. The binary mask contains N paths, and each corresponds to one direction of ARS. Each path contains 20 points, and the distance between any two adjacent points is 3 mm. Consequently, the length of each path is 60 mm. We empirically set this length to 60 mm since beyond this distance the facial data become unreliable due to the presence of hair or missing data (see Fig. 4a). The angle between a path and the x-axis and the distance from the points to the origin are predefined. As a result, it is easy to calculate the x and y coordinates of each point. Next, the depth value (ARS feature value) of these points is determined by bicubic interpolation at the x and y coordinates of their corresponding points in the facial range image. Therefore, an ARS is represented by a 1D feature vector of the computed depth values (see Fig. 4b). Finally, a 3D face is represented by N ARSs, which actually correspond to a set of N 1D feature vectors.

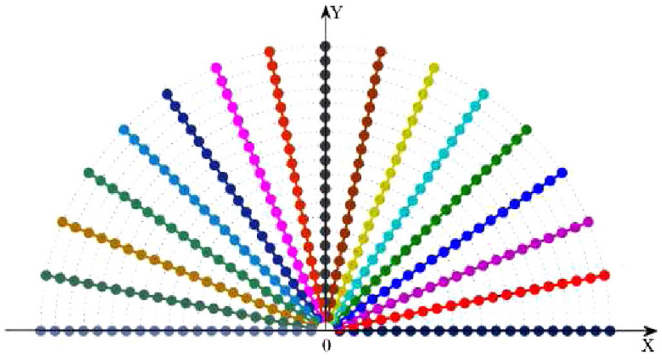

Fig. 3. The binary mask used to extract ARSs. The paths corresponding to different directions are differently colored. (Figure best seen in color).

Note that, we only extract ARSs from the semi-rigid region of the face. The ones from the non-rigid regions are not robust to

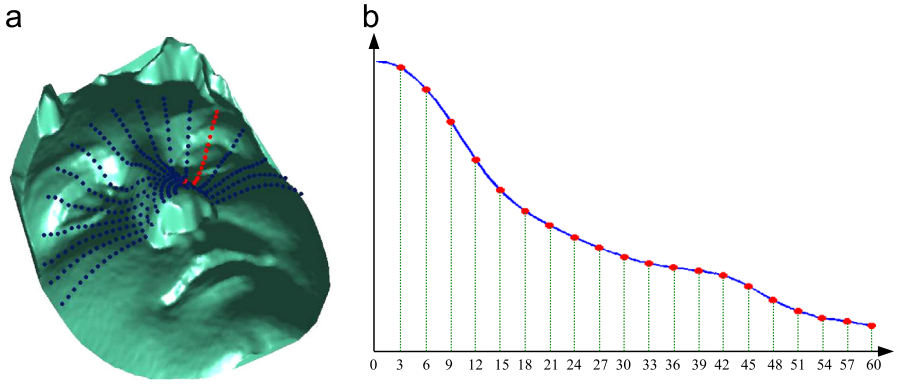

Fig. 4. The extraction of ARSs. (a) 17 ARSs are extracted from the semi-rigid region of the face and (b) an enlarged view of the ARS colored red in (a). (Figure best seen in color). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

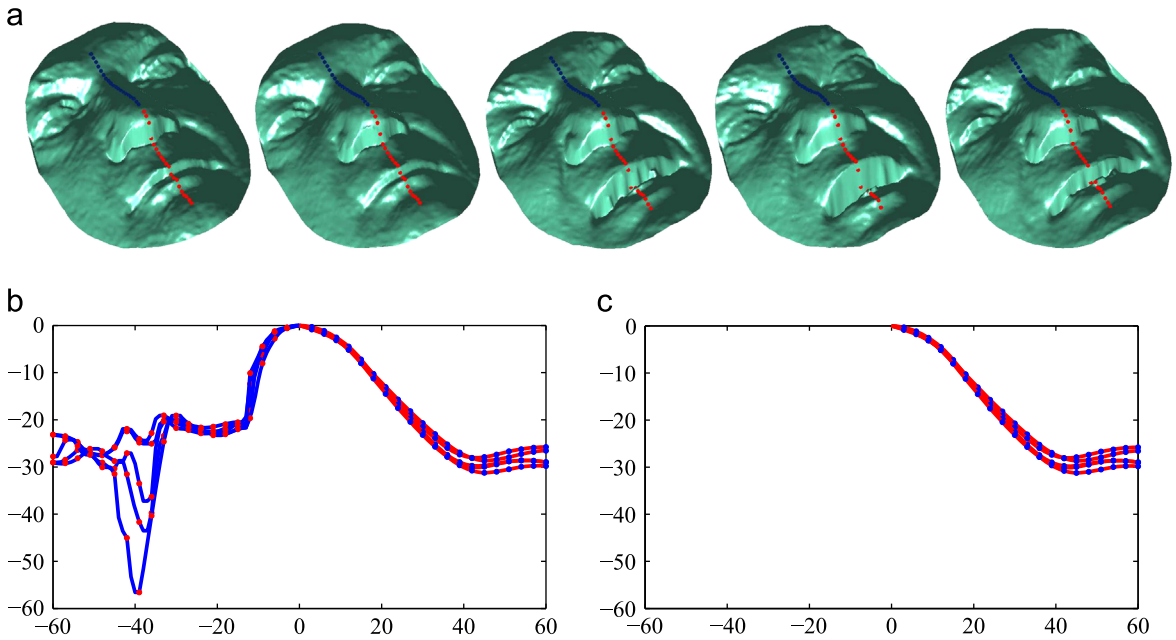

Fig. 5. A comparison between the ARSs extracted from the semi-rigid and the non-rigid regions of a face. (a) Five facial scans of one individual under different expressions. The blue ARSs represent the semi-rigid region while the red ARSs correspond to the non-rigid region. (b) and (c) Enlarged views of ARSs extracted from the semi-rigid and non-rigid regions. (Figure best seen in color). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

facial expressions (e.g. case of mouth movements). A visualization of the ARSs extracted from the semi-rigid and non-rigid regions is presented in Fig. 5. An experimental comparison on the performance of the system using the ARSs extracted from the two regions is further discussed in Section 6.1.3. Increasing the number of N can provide a more accurate representation of the 3D face, but at the expense of an additional computational cost and information redundancy. Taking into consideration the trade-off between computational efficiency and accurate face representation, only 17 ARSs are extracted to represent a 3D face. The effect on the selection of the number of ARSs on the accuracy of face recognition is discussed in Section 6.1.2.

Although the facial scans have been preprocessed, the ARSs can still be affected by minor spike-like noise. A simple average filter is used to smooth the extracted ARSs. In Fig. 6, we extracted the ARSs of two individuals (shown in different colors) in four directions under different expressions (for a better view we converted both facial range images and their corresponding ARSs to shaded pointclouds). It is obvious that the ARSs of the same individual are notably similar even under the distortions caused by expressions. This can therefore be used to discriminate one individual from another. However, the ARS features are still close in the input space, which can lead to a linearly inseparable classification problem. This is particularly the case when the number of samples is excessively large. It suggests that a transformation of the ARS features to a more discriminating form is required. This will be discussed in detail in Section 5.1 .

Other 3D face recognition approaches also used radial facial curves for their facial representation, e.g. in [31,16]. However, their curve extraction was based on the mesh-cutting algorithm. This algorithm crops the curve points defined as the intersections of a predefined slicing plane and the facial mesh triangles. Our proposed ARS based approach has two advantages compared to their work. First, the complicated mesh-cutting and the extraction of the intersection points is not required in our case. This results in a substantial reduction of the computational cost (a detailed analysis of computational cost is given in Section 6.4.2). In addition, rather than using the intersections of mesh triangles, the ARS simply extracts its points using a bicubic interpolation at the locations determined from the binary mask. This results in an exact correspondence with ARSs extracted from different faces, since they all have exactly the same xy-plane projection.

5. A two-stage mapping-based 3D face recognition

In this section, we describe the framework of the proposed two-stage mapping-based 3D face recognition strategy. In the first-stage mapping, N KPCAs are trained (one for each ARS) to transform the ARSs to mid-level feature representations. In the next stage, the resulting mid-level features are combined under a new single feature vector and fed into an SVM to perform face recognition.

The main strengths of our proposed strategy can be summarized as follows. First, we address the linearly inseparable classification problem by transforming the ARSs to a high dimensional nonlinear space using multiple KPCAs. More specifically, due to the high geometric correlations of the facial surfaces, the extracted ARSs are very close to each other in the original feature space. This may lead to a linearly inseparable classification problem. Transforming the ARSs (linearly inseparable data) into a high dimensional nonlinear feature space will resolve the linearly inseparable problem [32]. Second, unlike previous works which commonly train a single PCA model for the complete data, e.g. in [14], we train multiple KPCA models. In our case, the ARS feature vectors which correspond to the same direction are extracted from different faces. As a result, the generation of multiple KPCA models (one for each direction of ARS) results in a more discriminative mapping. Third, the SVM based classifier is capable of handling facial expressions as it is trained using facial scans under various facial expressions (both neutral and nonneutral).

5.1. Kernel principal component analysis (KPCA)

PCA transforms a large number of correlated variables into uncorrelated variables. It has been comprehensively used in face recognition [33,14,34]. Despite its widespread use, PCA still faces some limitations, e.g. it cannot effectively model the nonlinear relationships among variables. To address this problem, PCA has been extended to a kernel form that maps the input data to a high-dimensional space (using the kernel trick) where PCA is applied.

Let $\mathbf{X}=[\mathbf{x}_1 \ldots \mathbf{x}_M] \in \mathbb{R}^{D\times M}$ be the input data and φ be a nonlinear function that maps the input space $\mathbb{R}^D$ to a high-dimensional feature space $\mathbb{F}$, that is $\varphi : \mathbb{R}^D \rightarrow \mathbb{F}$. $\varphi(\mathbf{X})$ is used to represent the mapped data as: $\varphi(\mathbf{X})=[\varphi(\mathbf{x}_1) \ldots \varphi(\mathbf{x}_M)]$. Then, a kernel matrix $\mathbf{K} \in \mathbb{R}^{M\times M}$ is defined by [35]

$$\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j) = (\varphi(\mathbf{x}_i) \cdot \varphi(\mathbf{x}_j)), \quad i,j = 1 \ldots M. \tag{1}$$

Consequently, the nonlinear mapping between the input space and the high-dimensional space is implicitly defined by the kernel. In general, three major types of kernels are commonly used, i.e. polynomial, sigmoid and Radial Basis Function (RBF). The RBF is the most commonly used in the literature due to its high flexibility. In our work, we have selected the RBF kernel as our experiments showed a superior performance of the RBF kernel compared with other kernels (the effect of different kernels on the face recognition performance is discussed in Section 6.2.2). The RBF kernel is defined by

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{2\sigma^2}\right), \tag{2}$$

where $\sigma > 0$. Similar to the conventional PCA, $\mathbf{K}$ is centered to become $\mathbf{K}'$ [35]. The eigenvalues $\lambda=[\lambda_1 \ldots \lambda_M]$ and their corresponding eigenvectors $\mathbf{V}=[\mathbf{v}_1 \ldots \mathbf{v}_M]$ of $\mathbf{K}'$ can be obtained by solving the following eigenvalue problem [35]: $M\lambda\mathbf{A}=\mathbf{K}'\mathbf{a}$. $\mathbf{A}=[\mathbf{a}_1 \ldots \mathbf{a}_M]$ is an orthogonal matrix that is used to derive the eigenvector matrix as $\mathbf{V}=\sum_{i=1}^{M} \mathbf{a}_i \varphi(\mathbf{x}_j)$. For a given data, its corresponding principal components can be derived by projecting its mapped data onto the eigenvectors $\mathbf{V}$. In practice, the dimensionality of the data can be reduced by only retaining the $L(L<M)$ largest eigenvectors. However, in our case the input ARS features are already low dimensional (only 20). Thus, dimensionality reduction is not the major concern. Rather, the aim of applying KPCA here is to solve the linearly inseparable classification problem in the input space. As a result, we retain the largest 20 eigenvectors, and keep the dimensionality of the resulting mid-level feature to be the same as the ARS.

For each direction of the ARSs, we generate one KPCA model (using the faces from the training set) to map all the data (both training and testing faces) to a mid-level feature space in order to obtain a more discriminative mapping. Specifically, since ARSs from 17 different directions are extracted from a 3D face (as described in Section 4), we generate 17 KPCA models to ensure that each KPCA model corresponds to an individual direction. Finally, all mid-level features are concatenated into a single feature vector to represent a face. The resulting feature vector is subsequently fed into an SVM to perform face recognition.

5.2. Support vector machine (SVM)

In recent years, SVM has been widely used due to its excellent classification and generalization capabilities [36,37,12]. Compared with other classification techniques, SVM can minimize the risk of

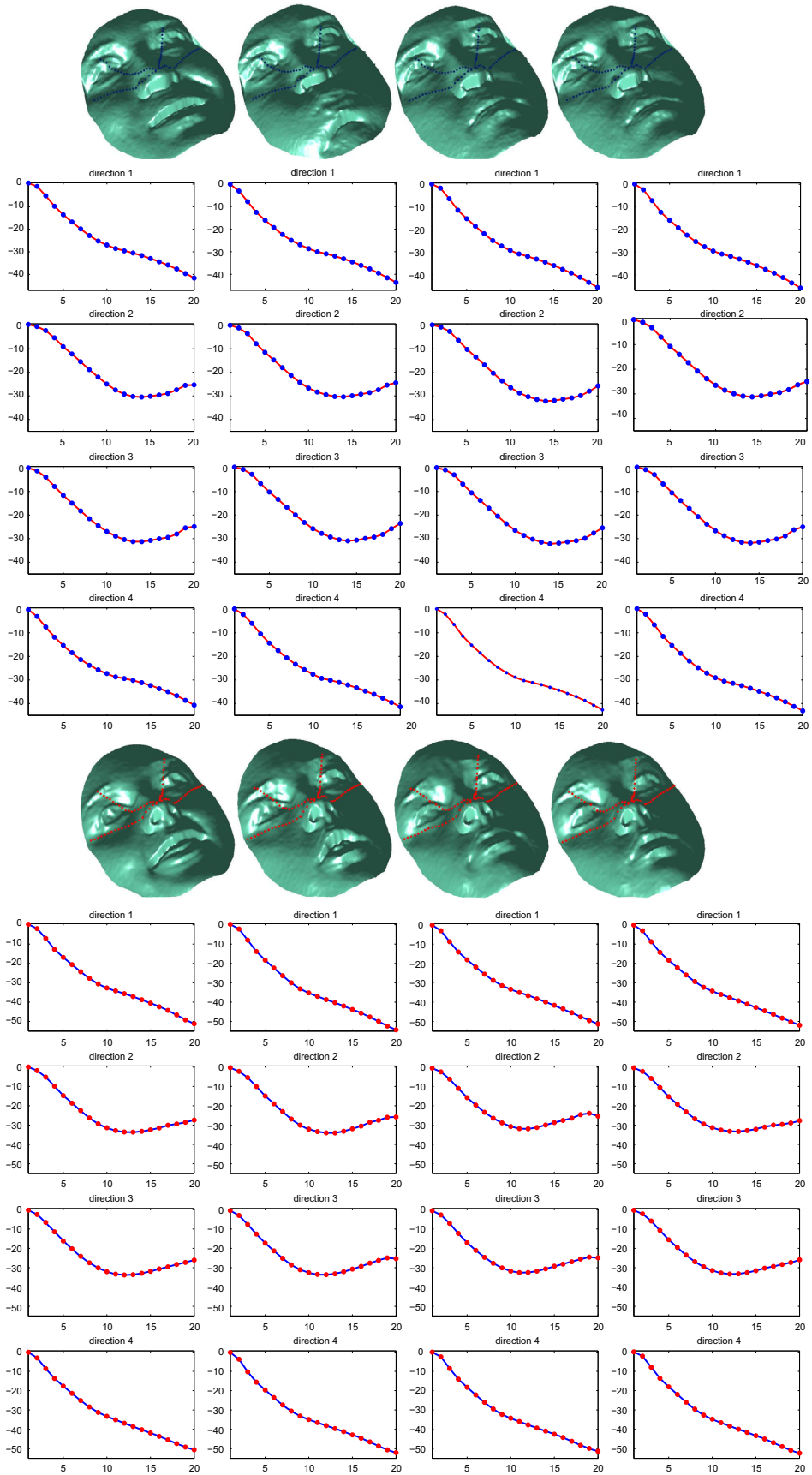

Fig. 6. The ARSs of two individuals extracted in four directions (each individual has four facial scans which are under different facial expressions). A clearer illustration is given in Fig. 7, which shows the similarity of the ARSs belonging to the same individual (Figure best seen in color).

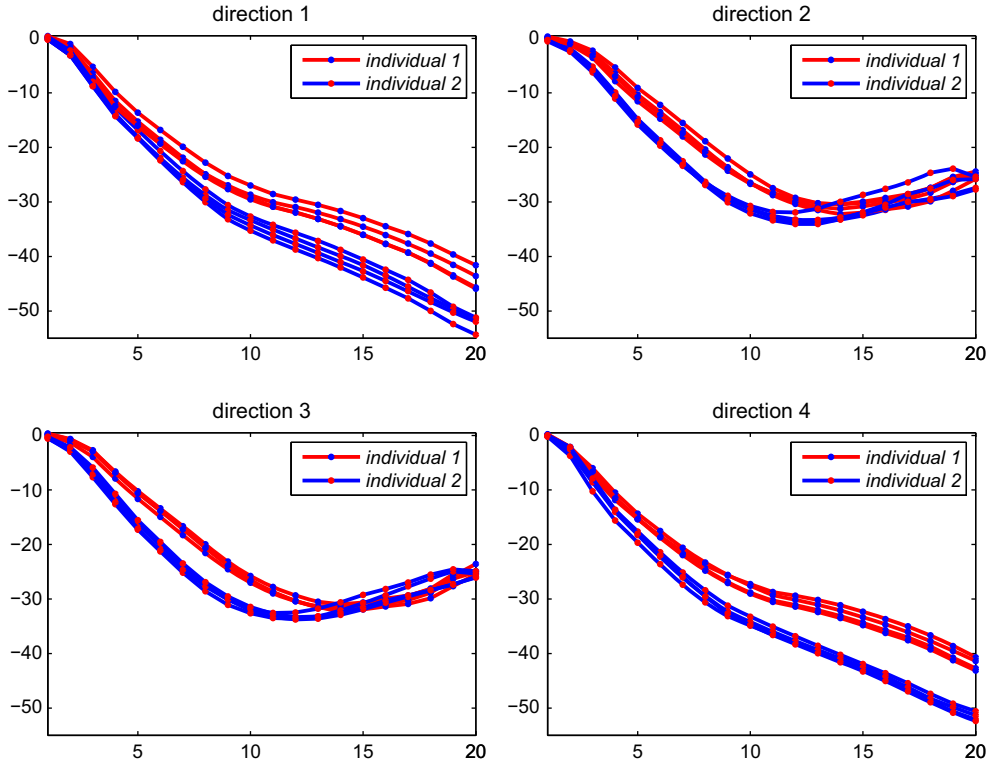

Fig. 7. The ARSs of two individuals plotted according to their directions. It demonstrates that the ARSs corresponding to the same individual cluster together. (Figure best seen in color).

overfitting caused by a small sample size and high-dimensional feature vectors. We first briefly discuss SVM for the case of binary classification. To find a hyper-plane to separate q labeled binary samples, $\mathbf{x}_i \in \mathbb{R}^D$, $i=1,\ldots,q$ and their labels $y_i \in [+1,-1]$, SVM solves the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{w},b,\xi} \quad \left(\frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{l}\xi_i\right) \\ &\text{s.t.} \quad y_i(\mathbf{w}^T\varphi(\mathbf{x}_i)+b) \geq 1-\xi_i \end{aligned}, \tag{3}$$

where $C>0$ is a penalty parameter, $\mathbf{w}$ is the vector of regression coefficients, b is a constant and $\xi_i \geq 0$ are the slack variables. To facilitate the separation of data, the function $\varphi(\mathbf{x})$ is also used to linearly or nonlinearly map the input data to a high-dimensional feature space. In practice, $\varphi(\mathbf{x})$ is also written by a kernel function as: $k(\mathbf{x}_i,\mathbf{x}_j)=\varphi(\mathbf{x}_i)\cdot\varphi(\mathbf{x}_j)$. The RBF kernel is commonly used as it generally shows a reasonable performance [38], particularly when dealing with linearly inseparable data. However, in this work, the proposed ARS features are nonlinearly mapped using KPCA. Therefore, the resulting mid-level features can be expected to become linearly separable. Consequently, the linear kernel is simply used in this work for classification (an experimental evaluation on the effect of the face recognition performance, under different kernels, is discussed in Section 6.2.2). A linear kernel is defined as

$$k(\mathbf{x}_i,\mathbf{x}_j)=\mathbf{x}_i^T\mathbf{x}_j. \tag{4}$$

In addition, the linear kernel is selected because it does not require any parameter selection. It is well known that the performance of SVM is affected by the kernel parameters. These parameters are determined empirically by performing a "grid-search" over a range of values for a cross validation testing set. The whole process is computationally expensive and is avoided in our work by using a linear kernel based SVM classification.

The final goal of SVM is to maximize the distance between the support vectors and the hyper-plane: $f(\mathbf{x})=\mathbf{w}\cdot\mathbf{x}+b$. It is used to separate the data in two classes. For any testing data $\mathbf{x}$, its normalized distance to the hyper-plane can be calculated as $d(\mathbf{x})=(\mathbf{w}\cdot\mathbf{x}+b)/\|\mathbf{w}\|$. Its class label is determined by the sign of $d(\mathbf{x})$ (+1 or −1), which can be viewed as a similarity score (positive polarity).

To extend SVM to solve the multi-class classification problem, two methods are commonly used, i.e. "one-vs-one" and "one-vs-all". A comparison [38] of the two methods shows that they provide similar performance. However, considering the training effort required, the "one-vs-all" method is selected for its low computational cost. Assuming that there are q individuals in the training set, q corresponding SVMs need to be trained (one for each individual). Let $\mathbf{x}$ be a testing face, and $d_k(\mathbf{x})$ ($k=1,2,\ldots q$) be the calculated distance to the kth individual.

For face verification, given a threshold η:

$$\begin{aligned} &\text{Accept} \quad \text{if } \eta - d_k(\mathbf{x}) \geq 0 \\ &\text{Reject} \quad \text{if } \eta - d_k(\mathbf{x}) < 0 \end{aligned}. \tag{5}$$

If the threshold η raises, the Verification Rate (VR) increases, but the False Acceptance Rate (FAR) also increases, and vice versa.

For face identification, the recognition label $y(\mathbf{x})$ can be computed as

$$y(\mathbf{x}) = \arg\max_{1\leq k\leq q}(d_k(\mathbf{x})), \tag{6}$$

where the recognition label is assigned with the individual that yields the largest $d_k(\mathbf{x})$.

6. Experimental results

In this section, we present the performance of our proposed approach when tested on the FRGC v2.0 and the SHREC 2008

datasets and analyze the following aspects: (1) we evaluate the performance of our proposed feature extraction approach, e.g. assessment of the sensitivity of the extracted features with respect to the parameters of the feature extraction algorithm (see details in Section 6.1); (2) we evaluate the two-stage classification module (Section 6.2); (3) we evaluate the robustness of our approach with respect to the nosetip detection errors, variations in expression deformations and environmental effects, e.g. scene background, illumination, and face location (Section 6.3); (4) we finally provide a comparison with the state-of-the-arts along with an analysis of the computational cost (Section 6.4).

FRGC v2.0 is one of the largest publicly available 2D and 3D human face dataset. Its validation partition includes 4007 (fall 2003 and spring 2004 sessions) nearly frontal 3D facial scans belonging to 466 individuals. 2410 facial scans out of 4007 are with a neutral expression. The other 1597 are with nonneutral expressions, including surprise, happy, puffy cheeks, and anger. The 3D face track of SHREC 2008 is comprised of 3D facial scans of 61 individuals (45 males and 16 females). For each individual, seven facial scans were taken under different acquisition viewpoints and various expressions. In particular, two frontal facial scans were acquired in neutral expression, another two neutral scans with slightly rotated postures (looking up or looking down), and the other three scans were acquired with nonneutral expressions (smile, laugh or a random expression). In brief, there are 244 neutral scans while the remaining 183 scans have different facial expressions.

6.1. Evaluation of the proposed feature extraction module

In this section, we assess the effect of the following aspects on the performance of our proposed feature extraction approach. First, we evaluate the recognition performance with different values of the RBF kernel parameter of KPCA. Second, we evaluate the effect of changing the number of the extracted ARSs on the recognition performance. Third, we evaluate the effect of the selected facial expression sensitive regions on the recognition performance.

6.1.1. Effect of the value of the kernel parameter

The first experiment assesses the performance of face recognition under different values of the RBF kernel parameter, σ, of KPCA (see Eq. (2) in Section 5.1). This experiment was performed on the FRGC v2.0 dataset by randomly selecting a maximum of five scans per individual for training and all the remaining scans for testing. Note that, some individuals only have less than five facial scans. For these individuals, we randomly selected one scan for testing and all the remaining scans for training.

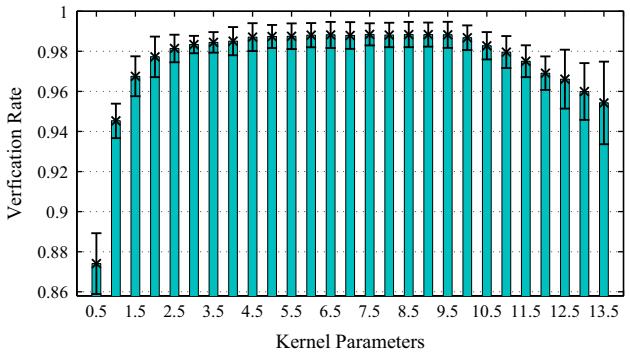

Fig. 8. The average 0.1% FAR VRs and their corresponding 95% CIs versus different RBF kernel parameter of KPCA.

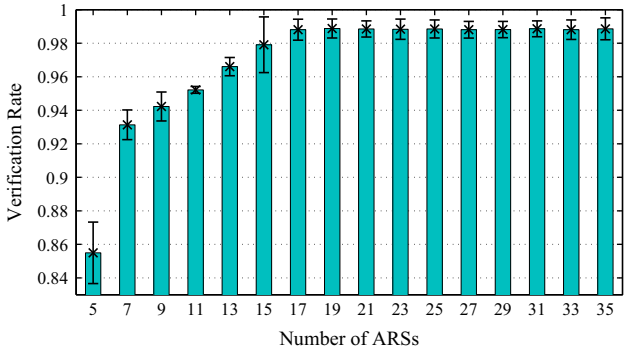

Fig. 9. The average 0.1% FAR VRs and their corresponding 95% CIs versus different numbers of ARSs.

We repeated the experiment 20 times based on the random division of the training/testing data. The KPCA has σ ranging from 0.5 to 20 with an increment of 0.5. The average VRs at 0.1% FAR and the 95% Confidence Intervals (CIs) are reported in Fig. 8. It can be observed that the recognition accuracy has noticeably increased when the value of σ is very low (from 0.5 to 1.5). Then, it levels off at the best performance with σ ranging from 4.5 to 10. Minor performance degradations are noticed if σ continues to increase beyond this range. Because the recognition accuracy is consistent over a wide range of σ, the performance of the proposed system is not highly dependent on the kernel parameter selection. In the following experiments, we use a KPCA with a parameter σ of 6.5.

6.1.2. Effect of the number of ARSs

In this section, we analyze the recognition performance with respect to the number of ARSs. A large number of ARSs is not necessarily desirable as it may include redundant information, which will be time-consuming to analyze. On the other hand, a small number of ARSs will introduce losses in the facial representation which becomes less discriminative, thus affecting the accuracy of face recognition.

The experiment was performed using the same experimental setup as the one discussed in Section 6.1.1. We also repeated the experiment 20 times using the same division (random) of the training/testing sets. The average 0.1% FAR VRs and the corresponding 95% CIs are reported in Fig. 9. We can observe that initially, the average VR increases quickly when the number of ARSs is small (below 15). This is because more ARSs bring in more discriminating information. The figure shows that the VR levels out at its best performance when the number of ARSs is more than 17. This may be due to the reduction of the contribution of the redundant information caused by a large number of ARSs. Based on this experiment and the observations above, 17 ARSs are sufficient to simultaneously provide a high recognition accuracy together with a high computational efficiency. Any further increase in the number of ARSs will only lead to an increased computational cost with no contribution to the recognition performance.

6.1.3. Effect of the contribution of the different facial regions

In this section, a comparative experiment was conducted to demonstrate the face recognition performance using different facial regions, namely, semi-rigid (upper region), non-rigid (lower region) and all (upper+lower regions). The experimental setup was the same as the one described in Section 6.1.1. Note that, we also extracted 17 ARSs from the non-rigid region using the same method as for the semi-rigid region. We repeated the experiments 20 times and report the VRs at 0.1% FAR along with the corresponding 95% CIs in Fig. 10. Two observations can be made. First, the performance of using the semi-rigid region significantly

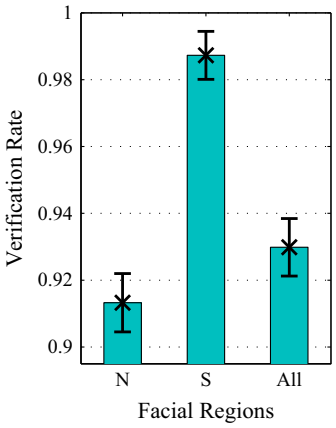

Fig. 10. The average 0.1% FAR VRs and the 95% CIs versus different facial regions. Namely, non-rigid, semi-rigid and all regions (non-rigid+semi-rigid).

outperforms the performance of the non-rigid region in terms of recognition accuracy (an average 0.1% FAR VR of 98.7% versus 91.3%). Second, face recognition using all regions yields a lower accuracy compared to the semi-rigid region, but a higher accuracy than the non-rigid region (an average 0.1% FAR VR of 93.0%). This shows that the non-rigid region highly affects and causes a degradation in the recognition performance.

6.2. Evaluation of the classification module

In this section, we evaluate the effect of the proposed two-stage classification strategy on the overall performance via two experiments. First, we evaluate the recognition performance under different numbers of training/testing samples. Second, we analyze the recognition performance using different kernels for both KPCA and SVM.

6.2.1. Different number of training scans

Since our proposed 3D face recognition approach is based on the machine learning techniques (KPCA and SVM), its performance depends on the diversity and the total number of samples in the training set. By using training faces with both neutral and non-neutral expressions, the learned model should be able to deal with the expression variation problem. As a result, increasing the number of training samples (randomly selecting faces mixed with neutral and nonneutral expressions) can result in an improved recognition performance. We therefore need to explore the effect on the recognition performance under different numbers of training and testing samples. This experiment was also conducted on the FRGC v2.0 dataset, in which the number of facial scans belonging to each individual varies from 1 to 22 (see details in Table 1). We divided the dataset as follows. We randomly included m (varying from 2 to 7) scans per individual in the training set and the remaining scans were included in the testing set. For those individuals whose facial scans are less than or equal to m, we also randomly selected one scan for testing and all the remaining scans were used for training (following the same procedure in Section 6.1.1). We repeated the experiment 20 times using a random division of training and testing data with m varying from 2 to 7. The average VRs at 0.1% FAR and their corresponding 95% CIs are shown in Fig. 11.

It can be observed that the proposed approach achieves the lowest average VR for the training set with a maximum of two training scans per individual. By increasing the number of training scans, the recognition accuracy has improved. When a maximum of five scans per individual are used for training, there is a high probability of having a training set mixed with both neutral and nonneutral faces for each individual. Thus, the trained SVM model is able to respond well to different expressions of faces. The recognition accuracy (when a maximum of five scans per individual are used for training) is significantly improved compared to when a lower value of m is used (a VR of 98.7% at 0.1% FAR). Nevertheless, our approach still achieves a VR of 91.5% at 0.1% FAR for the first and the most challenging experiment (a maximum of two scans per individual for training), which clearly demonstrates the robustness of our proposed approach.

Table 1
Statistical information of the FRGC v2.0 dataset. n is the number of the individuals with at least m scans.

m	1	2	3	4	5	6	7	8	10	15	22
n	466	410	384	352	316	285	257	230	186	95	3

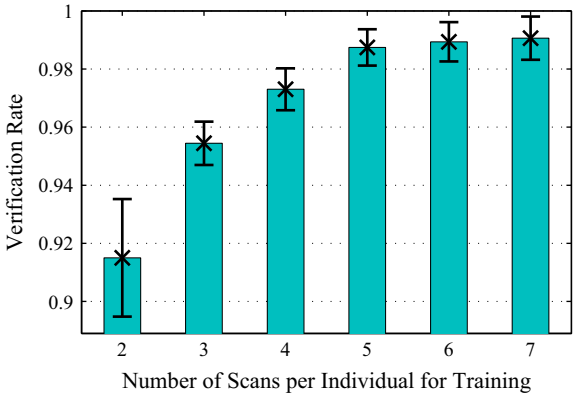

Fig. 11. The average 0.1% FAR VRs and their 95% CIs versus the number of facial scans per individual for training.

6.2.2. Effect of the choice of the kernels

In this section, we analyze the performance of face recognition under different kernels, for KPCA and SVM. The detailed cross comparison results are listed in Table 2. The kernels listed vertically correspond to the KPCA and those listed horizontally correspond to the SVM. We also report the results when using the conventional PCA and the raw ARS features without any mid-level feature extraction. For SVM, a comparison between three kernels (linear, polynomial and RBF) is presented. This comparative experiment was also performed on the FRGC v2.0 dataset following the experimental protocol described in Section 6.1.1. We repeated the experiment 20 times and report the average VRs at 0.1% FAR and their corresponding 95% CIs in Table 2.

Several observations can be made based on these experimental results. First, the RBF kernel outperforms other kernels in terms of recognition accuracy. Second, the results prove that the extraction of mid-level features can significantly improve the face recognition performance. The results also show that the input ARS features that are nonlinearly mapped by RBF kernel (KPCA) to a high-dimensional space, can solve the linearly inseparable classification problem in the input feature space. Third, the combinations of R-PCA+L-SVM and R-PCA+R-SVM achieve nearly the same recognition performance. However, taking into consideration the required effort of parameter selection for SVM, the R-PCA+L-SVM is a promising choice to achieve computational efficiency. Finally, based on our experimental results, the polynomial and sigmoid kernels have revealed not to be a suitable choice.

Table 2
The average 0.1% FAR VRs and the 95% CIs using different kernels, i.e. linear (L), polynomial (P), sigmoid (S) and RBF (R), for both KPCA and SVM. The kernels listed vertically correspond to KPCA while the horizontal ones correspond to SVM. We also compare the face recognition performance of our approach with the conventional PCA and raw ARS features.

	L-SVM	P-SVM	S-SVM	R-SVM
Non	88.5% (87.2%-89.7%)	45.7% (39.5%–51.9%)	29.9% (24.1%–35.7%)	91.4% (90.7%–92.1%)
PCA	90.2% (89.3%–91.1%)	36.3% (24.5%–48.2%)	88.3% (86.9%–89.7%)	93.9% (93.2%–94.5%)
P-PCA	87.2% (80.8%–93.6%)	31.5% (24.4%–38.7%)	56.6% (45.7%–67.6%)	89.8% (88.1%–91.4%)
S-PCA	71.7% (60.9%–82.5%)	56.6% (48.2%–65.1%)	67.8% (63.0%–72.5%)	77.6% (71.9%–83.3%)
R-PCA	**98.7%** **(97.9%–99.5%)**	74.3% (65.3%–83.4%)	91.2% (89.5%–92.9%)	**98.5%** **(97.8%–99.2%)**

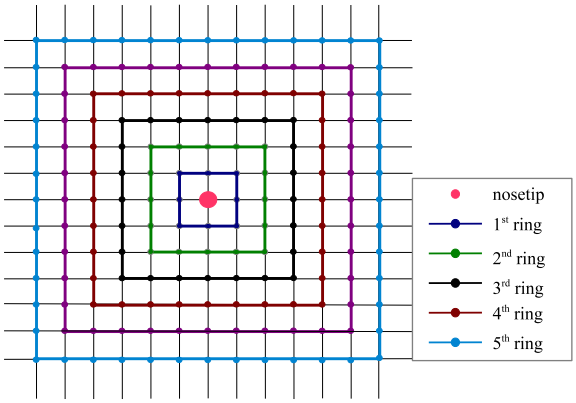

Fig. 12. The five nosetip relocation rings. For the *j*th ring, the center of all training and testing images are randomly displaced to a new position which lies at an arbitrary point on that ring. (Figure best seen in color).

6.3. System robustness analysis

In this section, we assess the robustness of our proposed approach through three experiments. We first analyze the robustness of our approach to the errors of the nosetip detection. Second, we test the proposed approach under facial expression variations. Third, we report the robustness of the proposed approach under environmental variations (e.g. background variations and illumination conditions).

6.3.1. Robustness to nosetip detection errors

The proposed approach requires the detection of the nosetip (used as a fiducial point). An experiment was conducted to demonstrate the recognition performance with respect to the errors of nosetip detection. We manually altered the positions of the nosetips of all the normalized facial scans from their actually detected positions. Rings have been designed on a uniform square grid for this manual alteration for the nosetip locations. An illustration of this nosetip relocation scheme is shown in Fig. 12. Five experiments were conducted to test the effect of five different rings (with a gradually increasing ring size) on the recognition performance. For an experiment corresponding to the *j*th ring (where $j=1,2,\ldots,5$), the center (nosetip position) of all the training and testing facial range images were randomly displaced to a new position which lied at an arbitrary point on that ring. This manual random displacement introduced errors in nosetip locations which in turn should affect the face recognition performance. An increase of the ring size results in a more significant error in the nosetip location and consequently a more challenging experimental setup.

Table 3
The average 0.1% FAR VRs and the 95% CIs produced by relocating the nosetip at a random point at each ring.

Rings	Average VR at 0.1% FAR	95% CI
ring-1	98.0%	(97.4%–98.7%)
ring-2	97.0%	(96.2%–97.8%)
ring-3	96.9%	(96.2%–97.5%)
ring-4	95.6%	(94.9%–96.4%)
ring-5	95.3%	(93.0%–97.7%)

We then randomly selected a maximum of five scans per individual (following the procedure in Section 6.1.1) to form a training set. The remaining scans were used for the testing set. We also repeated each of the five experiments 20 times and report the average VRs at 0.1% FAR and the corresponding 95% CI in Table 3 using the facial scans with the relocated nosetip on the first, second, third, fourth, and fifth rings (i.e. with an increasing level of difficulty). As expected, the errors in the detection of the nosetip do affect the recognition accuracy of our proposed approach. However, in the most challenging part of this experiment (the 5th ring), our proposed approach still achieves an average 0.1% FAR VR of 95.3%. This clearly demonstrates the robustness of our proposed approach against the errors in detecting and locating the position of the nosetip.

6.3.2. Effect of facial expressions

In this section, we perform experiments to assess the effect of expression variations on the recognition performance. First, the 2410 neutral scans of 466 individuals from the FRGC v2.0 dataset and the 244 neutral scans of 61 individuals from the SHREC 2008 dataset were used to form the training sets. The remaining 1597 nonneutral FRGC v2.0 scans and the 183 nonneutral SHREC 2008 scans were used to form the testing sets. This type of arrangement is called the "neutral vs. nonneutral" experiment. Second, by swapping the training and testing sets of the "neutral vs. nonneutral" experiment, we performed the so called "nonneutral vs. neutral" experiment.

The Cumulative Match Characteristic (CMC) curves of the face identification experiments are shown in Fig. 13. The rank-1 Identification Rates (IRs) obtained in the cases of "neutral vs. nonneutral" and "nonneutral vs. neutral" on the FRGC v2.0 dataset were 93.4% and 94.1%, respectively. On the SHREC 2008 dataset, we obtained rank-1 IRs of 90.7% and 88.1% under the same cases, respectively. Fig. 14 illustrates the Receiver Operation Curves (ROCs) of the face verification experiments. When the proposed approach was tested on the FRGC v2.0 dataset, it achieved a VR of

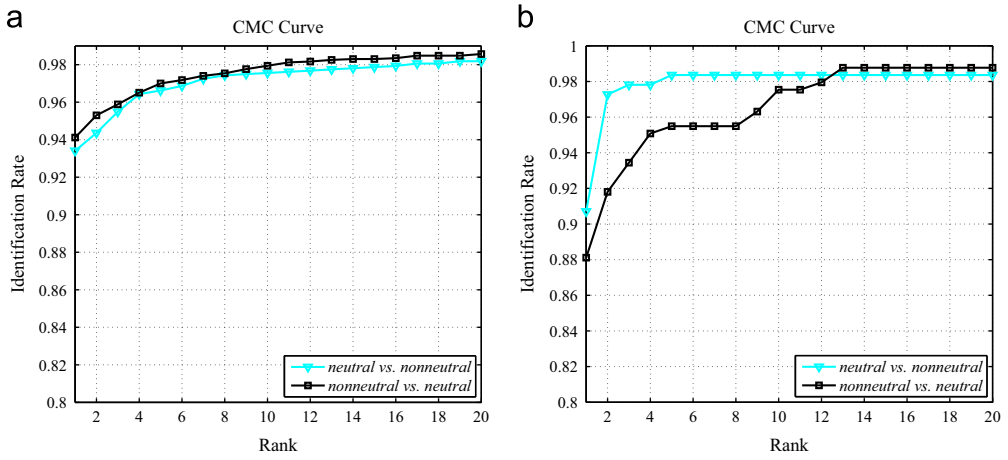

Fig. 13. The CMC curves of the "neutral vs. nonneutral" and "nonneutral vs. neutral" experiments. (a) The results on the FRGC v2.0 dataset and (b) the results on the SHREC 2008 dataset. The resulting rank-1 IRs of the "neutral vs. nonneutral" experiment are 93.4% and 90.7% respectively.

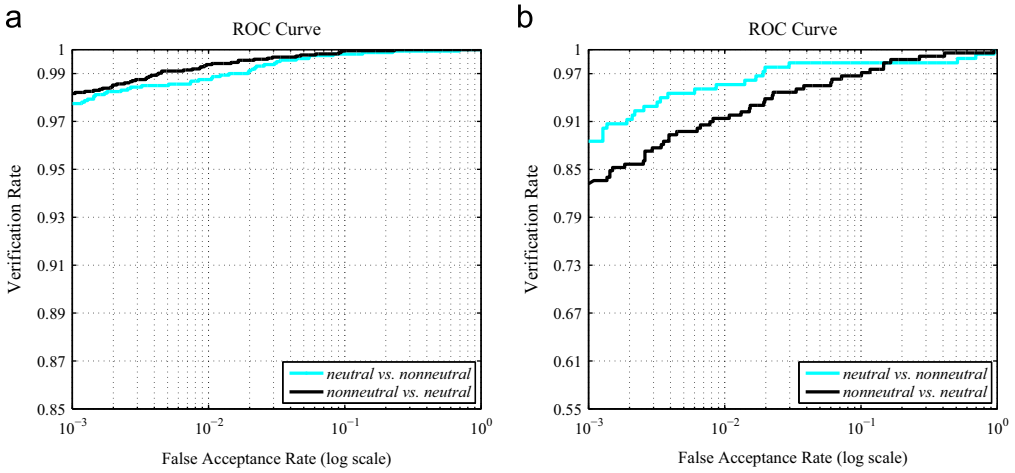

Fig. 14. The ROC curves of the "neutral vs. nonneutral" and "nonneutral vs. neutral" experiments. (a) The results on the FRGC v2.0 dataset and (b) the results on the SHREC 2008 dataset. The resulting 0.1% FAR VRs of "neutral vs. nonneutral" experiment are 97.8% and 88.5% respectively.

Fig. 15. Examples of the selected faces under large facial expression deformations from the FRGC v2.0 dataset.

97.8% and 98.2% at 0.1% FAR for the "neutral vs. nonneutral" and "nonneutral vs. neutral" experiments, respectively. On the SHREC 2008 dataset, we obtained 0.1% FAR VRs of 88.5% and 83.2% under the same cases, respectively.

Furthermore, in order to evaluate the robustness of our proposed approach with respect to the large expression deformations, we conducted an experiment that was first proposed by Berretti et al. [15]. Large facial expressions are selected to correspond to the facial scans with an open mouth because it can significantly change the topology of the facial surface (see Fig. 15). We used all the neutral scans of the FRGC v2.0 dataset as the training set and the 816 facial scans with an open mouth as the testing set to perform face

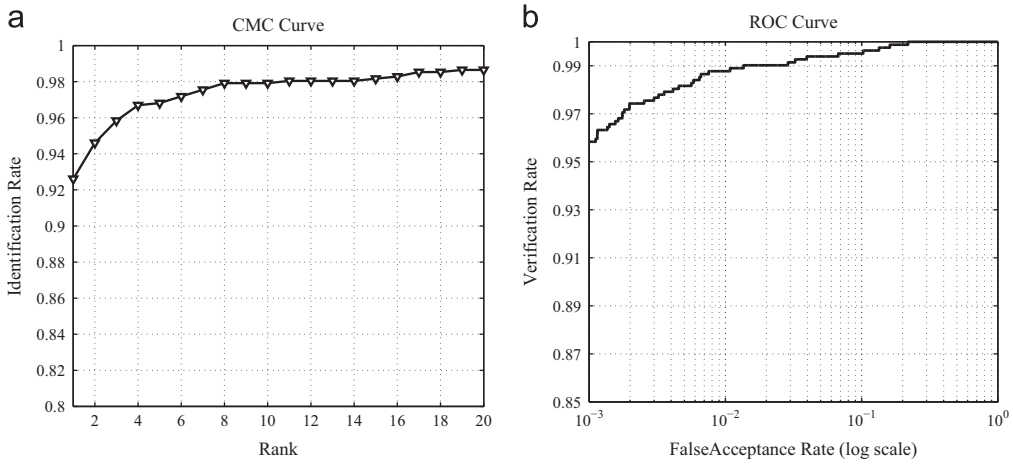

Fig. 16. The "neutral vs. large mouth deformation" experiment. (a) The CMC curve and (b) the ROC curve. The testing set comprised of 816 selected facial scans with an open mouth from the FRGC v2.0 dataset. The resulting rank-1 IR and 0.1% FAR VR are 92.6% and 95.8% respectively.

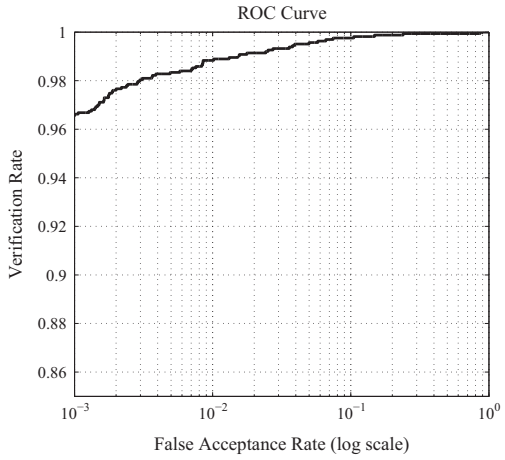

Fig. 17. ROC curve of the ROC III experiment of the FRGC v2.0 program. The resulting 0.1% FAR VR is 96.7%.

recognition.[1] The ROC and CMC curves are presented in Fig. 16. Compared to the "neutral vs. nonneutral" experiment, the resulting rank-1 IR and the 0.1% FAR VR decrease by 0.8% and 2%, respectively. However, the proposed approach still achieves a rank-1 IR of 92.6% and a 0.1% FAR VR of 95.8% and is able to accurately recognize the facial scans with an open mouth. This illustrates the robustness of the proposed approach under severe facial deformations.

6.3.3. *Effect of environmental variations*

This experiment is actually part of the standard FRGC v2.0 protocols (ROC III experiment) as described in [25]. The training set was generated using the facial scans taken during the fall of 2003, and the testing set contained facial scans taken during the spring of 2004. The intentional variations (e.g. the type of background, the location of the face, and the intentional lighting variations) as well as the time interval between both sets make the experiment challenging. We conducted the face verification experiment and the ROC curve is shown in Fig. 17. In the ROC III experiment, the proposed approach achieves a 0.1% FAR VR of 96.7%.

[1] All 816 facial scans were manually selected by Berretti et al. [15], and a list is available at: http://www.dsi.unifi.it/~berretti/frgcv2.0/

6.4. *Performance comparison and timing analysis*

In this section, we compare our proposed 3D face recognition approach with the state-of-the-arts along with a computational analysis of each step.

6.4.1. *Comparison with the state-of-the-arts*

For a fair comparison, we only focus on the approaches which were tested on the FRGC v2.0 and SHREC 2008 datasets. Table 4 provides a comparison of 0.1% FAR VRs with the state-of-the-arts in the cases of "neutral vs. nonneutral" and the ROC III experiments on the FRGC v2.0 dataset along with the "neutral vs. nonneutral" experiment on the SHREC 2008 dataset. The results of other approaches are directly reported from their publications. Also, for a fair comparison, in the case of 3D+2D multimodal approaches, we only consider the results of the 3D modality.

For the FRGC v2.0 standard protocol "neutral vs. nonneutral" and ROC III experiments, we obtained a 0.1% FAR VR of 97.8% and 96.7%, respectively. These results are comparable to, or even better than, the published results as shown in Table 4a. Note that, this comparison is only done with approaches which use "neutral vs. nonneutral" and ROC III experiments, while the aim of this paper is to propose a 3D face recognition system using both neutral and nonneutral faces of different individuals for training in order to address the expression variation problem. In this case (i.e. when we combine the neural and nonneutral scans in our training set), our approach achieves a superior recognition performance compared to the case where we only use neutral scans for training (Section 6.2.1). For the "neutral vs. nonneutral" experiment of SHREC 2008, we obtained a 0.1% FAR VR of 88.5%, outperforming all the other approaches.

6.4.2. *Computational performance*

Most of the existing 3D face recognition approaches were based on the surface registration or complex feature matching techniques. Although some of them could perform face verification in near real-time, face identification still remains a challenge, particularly when the size of the gallery is large. In fact, each probe facial scan needs to be matched one-to-one with every facial scan in the gallery. Moreover, most of the existing approaches only reported the time cost of the testing phase, overlooking the time cost of training. However, the computational cost during the training phase is also an important concern for an efficient and practical 3D face recognition system especially in real-wold

Table 4
Comparison with the state-of-the-arts. (a) 0.1% FAR VRs of the "neutral vs. nonneutral" and the ROC III experiments on the FRGC v2.0 dataset. (b) 0.1% FAR VRs of the "neutral vs. nonneutral" experiment on the SHREC 2008 dataset.

(a)

Approach	Neutral vs. nonneutral	ROC III
Mian et al. [14]*	92.7%	n/a
Mian et al. [9]*	97.0%	n/a
Al-Olsaimi et al. [19]	97.8%	94.1%
Xu et al. [24]	n/a	95.3%
Queirolo et al. [39]	n/a	96.6%
Faltermier et al. [7]*	n/a	94.8%
Berretti et al. [15]	91.4%	n/a
Kakadiaris et al. [17]	n/a	97.0%
This paper	**97.8%**	**96.7%**

(b)

Approach	Neutral vs. nonneutral (%)
ter Haar et al. [40]	39.3
Nair et al. [41]	42.6
Amberg et al. [42]	74.9
Berretti et al. [15]	80.9
This paper	**88.5**

The mark "*" means that only the results of their 3D modality are shown from their 3D+2D multimodal approaches.

Table 5
Face identification time of each step of our proposed 3D face recognition approach.

Step		Number of faces	Individuals	Time cost (s)
Face normalization		1	1	0.28
ARSs extraction				0.034
Training	KPCA	2410	466	16.8
	SVM			6.5
Testing	KPCA	1597		0.67
	SVM			5.4

applications, where the gallery needs to be updated with new enrollments quite often.

We report the consumed time of each step of our proposed approach in the case of "neutral vs. nonneutral" experiment on the FRGC v2.0 dataset. In addition to the time cost for model training and testing, we also report the time cost for the face normalization and ARSs extraction. All the reported consumed time was obtained using a MATLAB and C++ implementation running on a PC with an INTEL Core 2 Quad CPU and 8 GB RAM, and the results are shown in Table 5. These results show that the face normalization and ARSs extraction are very fast (0.28 s and 0.034 s, respectively for each face). During the training phase, the first-stage mapping requires 16.8 s to convert the ARS features of 2410 facial scans to mid-level features. Then, in the second-stage mapping, the SVM takes 6.5 s to train the model using 2410 samples belonging to 466 individuals (classes). During the testing phase, the first-stage mapping requires only 0.67 s to process 1597 facial scans, and in the second-stage mapping SVM takes only 5.4 s to identify 1597 facial scans of 466 individuals. Therefore, our proposed approach only requires 6.07 s in total in the testing phase of the "neutral vs. nonneutral" face identification experiment. In order to make a further comparison, we analyze the computational efficiency of several state-of-the-arts in the following paragraph.

The approach proposed by Kakadiaris et al. [17] is relatively computationally efficient as well as accurate for the task of face recognition. It requires 15 s for data preprocessing and can perform 1000 one-to-one face matching per second. Therefore, such approach requires a total of 744.2 s to perform the same face identification experiment as adopted in this paper. Note that, we have not considered the time required to rank the obtained similarity scores. Mian et al. [9] proposed a rejection classification algorithm to improve the computational efficiency, but their technique involves the usage of the ICP algorithm for 3D face matching. The ICP algorithm is an iterative and computationally expensive process. Their approach requires 4 s for matching a single probe with a gallery of 466 faces. Their approach would therefore require 6388 s if evaluated on our face identification experiment. Another highly accurate system [7] also used the ICP algorithm to match multiple regions of the faces. The authors reported a time of 7.5 s for preprocessing and 2.3 s for the matching of a probe-gallery pair. Following our experimental settings, their approach would therefore require 1,711,664.6 s. This approach is very slow for face identification, especially, when the number of scans in the gallery is quite large (466 in our case). Recently, Wang et al. [11] proposed a local shape difference boosting based 3D face recognition approach. It requires 0.66 s for matching a probe with a gallery of 466 (FRGC v2.0 dataset). Their approach costs a total of 1054 s for the same face identification experiment as conducted in this paper. As an extension of their previous work [43], Ballihi et al. [16] proposed a 3D face recognition approach based on the differential geometrical facial curves. Their approach involves the AdaBoost algorithm to improve the efficiency. However, it still requires 0.68 s to compare one pair of faces. Therefore, such algorithm needs 506,057.4 s to perform our face identification experiment (the time for similarity scores ranking is not considered), which is more than 80,000 times slower than our proposed approach.

7. Conclusions

In this work, we proposed a novel 3D face recognition approach based on local geometrical signatures called facial Angular Radial Signature to effectively approximate the semi-rigid region of a 3D face. We employed KPCA to map the raw ARS facial features to mid-level features to improve the discriminating power. Next, the resulting mid-level features were combined into one single feature vector and fed into an SVM to perform face recognition. We conducted comprehensive experiments on the FRGC v2.0 and SHREC 2008 datasets, and a superior recognition performance was achieved. Based on our experimental results, three conclusions can be drawn. First, using facial scans with different

expressions for training, the proposed ARS based framework can accommodate expression variations when recognizing an input face. Second, the semi-rigid region of the face contains the most effective and discriminating information, and a more robust 3D face recognition can be achieved when dealing with expressions. Moreover, a nonlinear mapping of the features (that are close in the input space) to a high-dimensional space can resolve the linearly inseparable classification problem, and results in an improved discriminating power. We also provided a computational performance analysis that demonstrates the efficiency of our proposed approach. Because the one-to-one surface registration or complex feature matching between a probe face to every gallery face is not required, the proposed approach enables rapid face recognition and can be applied in practical systems.

Conflict of interest

None declared.

Acknowledgments

The first author would like to thank the China Scholarship Council (CSC) for its financial support toward a PhD scholarship. We also thank the Australian Research Council (ARC) for its financial support under the Discovery Grant DP110102166 and the Linkage Project LP120100595.

References

[1] W. Zhao, R. Chellappa, P. Philips, A. Rosenfeld, Face recognition: a literature survey, ACM Computing Surveys 35 (4) (2003) 399–458.
[2] A. Abate, M. Nappi, D. Riccio, G. Sabatino, 2D and 3D face recognition: a survey, Pattern Recognition Letters 28 (14) (2007) 1885–1906.
[3] C. Xu, Y. Wang, T. Tan, L. Quan, Depth vs. Intensity: which is more important for face recognition?, in: International Conference on Pattern Recognition, vol. 4, 2004, pp. 342–345.
[4] K. Bowyer, K. Chang, P. Flynn, A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition, Computer Vision and Image Understanding 101 (1) (2006) 1–15.
[5] P. Besl, N. Mckay, Reconstruction of real-world objects via simultaneous registration and robust combination of multiple range images, IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (2) (1992) 239–256.
[6] K. Chang, K. Bowyer, P. Flynn, Multiple nose region matching for 3D face recognition under varying facial expression, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (10) (2006) 1695–1700.
[7] C. Faltemier, K. Bowyer, P. Flynn, A region ensemble for 3D face recognition, IEEE Transactions on Information Forensics and Security 13 (1) (2008) 62–73.
[8] R. Sablatnig, M. Kampel, Model-based registration of front- and backviews of rotationally symmetric objects, Computer Vision and Image Understanding 87 (1–3) (2002) 90–103.
[9] A. Mian, M. Bennamoun, R. Owens, An efficient multimodal 2D and 3D hybrid approach to automatic face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (11) (2007) 1927–1943.
[10] F. Al-Osaimi, M. Bennamoun, A. Mian, Integration of local and global geometrical cues for 3D face recognition, Pattern Recognition 41 (3) (2008) 1030–1040.
[11] Y. Wang, J. Liu, X. Tang, Robust 3D face recognition by local shape difference boosting, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (10) (2010) 1858–1870.
[12] Y. Lei, M. Bennamoun, A. El-Sallam, An efficient 3D face recognition approach based on the fusion of novel local low-level features, Pattern Recognition 46 (1) (2013) 24–37.
[13] Y. Guo, F. Sohel, M. Bennamoun, M. Lu, J. Wan, Rotational projection statistics for 3D local surface description and object recognition, International Journal of Computer Vision. 105 (1) (2013) 63–86.
[14] A. Mian, M. Bennamoun, R. Owens, Keypoint detection and local feature matching for textured 3D face recognition, International Journal on Computer Vision 79 (1) (2007) 1–12.
[15] S. Berretti, A. Del Bimbo, P. Pala, 3D face recognition using isogeodesic stripes, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (12) (2010) 2162–2177.
[16] L. Ballihi, B. Ben Amor, M. Daoudi, A. Srivastava, D. Aboutajdine, Boosting 3D-geometric features for efficient face recognition and gender classification, IEEE Transactions on Information Forensics and Security 7 (6) (2012) 1766–1779.
[17] I. Kakadiaris, G. Passalis, G. Toderici, N. Murtuza, T. Theoharis, Three-dimensional face recognition in the presence of facial expression: an annotated deformable model approach, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (4) (2007) 640–649.
[18] X. Lu, A. Jain, Deformation modeling for robust 3D face matching, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (8) (2008) 1346–1357.
[19] F. Al-Osaimi, M. Bennamoun, A. Mian, An expression deformation approach to non-rigid 3D face recognition, International Journal of Computer Vision 81 (3) (2009) 302–316.
[20] C. Zhong, Z. Sun, T. Tan, Robust 3D face recognition using learned visual codebook, in: IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–6.
[21] X. Lu, D. Colbry, A. Jain, Three-dimensional model based face recognition, in: International Conference on Pattern Recognition, vol. 1, 2004, pp. 362–366.
[22] S. Gupta, M. Markey, A. Bovik, Anthropometric 3D face recognition, International Journal of Computer Vision 90 (3) (2010) 331–349.
[23] Y. Sun, X. Chen, M. Rosato, L. Yin, An evaluation of multi-model 2D+3D bimetrics, Man and Cybernetics, Part A: Systems and Humans 40 (3) (2010) 461–474.
[24] C. Xu, S. Li, L. Tan, Automatic 3D face recognition from depth and intensity Gabor features, Pattern Recognition 42 (9) (2009) 1895–1905.
[25] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, IEEE, 2005, pp. 947–954.
[26] L. Yin, Z. Wei, Y. Sun, J. Wang, M. Rosato, A 3D facial expression database for facial behavior research, in: International Conference on Automatic Face and Gesture Recognition, 2006, pp. 211–216.
[27] X. Peng, M. Bennamoun, A. Mian, A training-free nose tip detection method from face range images, Pattern Recognition 44 (3) (2011) 544–558.
[28] C. Liu, H. Wechsler, Independent component analysis of Gabor features for face recognition, IEEE Transactions on Neural Networks 14 (4) (2003) 919–928.
[29] H. Mohammadzade, D. Hatzinakos, Iterative closest normal point for 3D face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (2) (2012) 381–397.
[30] F. Al-Osaimi, M. Bennamoun, A. Mian, Spatially optimized data-Level fusion of texture and shape for face recognition, IEEE Transactions on Image Processing 21 (2) (2012) 859–872.
[31] F. ter Haar, R. Veltkamp, A 3D face matching framework for facial curves, Graphical Models 71 (2) (2009) 77–91.
[32] S. Haykin, Neural Networks: A Comprehensive Foundation, 2nd edition, Prentice Hall, 1999.
[33] K. Chang, K. Bowyer, P. Flynn, Face recognition using 2D and 3D facial data, in: ACM Workshop on Multimodal User Authentication, 2003, pp. 25–32.
[34] L. Qiao, S. Chen, X. Tan, Sparsity preserving projections with applications to face recognition, Pattern Recognition 43 (1) (2010) 331–341.
[35] B. Schölkopf, A. Smola, K. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural computation 10 (5) (1998) 1299–1319.
[36] S. Kim, Y. Park, K. Toh, S. Lee, SVM-based feature extraction for face recognition, Pattern Recognition 43 (8) (2010) 2871–2881.
[37] N. Khan, R. Ksantini, I. Ahmad, B. Boufama, A novel SVM+NDA model for classification with an application to face recognition, Pattern Recognition 45 (1) (2012) 66–79.
[38] C. Hsu, C. Lin, A comparison of methods for multi-class support vector machines, IEEE Transactions on Neural Networks 13 (2002) 415–425.
[39] C. Queirolo, L. Silva, O. Bellon, 3D face recognition using simulated annealing and the surface interpenetration measure, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (2) (2010) 206–219.
[40] F. ter Haar, R. Veltkamp, A 3D face matching framework, in: IEEE International Conference on Shape Modelling and Applications, IEEE, 2008, pp. 103–110.
[41] P. Nair, A. Cavallaro, Registration and retrieval of 3D faces using a point distribution model, in: IEEE International Conference on Shape Modelling and Applications, IEEE, 2008, pp. 257–258.
[42] B. Amberg, R. Knothe, T. Vetter, Expression invariant 3D face recognition with a morphable model, in: IEEE International Conference on Automatic Face and Gesture Recognition, IEEE, 2008, pp. 1–6.
[43] L. Ballihi, B. Ben Amor, M. Daoudi, A. Srivastava, D. Aboutajdine, Selecting 3D curves on the nasal surface using AdaBoost for person authentication, in: Eurographics Workshop on 3D Object Retrieval, 2011, pp. 1–5.

Yinjie Lei received his MS degree from Sichuan University, P.R. China in the area of image processing. He received his Ph.D degree in computer science from the University of Western Australia in 2013. His research interests include image and text understanding, 3D face processing and recognition, 3D modeling, machine learning and statistical pattern recognition.

Mohammed Bennamoun received the MS Degree in control theory from Queen's University, Kingston, Canada, and the Ph.D. degree in computer vision from Queen's/QUT, Brisbane, Australia. He is currently a Winthrop Professor at the University of Western Australia, Crawley, Australia. His research interests include control theory, robotics, obstacle avoidance, face/object recognition, artificial neural networks, signal/image processing, and computer vision. He published more than 150 journal and conference publications.

Munawar Hayat completed his BE in Avionics from National University of Science and Technology (NUST) Pakistan in 2009. He is currently a research student at the University of Western Australia. His research interests include 3D video based facial analysis, pattern recognition and machine learning.

Yulan Guo received his BE in communication engineering from National University of Defense Technology, P.R. China. He is currently working toward his Ph.D. degree at National University of Defense Technology. He is also a visiting (joint) Ph.D. student at the University of Western Australia. His research interests include 3D object recognition, 3D modeling, pattern recognition, remote sensing and signal processing.