Information Sciences 320 (2015) 406–417

Contents lists available at ScienceDirect

Information Sciences

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

Bayesian multi-distribution-based discriminative feature extraction for 3D face recognition

Ronghua Liang *, Wenjia Shen, Xiao-Xin Li, Haixia Wang

College of Information Engineering, Zhejiang University of Technology, Hangzhou, China

ARTICLE INFO

Article history:
Received 18 June 2014
Received in revised form 21 March 2015
Accepted 27 March 2015
Available online 4 April 2015

Keywords:
3D face recognition
Single training sample per person
Depth image
Bayesian learning

ABSTRACT

Due to the difficulties associated with the collection of 3D samples, 3D face recognition technologies often have to work with smaller than desirable sample sizes. With the aim of enlarging the training number for each subject, we divide each training image into several patches. However, this immediately introduces two further problems for 3D models: high computational cost and dispersive features caused by the divided 3D image patches. We therefore first map 3D face images into 2D depth images, which greatly reduces the dimension of the samples. Though the depth images retain most of the robust features of 3D images, such as pose and illumination invariance, they lose many discriminative features of the original 3D samples. In this study, we propose a Bayesian learning framework to extract the discriminative features from the depth images. Specifically, we concentrate the features of the intra-class patches to a mean feature by maximizing the multivariate Gaussian likelihood function, and, simultaneously, enlarge the distances between the inter-class mean features by maximizing the exponential priori distribution of the mean features. For classification, we use the nearest neighbor classifier combined with the Mahalanobis distance to calculate the distance between the features of the test image and items in the training set. Experiments on two widely-used 3D face databases demonstrate the efficiency and accuracy of our proposed method compared to relevant state-of-the-art methods.

Published by Elsevier Inc.

1. Introduction

Face recognition has received considerable attention from the scientific and industrial communities over recent decades. Although 2D face recognition technologies have been extensively studied [46] and are effective under constrained conditions, the variations in illumination, pose and expression still present challenges to existing recognition approaches [1,46]. In this respect, 3D face recognition can be advantageous, since a 3D face model represents the geometry underlying the face image and is thus independent of the ambient illumination and viewpoint [30].

3D face recognition methods can be roughly classified into two categories: spatial matching methods and feature-based matching methods [30]. The former match 3D faces directly by comparing the surface similarity; they include Hausdorff distance [2,17,18,36], iterative closest point (ICP) and its extension [9,32,47]. The main problem of spatial matching methods lies in their high computational cost since 3D face models are usually very large, frequently containing millions of points.

* Corresponding author. Tel./fax: +86 571 85290565.
E-mail address: rhliang@zjut.edu.cn (R. Liang).

http://dx.doi.org/10.1016/j.ins.2015.03.063
0020-0255/Published by Elsevier Inc.

As we are seeking computational efficiency, we shall concentrate on feature-based matching methods which are more efficient as they generally describe the 3D models as multiple 2D features [14–16,19,27,44] – commonly utilized 2D features include raw depth images and surface properties such as gradient and curvature [30].

The surface properties, especially curvature, are widely exploited from different views [5,11,35,39]. However, surface properties are difficult to exploit as the discrete surface property values are calculated from 3D face models, and are thus affected by different levels of noise in the acquisition settings. To enhance the robustness, pre-processing methods such as key point extraction and region segmentation have to be used to deal with the surface properties [29,31], which increases the complexity of the algorithms.

Depth images are usually less subject to noise [30], are usually robust to changes of illumination and viewpoint, and have good computational efficiency [21]. As a result, depth images have been widely used in 3D face recognition, frequently being integrated with various 2D recognition methods [13,20–23,33,34]. Achermann et al. [3] conducted early work on PCA-based depth images, while Lin et al. [26] extracted semi-local summation invariant features from a rectangular region surrounding the nose of a 3D facial depth image. Tsalakanidou et al. [40] and Cook et al. [12] both used depth images combined with intensity information – Tsalakanidou et al. [40] constructed embedded hidden Markov model techniques for face recognition, whereas Cook et al. [12] presented a method based on Log-Gabor templates which were insensitive to variations in expression. Further, Queirolo et al. [33] used a simulated annealing-based approach for depth image registration, with surface interpenetration as a similarity measure, in order to match two face images.

However, a disadvantage of methods based on depth images is that they may lose important 3D face features, such as information on the geometric structure; depth images from different subjects can be similar, which increases the difficulty of recognition. Moreover, for 3D face models, the number of training samples from the same subject is usually very small, as the collection and storage costs of 3D samples are high [38]. In fact, the popular 3D-face databases 3D_RMA [6] and CASIA HFB [25] each contain only a single sample per person (SSPP), which provides insufficient discriminative features from that individual [37]. These issues increase the difficulty of performing accurate recognition.

To address the two problems mentioned above, this paper proposes a novel discriminant analysis method for robust 3D face recognition. To solve the SSPP problem, we divide each depth image into several patches to enlarge the sample size per person. For the problem of similarity of depth image from different subjects, we build a Bayesian multi-distribution model which imposes different distributions on intra-classes and inter-classes, respectively, and hence extract the discriminative features from depth image patches. Experiments on two widely-used 3D face databases have validated the efficiency of the approach proposed.

The main contributions of our proposed method are as follows.

(1) An approach based on a depth image obtained from the 3D face model is adopted for the sake of computational efficiency. By dividing the depth image into local patches, the size of the training sample is enlarged; as a result, we can generate enhanced intra-class information for feature extraction.
(2) Our feature extraction method can identify discriminative features from 3D depth image patches by use of discriminative Bayesian multi-distribution analysis. By this means, we can detect increasingly discriminative inter- and intra-class information related to an individual and thus enhance the recognition rate.

This paper is organized as follows. In Section 2, we describe the motivation of our method, Section 3 introduces the method in detail, and Section 4 presents the experimental results. We summarize our conclusions and indicate our directions for future work in Section 5.

2. Motivation

The SSPP problem mentioned in Section 1 is associated with two main challenges: how to represent the 3D face model with a reduced-dimension method, and how to uncover more discriminative features from the small-sized sample.

2.1. Depth images

Feature-based methods for 3D face recognition typically use depth images to represent the 3D face models. These images are extracted from projections of the 3D models and can provide depth information (i.e., distance from the clipping plane) as well as silhouette information of the 3D models [30]. Being based on 2D, this approach provides greater computational efficiency than 3D spatial matching methods [8]. Furthermore, a depth image is independent of illumination and pose so is advantageous over other 2D representations of 3D faces.

However, 3D face recognition based on depth images has limitations [30] as some 3D face information is lost during the projection that creates the depth image. Fig. 1 presents a 3D model and the corresponding depth image from two different subjects. It can be seen that the depth images of different subjects are similar, which illustrates the difficulties associated with precise recognition from a depth image.

In overcoming these limitations, our method addresses two issues: acquiring a more reliable depth image from the original 3D face model databases is described below; investigating discriminative features between different subjects is discussed in Section 2.2.

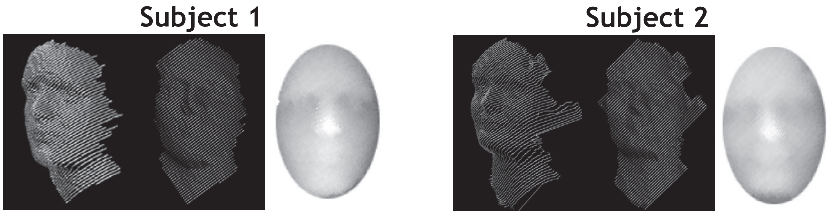


Fig. 1. 3D model (two different views in the left) and its corresponding depth image (in the right) for two subjects from the 3D_RMA database.

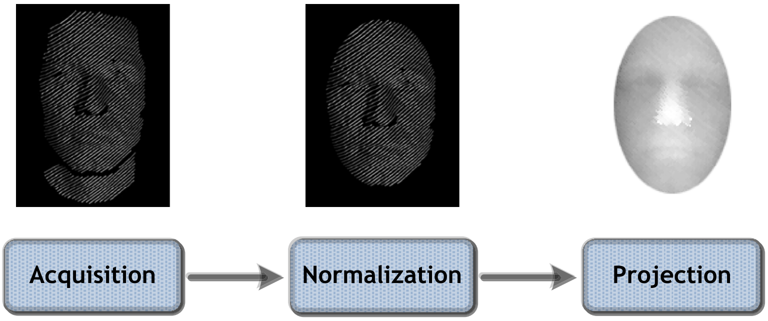


Fig. 2. Depth image produced by three preprocessing steps.

The new depth images are obtained from the three preprocessing steps indicated in Fig. 2. The first step is data acquisition, which is very important because 3D models contain imperfections such as holes and spikes, as well as undesirable data such as clothes, necks, ears, and hair. In this step, the redundant regions are discarded, noise is removed from the 3D data, and holes are filled using linear interpolation. As our feature selection method is pose-sensitive as are other 3D face recognition methods, the second step (normalization) adjusts the pose to use the tip of the nose as a reference point. Finally, the 3D points are projected to produce depth maps on a single reference plane.

2.2. Single sample per person problem

With respect to the SSPP problem, only one sample is available for data training. Several approaches [6,25,38] have tried to overcome this shortcoming for 3D face models; they usually require some human interactivity such as manually setting major feature points [48], which complicates the pre-processing for both data training and recognition. Some studies reduce the complexity by using 2D face recognition [4,28]. Wang and Tang [42] proposed a generic learning framework for discriminative feature extraction based on appearance; this discriminative information can be exploited according to the inter- and intra-class variations. Chen et al. [10] employed LDA by partitioning each facial image into several local patches, and Kanan et al. [24] developed a weighted pseudo-Zernike method to extract the discriminative features of each local patch.

The concepts involved in addressing the SSPP problem in terms of 2D faces can be improved and extended to 3D by the use of depth images. In our method, to acquire additional information for data training, we partition each depth image into non-overlapping patches to construct an image set for each individual sample and then consider the inter- and intra-class discriminations of these patches.

We note that the different semantic patches created from a subject, such as eyes and nose, will differ, thus the various semantic patches of a single subject can be separated based on distance information. To examine the distribution of these local patches, we randomly select a subject from the CASIA HFB database [25] and visualize the distribution structure of its local patches. Fig. 3b depicts a 128 × 128 depth image, which is partitioned into 4 × 4 non-overlapping patches. Fig. 3a shows the histogram distribution of the Euclidean distance between any two patches of the same subject. We observe that some distance values lie near the origin point, but more distance values are dispersed. This finding validates our observation that the patches in a single subject are distant and thereby the images are indistinguishable. However, the intra-class patches from the same subject should be similar for feature extraction and classification. Thus we consider the vectors of all patches from a single subject to be a multivariate Gaussian distribution. And for the concentricity of Gaussian distribution, most of the patches from the same subject will be clustered into intra-class variations. A detailed description of this is provided in Section 3.1.

As discussed in Section 2.1, the depth images of different subjects can be highly similar, which suggests that the patches in the distance images of different subjects located in the same area (hereafter referred to the same semantic patch) will be similar, too. Fig. 4a shows data from the face images of 16 subjects to verify this observation. Each subject is represented by

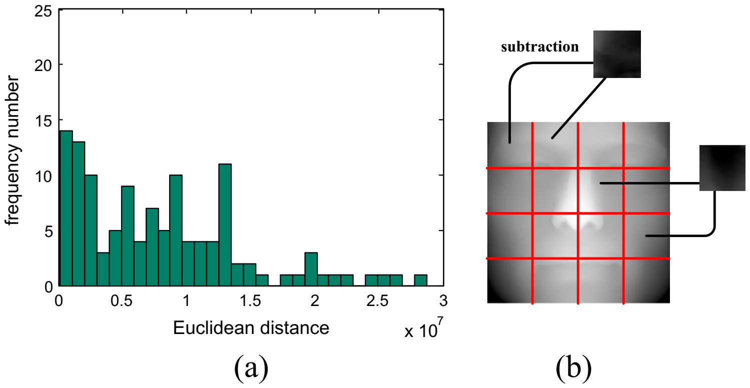


Fig. 3. (a) Frequency number histogram of the Euclidean distance between two patches from the same subject. (b) Estimated Euclidean distance after two patches from the same subject are subtracted.

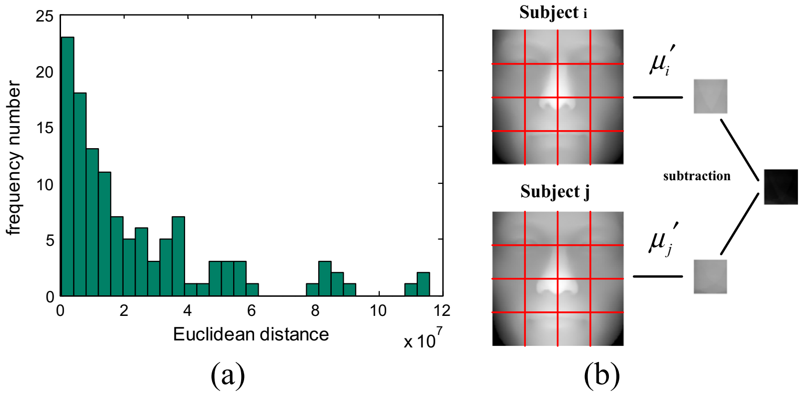


Fig. 4. (a) Frequency number distribution histogram of the Euclidean distance numbers between each pair of mean vectors from different subjects. (b) Estimated Euclidean distance after the two mean vectors of each patch in different subjects are subtracted.

the mean vector of all patches, denoted by $\boldsymbol{\mu}'$. For these subjects, Fig. 4a shows the frequency number distribution of the Euclidean distance between each pair of $\boldsymbol{\mu}'$. Most distance values range from 0 to 4, which is similar to the distance of each pair of patches from the same subject (as shown in Fig. 3a). Hence, those $\boldsymbol{\mu}'$ cannot be used to distinguish between different subjects. Nonetheless, the inter-class patches from different subjects should be distinguished by the low-dimensional discriminative feature vector $\boldsymbol{\mu}$ for feature extraction and classification to address SSPP.

In summary, the same semantic patches from different subjects are usually more similar than different patches from the same subject. This implies that different patches from the same subject are separable in the original image, whereas similar semantic patches from different subjects are clustered. This observation runs counter to our objective which is that, after depth image mapping, the local patches from different subjects should be effectively separated while those from the same subjects are close.

Hence, we ensure that the subjects are represented by discriminative information that can be exploited for recognition. To achieve this objective, we propose feature extraction based on the Bayesian theorem. We estimate the discriminative $\boldsymbol{\mu}$, which is expressed as: $p(\boldsymbol{\mu}|\mathbf{x}) \propto p(\mathbf{x}|\boldsymbol{\mu})p(\boldsymbol{\mu})$, where $p(\mathbf{x}|\boldsymbol{\mu})$ is the likelihood function that controls intra-class information and $p(\boldsymbol{\mu})$ is the priori distribution that manages inter-class information.

3. Proposed approach

3.1. Overview of the proposed approach

Fig. 5 shows the stages of the method proposed. The first stage is data preprocessing, as described in Section 2.1. The redundant regions are discarded; noise is removed from the 3D data, and holes are filled using linear interpolation. The 3D points are normalized and projected into the depth images, which are then ready for feature extraction and recognition. In the second stage, each depth image is divided into non-overlapping patches as samples of each person for training. The third stage uses feature extraction based on Bayesian theorem with a multivariate Gaussian likelihood function and

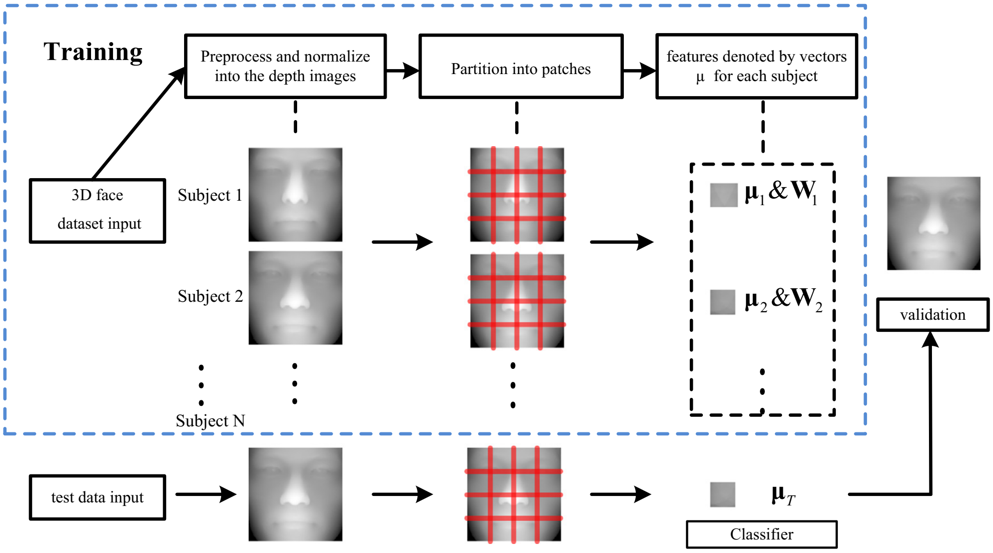


Fig. 5. Overview of the proposed method.

exponential priori distribution. Each sample can be represented by a set of lower dimensional vectors $\boldsymbol{\mu}$. Finally, we use the feature vectors for classification and recognition, with the Mahalanobis distance [43,45] between feature vectors of two images being employed as the similarity metric.

3.2. Feature extraction

The n training depth images with size $a \times b$ in the recognition database are denoted as $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n]$. Each image is divided into $N \times M$ non-overlapping patches. Thus, we obtain an image set $\mathbf{X}_i = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{it}](t = N \times M)$, where $\mathbf{x}_{ir}$ represents patch r of person i and $\mathbf{x}_{ir} \in R^m$, and $m = (a \times b)/(N \times M)$.

The aim of our method is to find n feature matrices $\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_n, \mathbf{W}_i \in R^{m \times m_i}, i = 1, 2, \ldots, n$, where m and m_i denote the feature dimension of each original image patch and the low dimensional features learned by $\mathbf{W}_i$, respectively. After that, we can obtain the mean vectors of the training data $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_n$, where $\boldsymbol{\mu}_i = \frac{1}{t}\sum_{r=1}^{t}\mathbf{W}_i^T\mathbf{x}_{ir}(i = 1, 2, \ldots, n)$.

As mentioned in Section 2.2, we need to ensure that different semantic patches of the same subject are separable, while intra-class patches in the same subject are clustered. We consider the vectors of all patches from the same subject as a multivariate Gaussian distribution. Thus, our approach can cluster most patches of the same subject together. Given a vector $\boldsymbol{\mu}_i \in R^{m_i}$, the associated target $\mathbf{W}_i^T\mathbf{x}_{ir}$ has a multivariate Gaussian distribution with mean $\boldsymbol{\mu}_i$ and variance Σ_i as

$$p\left(\mathbf{W}_i^T\mathbf{x}_{ir}\middle|\boldsymbol{\mu}_i, \sum_i\right) = \frac{1}{(2\pi)^{m_i/2}}\frac{1}{|\sum_i|^{1/2}}\exp\left\{-\frac{1}{2}\left(\mathbf{W}_i^T\mathbf{x}_{ir} - \boldsymbol{\mu}_i\right)^T\sum_i^{-1}\left(\mathbf{W}_i^T\mathbf{x}_{ir} - \boldsymbol{\mu}_i\right)\right\} \tag{1}$$

where $\boldsymbol{\mu}_i$ is an m_i-dimensional mean vector, Σ_i is an $m_i \times m_i$ covariance matrix, and $|\Sigma_i|$ denotes the determinant of $\Sigma_i, \sum = \left[\sum_1, \sum_2, \ldots, \sum_n\right]^T \in R^n$. As shown in Eq. (1), Σ_i represents the degree of the polymerization from the different semantic patches in the same subject. As we decrease the value of $|\sum_i|$, the patches from one subject become closer after calculation.

Assume that the data $\left\{\mathbf{W}_i^T\mathbf{x}_i, \boldsymbol{\mu}_i\right\}$ is independent of the distribution in Eq. (1), the likelihood function can be represented as

$$p\left(\mathbf{W}_i^T\mathbf{x}_i\middle|\boldsymbol{\mu}_i, \sum_i\right) = \prod_{r=1}^{t} p\left(\mathbf{W}_i^T\mathbf{x}_{ir}\middle|\boldsymbol{\mu}_i, \sum_i\right) \tag{2}$$

As discussed in Section 2.2, the patches of different subjects can be easily clustered. Meanwhile, from the viewpoint of classification, we aim to maximize the separability among different persons in the low-dimensional feature spaces. We intend to find the low-dimensional feature vector $\boldsymbol{\mu}$ to represent each subject, which is also discriminative from each other. We consider the inter-class priori on $\boldsymbol{\mu}$ to control the inter-class discrimination, where each two mean $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j$ are far apart. For this purpose, we assume that $\boldsymbol{\mu}_i$ is generated according to the following exponential priori [7]:

$$p\left(\boldsymbol{\mu}_i\middle|\boldsymbol{\mu}_j,\sum_i^{\prime}\right)\propto\exp\left(-\left\|\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\right\|\sum_i^{\prime-1}\right)\quad(i=1,\ldots,n,\ j\neq i)\tag{3}$$

where $\sum' = \left[\sum_1', \sum_2', \ldots, \sum_n'\right]^T \in R^n$. $\sum_i'$ represents the degree of polymerization between person i and the other subjects. As we increase the value of $\left|\sum_i'\right|$, the different subjects become more separable. Assume that $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_n$ are independent of the priori in Eq. (3), then the priori for $\boldsymbol{\mu}_i$ can be expressed as

$$p\left(\boldsymbol{\mu}_i\middle|\sum_i^{\prime}\right)=p\left(\boldsymbol{\mu}_i\middle|\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_j,\ldots\boldsymbol{\mu}_n,j\neq i,\sum_i^{\prime}\right)=\prod_{\substack{j=1\\j\neq i}}^{n}p\left(\boldsymbol{\mu}_i\middle|\boldsymbol{\mu}_j,\sum_i^{\prime}\right)\tag{4}$$

Subsequently, the posterior distribution for $\boldsymbol{\mu}_i$, which is proportional to the product of the priori and the likelihood function [7], is given by

$$p\left(\boldsymbol{\mu}_i\middle|\mathbf{W}_i^T\mathbf{x}_i,\sum_i,\sum_i^{\prime}\right)\propto p\left(\mathbf{W}_i^T\mathbf{x}_i\middle|\boldsymbol{\mu}_i,\sum_i\right)p\left(\boldsymbol{\mu}_i\middle|\sum_i^{\prime}\right)\tag{5}$$

To combine the logarithm of Eq. (5) with Eqs. (1)–(4), we obtain the maximum posterior estimation of $\boldsymbol{\mu}_i$ by minimizing

$$\sum_{r=1}^{t}\left(\frac{1}{2}\left(\mathbf{W}_i^T\mathbf{x}_{ir}-\boldsymbol{\mu}_i\right)^T\sum_i^{-1}\left(\mathbf{W}_i^T\mathbf{x}_{ir}-\boldsymbol{\mu}_i\right)\right)+\sum_{\substack{j=1\\j\neq i}}^{n}\left\|\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\right\|\sum_i^{\prime-1}\tag{6}$$

Thus, our objective function can be written as

$$\min_{\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_n} J = J_1 + J_2\tag{7}$$

where

$$J_1=\sum_{r=1}^{t}\frac{1}{2}\left(\mathbf{W}_i^T\mathbf{x}_{ir}-\boldsymbol{\mu}_i\right)^T\sum_i^{-1}\left(\mathbf{W}_i^T\mathbf{x}_{ir}-\boldsymbol{\mu}_i\right)$$

and

$$J_2=\sum_{\substack{j=1\\j\neq i}}^{n}\left\|\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\right\|\sum_i^{\prime-1}.$$

J_1 in Eq. (7) is obtained from multivariate Gaussian distribution in Eq. (1) to ensure the closeness of their low-dimensional representations if $\mathbf{x}_{ir}$ (r = 1,. . .,t; i = 1,. . .,n) are obtained from the same subject. In contrast, J_2 in Eq. (7) is from the exponential priori Eq. (3) to ensure the separability of their low-dimensional representations if $\boldsymbol{\mu}_i(i=1,\ldots,n)$ comes from different subjects.

To the best of our knowledge, there is no closed-form solution for the optimization problem defined in Eq. (7) as there are n projection matrices and n vectors to be achieved simultaneously. We solve this problem in an iterative manner to obtain the minimal value from the derivative of Eq. (7). The basic idea is to set $\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_n$ and $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_n$ with valid initial values and then solve for $\mathbf{W}_i$ and $\boldsymbol{\mu}_i$ sequentially.

To estimate $\boldsymbol{\mu}_i$, we consider the derivative of Eq. (7) with respect to $\boldsymbol{\mu}_i$.

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\mu}_i}J_1&=\sum_{r=1}^{t}\frac{\partial}{\partial\boldsymbol{\mu}_i}\left(\frac{1}{2}\left(\mathbf{W}_i^T\mathbf{x}_{ir}-\boldsymbol{\mu}_i\right)^T\sum_i^{-1}\left(\mathbf{W}_i^T\mathbf{x}_{ir}-\boldsymbol{\mu}_i\right)\right)=\frac{1}{2}\sum_{r=1}^{t}\left(-\left(\mathbf{x}_{ir}^T\mathbf{W}_i\sum_i^{-1}\right)^T-\sum_i^{-1}\mathbf{W}_i^T\mathbf{x}_{ir}+\sum_i^{-1}\boldsymbol{\mu}_i+\left(\boldsymbol{\mu}_i^T\sum_i^{-1}\right)^T\right)\\&=\sum_{r=1}^{t}\sum_i^{-1}\left(\boldsymbol{\mu}_i-\mathbf{W}_i^T\mathbf{x}_{ir}\right),\end{aligned}\tag{8}$$

and

$$\frac{\partial}{\partial\boldsymbol{\mu}_i}J_2=\frac{\partial}{\partial\boldsymbol{\mu}}\sum_{\substack{j=1\\j\neq i}}^{n}\left\|\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\right\|\sum_i^{\prime-1}=\sum_{\substack{j=1\\j\neq i}}^{n}\frac{\partial}{\partial\boldsymbol{\mu}}\left(\boldsymbol{\mu}_i^T\sum_i^{\prime-1}\boldsymbol{\mu}_i-\boldsymbol{\mu}_i^T\sum_i^{\prime-1}\boldsymbol{\mu}_j-\boldsymbol{\mu}_j^T\sum_i^{\prime-1}\boldsymbol{\mu}_i+\boldsymbol{\mu}_j^T\sum_i^{\prime-1}\boldsymbol{\mu}_j\right)=2\sum_{\substack{j=1\\j\neq i}}^{n}\sum_i^{\prime-1}\left(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\right).\tag{9}$$

The derivative is of J is found by combining Eqs. (8) and (9):

$$\frac{\partial}{\partial\boldsymbol{\mu}_i}J=\frac{\partial}{\partial\boldsymbol{\mu}_i}(J_1+J_2)=\sum_{r=1}^{t}\sum_i^{-1}\left(\boldsymbol{\mu}_i-\mathbf{W}_i^T\mathbf{x}_{ir}\right)+2\sum_{\substack{j=1\\j\neq i}}^{n}\sum_i^{\prime-1}\left(\boldsymbol{\mu}_i-\boldsymbol{\mu}_j\right).$$

If we set this derivative to zero, we obtain the solution for the maximum estimation of the mean given by

$$\left(2(n-1)\sum_{i}^{\prime-1}+t\sum_{i}^{-1}\right)\boldsymbol{\mu}_i=2\sum_{\substack{j=1\\j\neq i}}^{n}\sum_{i}^{\prime-1}\boldsymbol{\mu}_j+\sum_{r=1}^{t}\sum_{i}^{-1}\mathbf{W}_i^T\mathbf{x}_{ir}.$$

Let

$$A_i=\left(2(n-1)\sum_{i}^{\prime-1}+t\sum_{i}^{-1}\right)^{-1},\ \text{where}\ \sum_{i}^{\prime-1}\neq-\frac{t\sum_i^{-1}}{2(n-1)},\tag{10}$$

we then obtain:

$$\boldsymbol{\mu}_i=2A_i\sum_{i}^{\prime-1}\sum_{\substack{j=1\\j\neq i}}^{n}\boldsymbol{\mu}_j+A_i\sum_{i}^{-1}\sum_{r=1}^{t}\mathbf{W}_i^T\mathbf{x}_{ir}.\tag{11}$$

To estimate $\mathbf{W}_i$, the derivative of Eq. (7) with respect to $\mathbf{W}_i$ is given by

$$\frac{\partial}{\partial\mathbf{W}_i}J=\frac{\partial}{\partial\mathbf{W}_i}(J_1+J_2)=\frac{\partial}{\partial\mathbf{W}_i}J_1=\frac{\partial}{\partial\mathbf{W}_i}\left(\sum_{r=1}^{t}\frac{1}{2}\left(\mathbf{W}_i^T\mathbf{x}_{ir}-\boldsymbol{\mu}_i\right)^T\sum_{i}^{-1}\left(\mathbf{W}_i^T\mathbf{x}_{ir}-\boldsymbol{\mu}_i\right)\right)=\sum_{r=1}^{t}\mathbf{x}_{ir}\left(\mathbf{x}_{ir}^T\mathbf{W}_i-\boldsymbol{\mu}_i^T\right)\sum_{i}^{-1}.$$

Set this derivative to zero, and we obtain the solution for the maximum estimation of $\mathbf{W}_i$ given by

$$\left(\sum_{r=1}^{t}\mathbf{x}_{ir}^T\right)\mathbf{W}_i=t\boldsymbol{\mu}_i^T.\tag{12}$$

We thus iteratively solve the optimization problem in Eq. (7) by using Eqs. (11) and (12) to determine the n projection matrices $\mathbf{W}_1,\mathbf{W}_2,\ldots,\mathbf{W}_n$, and n means $\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\ldots,\boldsymbol{\mu}_n$. Our algorithm is summarized as follows.

Algorithm 1.

```
input: a training set of n people $\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n$; the covariance parameters $\Sigma,\Sigma'$.
initialization:
    Set $\mathbf{W}_i^0=I_{m\times m_i}(i=1,2,\ldots,n),\boldsymbol{\mu}_i^0=\frac{1}{t}\sum_{r=1}^{t}\left(\mathbf{W}_i^0\right)^T\mathbf{x}_{ir}\quad(i=1,2,\ldots,n)$.
parameters calculation:
    For each subject, calculate $A_i$, defined in Eq. (10).
optimization:
    Repeatedly:
    Compute $\boldsymbol{\mu}_i^q$ and $\mathbf{W}_i^q$ according to Eqs. (11) and (12), respectively.
    Until maximum iterations or convergence
output:
    Projection matrices $\mathbf{W}_i=\mathbf{W}_i^q$ and mean $\boldsymbol{\mu}_i=\boldsymbol{\mu}_i^q, i=1,\ldots,n$.
```

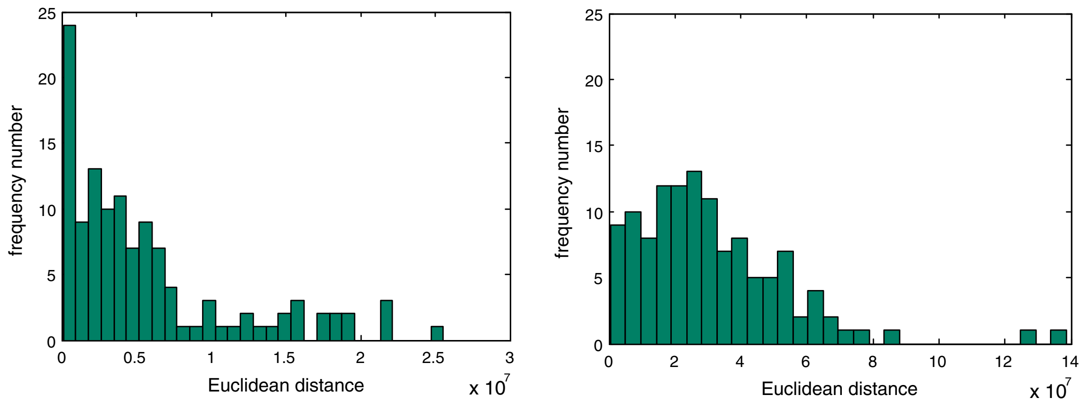
After $\mathbf{W}_i$ and n mean vectors $\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\ldots,\boldsymbol{\mu}_n$ are obtained, we can exploit them to represent the local patches for each subjects. Fig. 6a illustrates the low-dimensional features of the 16 different patches from one subject in Fig. 3a learned by our method. We see that most of the distance values between patches from the same subject lie near the origin, which shows that intra-class patches of the same subject are clustered by our method. Fig. 6b presents the low-dimensional features of the 16 different subjects in Fig. 4a learned by our method. We observe that most of the mean distance values between pairs of subjects are dispersed so our method separates inter-class patches. Therefore, in our approach of inter-class and intra-class variations, the discriminative information can be exploited for recognition.

3.3. Recognition

In Section 3.2, we produced the n projection matrices $\mathbf{W}_1,\mathbf{W}_2,\ldots,\mathbf{W}_n$ and n mean vectors $\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\ldots,\boldsymbol{\mu}_n$ by learning the training data with our Algorithm 1. Given a test face sample T, the division of t non-overlapping local patches can be modeled as $\mathbf{X}_T=[\mathbf{x}_{T1},\mathbf{x}_{T2},\ldots,\mathbf{x}_{Tt}]$, and we employ c to represent it as follows:

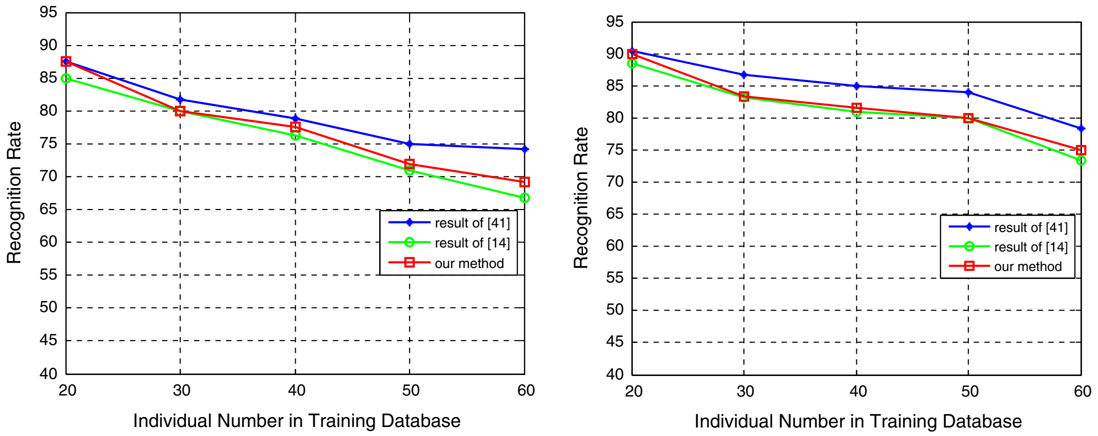
$$\text{identity}(c)=\arg\min_i\ d(\mathbf{X}_i,\mathbf{X}_T),\tag{13}$$

where $d(\mathbf{X}_i,\mathbf{X}_T)$ is the distance between the i^{th} subject and the test sample T.



(a) (b)

Fig. 6. (a) Frequency number histogram of the Euclidean distance for each pair of patches from one subject learned by our method compared with Fig. 3a. (b) Frequency number histogram of the mean Euclidean distance between each of two different subjects compared with Fig. 4a.



(a) Test Configuration T_1 (b) Test Configuration T_2

Fig. 7. Face recognition rate using different methods on the 3D_RMA database.

As described in Section 3.2, we gain low-dimensional features learned by $\mathbf{W}_i$. In other words, our training image $\mathbf{X}_i = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{it}]$ can be represented by a low-dimensional vector $\boldsymbol{\mu}_i$, which is the average of each patch $\mathbf{W}_i^T\mathbf{x}_{ir}$. Similarly, a low-dimensional vector $\boldsymbol{\mu}_T = \frac{1}{t}\sum_{r=1}^{t}\mathbf{W}_i^T\mathbf{x}_{Tr}(i = 1, 2, \ldots, n)$ for different subjects i can be generated for the test sample $\mathbf{X}_T$. The distance between subject i and the test sample T can be reformulated as $d(\boldsymbol{\mu}_i, \boldsymbol{\mu}_T)$ for recognition.

Many methods are available to calculate the distance between two vectors. The Mahalanobis distance is often chosen for its advantages [43,45] in the identification of similarity between an unknown sample and a known sample. Unlike Euclidean distance, the Mahalanobis distance considers correlations of the data set and is thereby scale-invariant. Thus, the Mahalanobis distance has a multivariate size. We define the distance between the subject i and the test sample T as

$$d(\mathbf{X}_i, \mathbf{X}_T)^2 \triangleq d(\boldsymbol{\mu}_i, \boldsymbol{\mu}_T)^2 = (\boldsymbol{\mu}_i - \boldsymbol{\mu}_T)^T \sum^{t-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_T)$$

for use in Eq. (13).

4. Experiments

Our approach was tested on two widely-used 3D face databases in which discriminative feature vectors are generated for further face recognition. 3D_RMA [6] originates from Belgium and consists of 360 face models, each of 3000 points, representing 120 subjects. Each person is associated with three models with different orientations of the head: forward, left or right, and upward or downward, respectively. CASIA HFB [25], collected by the Center for Biometrics and Security

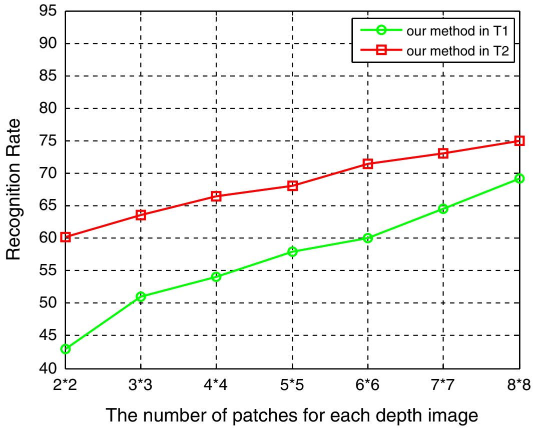


Fig. 8. Recognition rate with different numbers of patches in the 3D_RMA database.

Research, consists of visual (VIS), near infrared (NIR) and three-dimensional (3D) face models. The 3D face models from 92 persons in the CASIA HFB Database were tested – each person was represented by two models.

As mentioned in Section 2, in the pre-processing stage, redundant face regions for the 3D face models from both databases were discarded, noise removed, and holes filled using linear interpolation. Each 3D model was then normalized and sampled to construct one depth image. This depth image was resized to 128 × 128 pixels, normalized according to the position of the nose.

We first used the 3D_RMA database to compare our method with two typical 3D face recognition methods [14,41]. Fang et al. [14] presented a Markov Random Field model for the analysis of lattices in terms of the discriminative information of their vertices, and Wang et al. [41] introduced a method to learn the most discriminative local areas from the Signed Shape Difference Map (SSDM). In our tests, the size of each patch in the depth image was set to 16 × 16 and the parameters Σ, Σ' and m_i were empirically set to $0.5\mathbf{I}$, $2\mathbf{I}$ and 50, respectively, where $\mathbf{I}$ is an $m_i \times m_i$ identity matrix.

We constructed two test configurations to demonstrate the performance of our method on the SSPP problem. In configuration T_1, for each person we take one face model for training and the other two models for validation, while in configuration T_2, for each person we take two face models for training and use the other one for validation. In each configuration, we created five test subsets with 20, 30, 40, 50, 60 subjects from the database, respectively. In each subset, we randomly performed 20 experiments in the database to calculate the average recognition rate. Fig. 7 shows the recognition rates for our method and the methods from [14,41] on the five test subsets.

As shown in Fig. 7, our proposed method has a recognition rate that is higher than the method in [14] but slightly lower than that in [41]. Since our proposed method is implemented on 2D depth images, it is much faster than that in [14,41]; on a 2.6-GHz CPU, 3-GB RAM PC, our method was 5 times faster.

The method in [41] outperforms our method by using more 3D shape information. It selects the most discriminative local features from a Signed Shape Difference Map which is computed between two aligned 3D faces as a mediate representation for the shape comparison. The 3D shape information and three collective strong classifiers partly compensate for small sized training samples. However, there is a major drawback in that it requires much complex pre-processing and is thus inconvenient to use. In contrast, our method adopts high-efficiency preprocessing steps, requires only one classifier, involves less computational time, and is convenient to use, while achieving comparable results.

Comparing Fig. 7a and b, we note that the recognition rate in Fig. 7a using the T_1 configuration is slightly lower than, but comparable with, the recognition rate in Fig. 7b using T_2 configuration. This demonstrates the ability of our proposed method to solve the SSPP problem. Further, our proposed method provides a greater advantage in the SSPP case than in non-SSPP cases. On average, in T_1 our method outperforms [14] by up to 1.44%, but underperforms [41] by up to 2.2%. In T_2, our method outperforms [14] by up to 0.76%, while underperforming [41] by up to 2.92%. Therefore, our method is suitable for application to situations where there is difficulty in 3D sample collection.

Further tests investigated the influence of patch number in a depth image over the recognition rate, using configurations T_1 and T_2 on a subset of 60 randomly chosen subjects from the 3D_RMA database. The recognition rate is taken as the average of 20 test runs on this subset. As can be seen in Fig. 8, the recognition rate increases when the number of patches in each image increases – more partitions produce more training samples, and these improve the recognition rate on the SSPP problem in 3D face recognition. T_2 has more training data than T_1 so, not unexpectedly, the result in T_2 outperforms that in T_1 averagely by up to nearly 12%. We also observe that the number of partitions has less influence on the recognition rate in T_2 – the difference in recognition rate between T_1 and T_2 decreases as the number of patches in each image increases. Overall, our method has advantages in the SSPP problem when there are a greater number of patches.

The 3D_RMA database was also used to investigate the relationship between recognition rate and feature dimensions of the depth image after feature extraction. A subset of 60 randomly chosen subjects in database was used, and the final

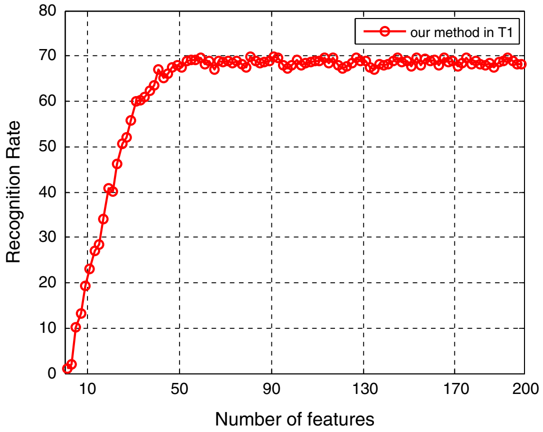


Fig. 9. Recognition rate versus the number of features in the 3D_RMA database.

recognition rate was computed by averaging the results from 5 tests. Fig. 9 shows that the recognition rate rapidly increases with the feature dimension at the beginning but steadies to a plateau for features above 50. The recognition rate in configurations T_1 was 69.2% with 50 features and 68.3% with 200 features. Therefore our proposed method can reduce the feature dimensions of the data effectively and extract the discriminative features to improve the recognition performance.

Further validation took place using the CASIA HFB Database. For each subject, we used one face model for training and the other face model for testing. Our method is compared with five typical learning-based methods used in face recognition problems, including PCA [34], 2DPCA (2D PCA) [13], Block PCA [23], 2D LDA [13], and LPP [22]. These methods set their parameters for the best performance to ensure a fair comparison. For PCA, the contribution parameter α was set to 0.95; for 2DPCA and 2DLDA, the dimension of the selected feature was set to 9 and 60, respectively; for LPP, the number of nearest neighbors was selected as 2 and t was set to 20 to calculate the similarity matrix. For our method, the parameters were the same as in the tests with the 3D_RMA database. For all block-based methods, such as Block PCA and our proposed method, the size of each block was set to 16×16. The tests covered five subsets with 20, 30, 50, 70 and 90 subjects from the database, respectively.

Table 1 shows the best recognition rate of PCA, 2DPCA (2D PCA), Block PCA, 2D LDA, LPP and our proposed method on the CASIA HFB Database. Our method highlighted with bold generally outperforms the other five methods on the CASIA HFB Database, achieving gains in accuracy of 2%, 5.6% and 3.3% compared with the best results of other methods in testing the subsets of 50, 70 and 90 subjects.

We make two observations from the results listed in Table 1.

1. For all methods, the recognition rates decrease as the number of subjects in the training database increases. An increase in the size of the training database increases the number of extracted features, which interferes with the recognition process. However, the rate of decrease of our method is the lowest among the methods. The most discriminative features extracted by the proposed method have eased the interference caused by the increase in training samples, so the proposed method is more practical than other methods.
2. Our method produces the best recognition performance across all of the experiments, which implies that both discriminative and geometrical information of local patches are important for recognition.

Since Block PCA generally performs better than the other four methods in Table 1, it was also used, alongside our method, to test the influence of partition numbers on the CASIA HFB Database. We randomly choose 30 subjects for training and testing and again took one 3D model per person for training and the other for testing. The results in Table 2 shows that

Table 1
Recognition rate (%) of different methods on the CASIA HFB database.

Method	Number of individuals in training database				
	20	30	50	70	90
PCA	65.0	53.3	40.0	30.0	30.0
2DPCA	65.0	56.7	52.0	44.3	40.0
Block PCA	65.0	60.0	54.0	48.6	40.0
2D LDA	60.0	56.7	50.0	44.3	38.9
LPP	65.0	66.7	52.0	42.9	41.1
Our method	**65.0**	**66.7**	**56.0**	**54.2**	**44.4**

Table 2
Recognition rate (%) of different methods with different number of patches on the CASIA HFB database.

Method	Number of patches in each image					
	2 × 2	3 × 3	4 × 4	6 × 6	7 × 7	8 × 8
Block PCA	43.2	43.5	47.6	53.3	53.8	60.0
Our method	46.7	50.0	63.2	65.3	66.7	66.9

our method has a higher recognition rate than Block PCA; this is because our proposed method can produce larger discrimination of the inter-class and intra-class information to extract features than Block PCA can. Table 2 also shows that the recognition rate increases as the number of patches in each depth image increases. The highest recognition rate obtained was 66.9% for our method using a partition size of 8 × 8. Based on Bayesian multi-distribution analysis of the patches in depth image, the results demonstrate that our method can efficiently deal with the SSPP problem in 3D face recognition.

5. Conclusion and future work

This paper proposes a multi-distribution-based discriminative feature extraction method for use in 3D face recognition which uses depth images to construct a classifier to improve the computational efficiency. To overcome the limitation of small sample sizes often found in 3D face model recognition, the depth image is divided into patches to produce more features, and a Bayesian learning framework is used to extract the discriminative features from the depth images. This method can obtain larger discriminative features than previous approaches by using inter-class and intra-class information. Experimental results on two widely used face databases have demonstrated its efficiency and effectiveness.

In the future, we will cover more 3D face databases and extend our method to deal with expressions, missing data, and problems associated with occlusion. Another interesting direction for possible future work is to combine local patch information and global holistic information of the depth images to improve the SSPP 3D face recognition rate.

Acknowledgment

This work is partially supported by National Science Foundation of China (61402411), Zhejiang Provincial Natural Science Foundation (LY14F020014 and LY14F020015), and Program for New Century Excellent Talents in University of China (NCET-12-1087).

References

[1] Face Recognition Vendor Test 2002. <http://www.frvt.org/>.
[2] B. Achermann, H. Bunke, Classifying range images of human faces with Hausdorff distance, in: Proc. 15th International Conference on Pattern Recognition, 2000, pp. 809–813.
[3] B. Achermann, X. Jiang, H. Bunke, Face recognition using range images, in: Proc. 1st International Conference on Virtual Systems and Multimedia, 1997, pp. 129–136.
[4] T. Ahonen, A. Hadid, M. Pietikainen, Face Recognition with Local Binary Patterns, in: Proc. 8th European Conference on Computer Vision, 2004, pp. 469–481.
[5] S. Berretti, A. Del Bimbo, P. Pala, 3D face recognition using Isogeodesic Stripes, IEEE Trans. Pattern Recognit. Mach. Intell. 32 (12) (2010) 2162–2177.
[6] C. Beumier, M. Acheroy, Automatic 3D face authentication, Image Vision Comput. 18 (4) (2000) 315–321.
[7] C.M. Bishop, in: Pattern Recognition and Machine Learning, Springer, 2006, pp. 78–113.
[8] B. Bustos, D.A. Keim, D. Saupe, T. Schreck, D.V. Vranic, Feature-based similarity search in 3D object databases, ACM Comput. Surv. 37 (4) (2005) 345–387.
[9] K.I. Chang, K.W. Bowyer, RJ. Flynn, Multiple nose region matching for 3D face recognition under varying facial expression, IEEE Trans. Pattern Recognit. Mach. Intell. 28 (10) (2006) 1695–1700.
[10] S. Chen, J. Liu, Z. Zhou, Making FLDA applicable to face recognition with one sample per person, Pattern Recognit. 37 (7) (2004) 1553–1555.
[11] A. Colombo, C. Cusano, R. Schettini, 3D face detection using curvature analysis, Pattern Recognit. 39 (3) (2006) 444–455.
[12] J. Cook, C. Mccool, V. Chandran, S. Sridharan, Combined 2D/3D face recognition using log-Gabor templates, in: Proc. IEEE Conference on Video and Signal Based Surveillance, 2006, pp. 83–90.
[13] A. Eftekhari, M. Forouzanfar, H.A. Moghaddam, J. Alirezaiee, Block-wise 2D kernel PCA/LDA for face recognition, Inf. Process. Lett. 110 (17) (2010) 761–766.
[14] T.H. Fang, S.K. Shah, I.A. Kakadiaris, 3D face discriminant analysis using gauss-markov posterior marginals, IEEE Trans. Pattern Anal. Mach. Intell. 35 (3) (2013) 728–739.
[15] Y. Gao, J. Tang, H. Li, Q. Dai, N. Zhang, View-based 3D modelretrieval with probabilistic graph model, Neurocomputing 73 (10–12) (2010) 1900–1905.
[16] Y. Gao, M. Wang, D. Tao, R. Ji, Q. Dai, 3D object retrieval and recognition with hypergraph analysis, IEEE Trans. Image Process. 21 (9) (2012) 4290–4303.
[17] Y. Gao, M. Wang, X. Wu, Q. Dai, 3-D object retrieval with hausdorff distance learning, IEEE Trans. Ind. Electron. 61 (4) (2014) 2088–2098.
[18] A. Gardner, J. Kanno, C.A. Duncan, R.Selmic, Measuring distance between unordered sets of different sizes, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 23–28.
[19] Y. Guo, F. Sohel, M. Bennamoun, J. Wan, M. Lu, A novel local surface feature for 3D object recognition under clutter and occlusion, Inf. Sci. 293 (2015) 196–213.
[20] S. Gupta, J.K. Aggarwal, M.K. Markey, A.C. Bovik, 3D face recognition founded on the structural diversity of human face, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–7.
[21] C. Hesher, A. Srivastava, G. Erlebacher, A novel technique for face recognition using range imaging, in: Proc. International Symposium on Signal Processing and Its Applications, 2003, pp. 201–204.
[22] X. He, P. Niyogi, Locality preserving projections, in: Advances in Neural Information Processing Systems, MIT Press, Cambridge, MA, USA, 2004.

[23] Q.C. Jiang, X.F. Yan, Monitoring multi-mode plant-wide processes by using mutual information-based multi-block PCA, joint probability, and Bayesian inference, Chemom. Intell. Lab. Syst. 136 (2014) 121–137.

[24] H. Kanan, K. Faez, Y. Gao, Face recognition using adaptively weighted patch PZM array from a single exemplar image per person, Pattern Recognit. 41 (12) (2008) 3799–3812.

[25] S.Z. Li, L. Zhen, The HFB face database for heterogeneous face biometrics research, in: Proc. 6th IEEE Workshop on Object Tracking and Classification Beyond and in the Visible Spectrum Miami, Florida, 2009.

[26] W. Lin, K. Wong, N. Boston, Y. Hu, Fusion of summation invariants in 3D human face recognition, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2006, pp. 1369–1376.

[27] K. Lu, Q. Wang, J. Xue, W. Pan, 3D model retrieval and classification by semi-supervised learning with content-based similarity, Inf. Sci. 281 (2014) 703–713.

[28] J. Lu, Y.P. Tan, G. Wang, Discriminative multimanifold analysis for face recognition from a single training sample per person, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 39–51.

[29] Y. Lee, J. Shim, Curvature-based human face recognition using depth-weighted Hausdorff distance, in: Proc. International Conference on Image Processing, 2004, pp. 1429–1432.

[30] A. Mian, N. Pears, in: 3D Imaging: Analysis and Applications, Springer, London, 2012, pp. 311–366.

[31] A.B. Moreno, A. Sanchez, J.F. Velez, F.J. Diaz, Face recognition using 3D surface-extracted descriptors, in: Proc. the Irish Machine Vision and Image Processing Conference, 2003.

[32] H. Mohammadzade, D. Hatzinakos, Iterative closest normal point for 3D face recognition, IEEE Trans. Pattern Recognit. Mach. Intell. 35 (2) (2013) 381–397.

[33] C.C. Queirolo, L. Silva, O.R.P. Bellon, M. R Segundo, 3D face recognition using simulated annealing and the surface interpenetration measure, IEEE Trans. Pattern Recognit. Mach. Intell. 32 (2) (2010) 206–219.

[34] T. Russ, C. Boehnen, T. Peters, 3D face recognition using 3D alignment for PCA, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2006, pp. 1391–1398.

[35] C. Samir, A. Srivastava, M. Daoudi, Three-dimensional face recognition using shapes of facial curves, IEEE Trans. Pattern Recognit. Mach. Intell. 28 (11) (2006) 1858–1863.

[36] E. Sangineto, Pose and expression independent facial landmark localization using dense-SURF and the hausdorff distance, IEEE Trans. Pattern Recognit. Mach. Intell. 35 (3) (2013) 624–638.

[37] G. Shakhnarovich, B. Moghaddam, Handbook of Face Recognition-Face Recognition in Subspaces, Springer, London, 2011. pp. 19–49.

[38] X. Tan, S. Chen, Z.H. Zhou, F. Zhang, Face recognition from a single image per person: a survey, Pattern Recogn. 39 (9) (2006) 1725–1745.

[39] H.T. Tanaka, M. Ikeda, H. Chiaki, Curvature-based face surface recognition using spherical correlation-principal directions for curved object recognition, in: Proc. International Conference on Automatic Face and Gesture Recognition, 1998, pp. 372–377.

[40] F. Tsalakanidou, S. Malassiotis, M.G. Strintzis, Face localization and authentication using color and depth images, IEEE Trans. Image Process. 14 (2) (2005) 152–168.

[41] Y.M. Wang, J.Z. Liu, X. Tang, Robust 3D face recognition by local shape difference boosting, IEEE Trans. Pattern Recognit. Mach. Intell. 32 (10) (2010) 1858–1870.

[42] X. Wang, X. Tang, Random sampling for subspace face recognition, Int. J. Comput. Vision 70 (1) (2006) 91–104.

[43] K.Q. Weinberger, L.K. Saul, Distance metric learning for large margin nearest neighbor classification, J. Mach. Learn. Res. Arch. 10 (2009) 207–244.

[44] Y. Xia, L. Zhang, W. Xu, Z. Shan, Y. Liu, Recognizing multi-view objects with occlusions using a deep architecture, Inf. Sci. (2015).

[45] S. Xiang, F. Nie, C. Zhang, Learning a Mahalanobis distance metric for data clustering and classification, Pattern Recogn. 41 (12) (2008) 3600–3612.

[46] W. Zhao, R. Chellappa, P. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Comput. Surv. 35 (4) (2003) 399–458.

[47] F. Zhou, F. De la Torre, Deformable Graph Matching, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 2922–2929.

[48] L. Zhang, D. Samaras, Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics, IEEE Trans. Pattern Recognit. Mach. Intell. 28 (3) (2006) 351–363.