

Research Article

Cross-Modality 2D-3D Face Recognition via Multiview Smooth Discriminant Analysis Based on ELM

Yi Jin,¹ Jiuwen Cao,² Qiuqi Ruan,¹ and Xueqiao Wang¹

¹ Beijing Key Lab of Traffic Data Analysis and Mining, School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

² Institute of Information and Control, Hangzhou Dianzi University, Zhejiang 310018, China

Correspondence should be addressed to Yi Jin; yjin@bjtu.edu.cn

Received 12 January 2014; Accepted 16 March 2014; Published 23 April 2014

Academic Editor: Jun Cheng

Copyright © 2014 Yi Jin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, 3D face recognition has attracted increasing attention from worldwide researchers. Rather than homogeneous face data, more and more applications require flexible input face data nowadays. In this paper, we propose a new approach for cross-modality 2D-3D face recognition (FR), which is called Multiview Smooth Discriminant Analysis (MSDA) based on Extreme Learning Machines (ELM). Adding the Laplacian penalty constrain for the multiview feature learning, the proposed MSDA is first proposed to extract the cross-modality 2D-3D face features. The MSDA aims at finding a multiview learning based common discriminative feature space and it can then fully utilize the underlying relationship of features from different views. To speed up the learning phase of the classifier, the recent popular algorithm named Extreme Learning Machine (ELM) is adopted to train the single hidden layer feedforward neural networks (SLFNs). To evaluate the effectiveness of our proposed FR framework, experimental results on a benchmark face recognition dataset are presented. Simulations show that our new proposed method generally outperforms several recent approaches with a fast training speed.

1. Introduction

During the past several decades, face recognition (FR) has gained a widespread attention due to its potential application values as well as theoretical challenges compared with other biometrics [1]. 2D face recognition has achieved many good results under certain conditions. However, 2D FR is usually affected by different poses, expressions, illuminations, and occlusions and still needs to further improve the recognition accuracy in many real world applications. With the development of 3D scans and 3D information acquisition technologies, 3D face recognition (3D FR) has been proven to be quite robust to face variation with different illuminations and pose changes in achieving high recognition accuracy [2].

Up to now, a number of 3D FR methods have been proposed to solve such kind of FR problems, such as illuminations, pose, and expressions. Generally, 3D FR algorithms consist of three main steps: 3D data preprocessing and normalization, 3D facial feature extraction, and pattern classification. 3D facial feature extraction plays a vital role in 3D

FR process. The common 3D feature representation methods can be naturally grouped into three categories: (1) holistic feature-based methods [3–8], (2) local feature-based methods [9–13], and (3) the hybrid methods [14–17].

Though current 3D face recognition algorithms make a good performance in illumination and pose varieties, their performance declines in many new applications. For instance, only visible 2D photographs or low resolution video images of subjects are available while the gallery database may only consist of 3D face models. Nowadays, most FR systems and algorithms are not specifically developed for cross-modality 2D-3D face matching. Little work has been presented on 2D-3D feature representation. Furthermore, 3D facial feature extraction methods are generally time-consuming. The high computational complex limits its applications on high dimensional face features and large face databases.

To address this problem, a novel approach named Multiview Smooth Discriminant Analysis based on the recent Extreme Learning Machines (ELM) (Figure 1) is presented for cross-modality 2D-3D FR in this paper. In this new

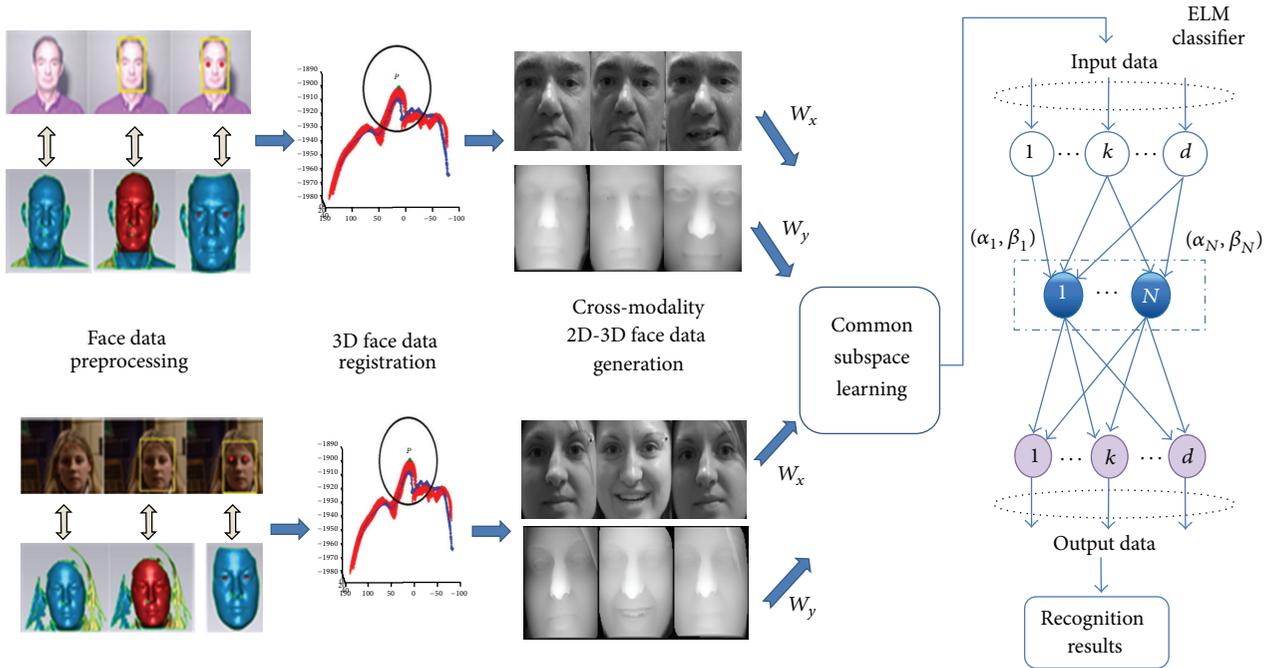


FIGURE 1: Intuitive explanation of our Multiview Smooth Discriminant Analysis based Extreme Learning Machines (ELM) approach. Multiview Smooth Discriminant Analysis is firstly used to obtain a more discriminant face representation; then ELM is utilized for pattern classification.

approach, Multiview Smooth Discriminant Analysis is first formulated to solve the multiple view-specific linear projections by adding the Laplacian smoothing constraint and to learn the common feature space for the cross-modality 2D-3D face features. Then, ELM is utilized as feature classifier by mapping the extracted feature to a high dimensional vector and treating the classification task as a regression problem.

The rest of this paper is organized as follows. Section 2 is the related works and Section 3 describes the Multiview Smooth Discriminant Analysis based on ELM method. Section 4 presents experimental results and discussions. The conclusion of this paper is drawn in Section 5.

2. Related Work

In this section, we briefly review some recent work related to our approach, such as 3D face recognition and Extreme Learning Machines (ELM).

2.1. 3D Face Recognition. 3D FR can be described into three steps: 3D data preprocessing and normalization, 3D facial feature extraction, and pattern classification. In general, 3D face feature representation methods are classified into: (1) holistic feature-based methods [3–8], (2) local feature-based methods [9–13], and (3) the hybrid methods [14–17].

The holistic-feature based methods usually use the whole face region as the input to extract facial features. Typical statistical methods, such as PCA [3] and LDA [4], which are popular 2D FR techniques, are also extended to 3D FR. PCA is used to extract the intrinsic discriminant feature vectors

on 2D intensity images and 3D depth images, respectively. Then, the fusion method is utilized to get the final results [5]. ICP-based 3D face recognition methods [6, 7], which utilize the entire facial surface directly as the holistic features, have been applied into 3D face registration [8], and the recognition result is achieved by using these separated rigid parts from the nonrigid parts. However, this kind of 3D FR methods somewhat fails to consider the local geometry which contains the intrinsic structure of the 3D data distribution. Thus, these methods are sensitive to expression, illumination, and pose variations.

The local-feature based methods utilize the 3D face appearance or regional geometric features to represent the 3D faces. Shapes, curvatures, and 3D facial landmarks, as well as other feature descriptors, are used as intrinsic 3D local structure to solve the FR problems. For instance, Queirolo et al. [9] used a simulated annealing-based approach (SA) and the surface interpenetration measure (SIM) to quantify the differences of two face images. Berretti et al. [10] explored the complete geometrical information of the 3D face model and use the iso-geodesic stripes to distinguish facial feature differences. Tang et al. [11, 12] proposed an expression insensitive 3D FR algorithm based on local pattern binary (LBP). Wang and Chua [13] proposed to use invariant 3D spherical Gabor filter (3D SGF) and the least trimmed square Hausdorff distance (LTS-HD) to handle the occlusions problems in 3D FR.

The hybrid methods jointly utilize the holistic and local features for 3D FR [16]. Compared with the other two categories, hybrid methods take advantage of both the 3D spatial information and the global statistic characteristics

and thus are demonstrated to be more robust in 3D FR. Spreuwers [14] proposed to tackle the face expression variations by dividing the face surface into partially overlapped small regions and then a decision-level fusion approach is applied on these regions. Ter Haar and Velkamp [15] preformed 3D face matching and evaluation using profile and contour of facial surface. Ming [17] proposed a 3D FR framework which utilizes the curvature information and orthogonal spectral regression for efficiently 3D discriminant feature extraction. However, most of these hybrid methods utilize two or more feature descriptors and thus have high computational complexity and cost.

Though new schemes have been proposed and achieved remarkable recognition performance on 3D face recognition, there are still some remaining unsolved problems that would affect FR performance. The performance of conventional 3D FR algorithms declines largely when there are only 2D images available for input test data. Besides, due to the expensive equipment, computational complexity, and the time-consuming 3D face preprocessing, it is difficult to perform on-line 3D FR in some real applications, such as airport or another security access control. So far as we know, there is very minimal work that has been done on this cross-modality 2D-3D face recognition [18–20]. Yang et al. [18] proposed a regularized kernel CCA method to learn the feature differences between 2D photos and 3D depth images. Jelsovka et al. [20] proposed a 3D range image and 2D-3D face images matching method by using the facial curves and CCA. Some similar approaches, such as deep CCA [21] and dictionary learning [22, 23], are also proposed to solve the cross-modality matching problem. CCA and its kernelization method are the typical approaches to learn a common subspace for two modality matching problem. However, CCA only learns the linear mapping by maximizing the total correlations between two views; it ignores the intraview and inter-view correlations. In other words, it has not taken the discriminative information into account. Recently, Kan et al. [24] proposed the Multiview Discriminant Analysis (MvDA) for cross-modality matching. Motivated by their work, we propose a Multiview Smooth Discriminant Analysis based on ELM. This new approach first uses a Laplacian smoothing constraint to make the mapping data spatially smooth and then takes advantage of the ELM as a high effective and less time-consuming classifier.

2.2. Extreme Learning Machines (ELM). In this subsection, we will briefly review the ELM and its applications on pattern classification [25, 26]. ELM is recently proposed for efficiently training single-hidden-layer feedforward neural networks (SLFNs). ELM performs classification by mapping data to a high dimensional vector and changes the classification task into a multioutput functional regression problem [27]. ELM provides better classification performance with a much shorter training time and the least human interference [26]. In [27], a voting based ELM has been developed to enhance the performance of multiclass classification. Employing multiple independent ELMs and making the final prediction with a majority voting method, V-ELM performs better than ELM with a higher classification rate. Huang et al. [28]

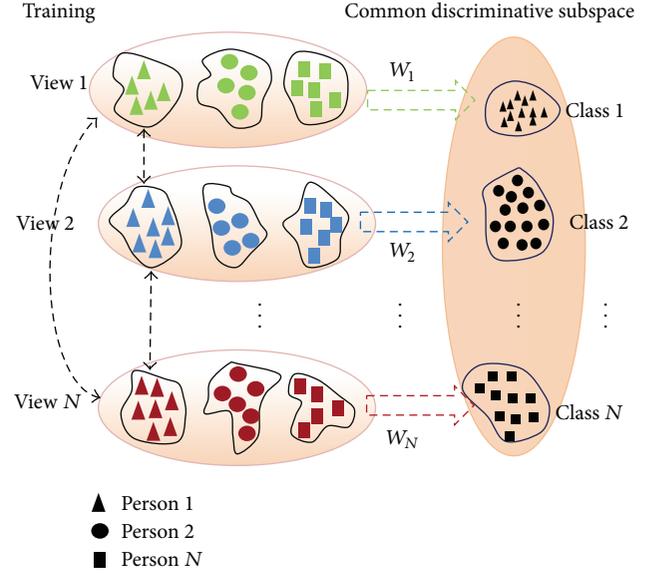


FIGURE 2: The principle of Multiview Smooth Discriminant Analysis (MSDA). The goal of MSDA is to build a smoothing common discriminative space into which the face samples from different views are projected. The different classes of samples are distributed in different views and the MSDA projects the samples into a common subspace to make the samples closer. Here, images with different labels are denoted in different shapes, such as triangle, circle, and square.

extended ELM to least square SVM (LS-SVM) [29] and proximal SVM (PSVM) [30] and provided a unified solution for multiclass classification. Kasun et al. [31] proposed an ELM-based auto encoder (ELM-AE) for big data application. Simulations on real world classification databases demonstrate that V-ELM generally outperforms several recent comparable methods with a fast training speed.

3. The Proposed Cross-Modality 2D-3D FR

In this section, we first introduce the basic idea and formulation of our proposed approach. Illustration of the Multiview Smooth Discriminant Analysis is shown in Figure 2. Then, we explain how to solve the optimization problem. Finally, the proposed 2D-3D FR approach is presented with data processing and ELM classification.

3.1. Multiview Smooth Discriminant Analysis. As it is shown in Figure 2, MSDA aims at finding n linear transform $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]$ which represent the projection matrixes from n views to the common discriminative space. Given the ν th view data samples as $X^{(\nu)} = \{x_{ij}^{(\nu)} \mid i = 1, 2, \dots, C; j = 1, 2, \dots, \rho_i^{(\nu)}\}$, where $x_{ij}^{(\nu)} \in \mathbb{R}^D$ is the i th sample from the ν th view of the j th class in the D dimensionality, C is the class number, and $\rho_i^{(\nu)}$ is the number of samples from the ν th view data, the projected data in the common space is denoted as $Y^{(\nu)} = \{y_{ij}^{(\nu)} = \alpha_\nu^T x_{ij}^{(\nu)} \mid i = 1, 2, \dots, C; j = 1, 2, \dots, \rho_i^{(\nu)}; \nu = 1, 2, \dots, n\}$. The Multiview Smooth Discriminant Analysis

(MSDA) aims at smoothing the basis vectors of the face data from different views by applying the Laplacian smoothing functional. The objective function of MSDA can be defined as the following:

$$(\alpha_1^*, \alpha_2^*, \dots, \alpha_n^*) = \arg \max_{\alpha_1, \dots, \alpha_n} \frac{(\alpha^T \mathbf{S}_B^x \alpha)}{(\alpha^T (\mathbf{1} - \lambda) \mathbf{S}_W^x \alpha + \lambda J(\alpha))}, \quad (1)$$

where \mathbf{S}_B^x and \mathbf{S}_W^x are the multiview between-class scatter matrix and within-class scatter matrix, respectively. $J(\alpha)$ is the discretized Laplacian regularization functional. λ is the smoothness-controlling parameter and $0 \leq \lambda \leq 1$.

(1) *Laplacian Smoothing*. The Laplacian operator is defined as the following [32]:

$$L\varphi(\tau) = \sum_{k=1}^{\varepsilon} \frac{\partial^2 \varphi}{\partial \tau_k^2}, \quad (2)$$

where φ is the function defined on the region-of-interest. And the Laplacian penalty function J , which measures the smoothness of the function φ , is defined as

$$J(\varphi) = \int_{\Omega} [L\varphi]^2 dt. \quad (3)$$

In our new proposed method, we mainly focus on face images, and thus we take discretized Laplacian smoothing method [33]. Let the $n_1 \times n_2$ face images be represented as vectors in \mathbb{R}^m and $\beta_i \in \mathbb{R}^m$ be the basis vectors to be smoothed. For an image, whose region-of-interest Ω is a two-dimensional lattice, let $\varepsilon = [\varepsilon_1, \varepsilon_2]$, where $\varepsilon_1 = 1/n_1, \varepsilon_2 = 1/n_2$ and the two-dimensional vectors $\tau_l = (\tau_{l1}, \tau_{l2})$, where $\tau_{lk} = (l_k - 0.5) \cdot \varepsilon_k, 1 \leq l_k \leq n_k, 1 \leq k \leq 2$. The total of grid points in the lattice is $n = n_1 \times n_2$. Suppose $\boldsymbol{\mu} = (\mu(\tau_1), \dots, \mu(\tau_n))$ is a n_j -dimensional vector which is a discretized version of function $\mu(\tau)$; then D_k , which is an $n_k \times n_k$ matrix that yields a discrete approximation to (2), has the property as follows:

$$[D_k \boldsymbol{\mu}]_l \approx \frac{\partial^2 \mu(\tau_l)}{\partial \tau^2}, \quad (4)$$

where $l = 1, \dots, n_k$. In our case, we choose the D_k to be the modified Neumann discretization [33, 34].

Given D_k , a discrete approximation for 2D Laplacian L is an $n \times n$ matrix and defined as

$$\Lambda = D_1 \otimes I_2 + D_2 \otimes I_1, \quad (5)$$

where I_k is $n_k \times n_k$ identity matrix for $k = 1, 2$. It is not difficult to prove that $\|\Lambda \cdot \alpha\|^2$ is directly related to the sum of the squared differences between nearby data points of α , which is an $n_1 \times n_2$ dimensional vector. And it also demonstrates that (5) measures the smoothness of α on the $n_1 \times n_2$ lattice.

(2) *The Algorithm and Solution*. In this part, we will illustrate the MSDA algorithm and its solution. According to the formulation of MSDA, the between-class and within-class

scatter matrixes are calculated from the samples of all the n views. Hence, the between-class and within-class scatter matrixes in (1) are formulated as the following:

$$\mathbf{S}_B^x = \begin{pmatrix} S_B^{(11)}, S_B^{(12)}, \dots, S_B^{(1n)} \\ S_B^{(21)}, S_B^{(22)}, \dots, S_B^{(2n)} \\ \vdots \\ S_B^{(n1)}, S_B^{(n2)}, \dots, S_B^{(nn)} \end{pmatrix}, \quad (6)$$

$$\mathbf{S}_W^x = \begin{pmatrix} S_W^{(11)}, S_W^{(12)}, \dots, S_W^{(1n)} \\ S_W^{(21)}, S_W^{(22)}, \dots, S_W^{(2n)} \\ \vdots \\ S_W^{(n1)}, S_W^{(n2)}, \dots, S_W^{(nn)} \end{pmatrix}.$$

The $S_B^{(vr)}$ and $S_W^{(vr)}$ are defined as

$$S_B^{(vr)} = \begin{pmatrix} \sum_{i=1}^C \frac{\rho_{ij} \rho_{iv}}{\rho_i} m_{ij}^{(x)} m_{iv}^{(x)T} \\ -\frac{1}{\rho} \left(\sum_{i=1}^C \rho_{ij} m_{ij}^{(x)} \right) \left(\sum_{i=1}^C \rho_{iv} m_{iv}^{(x)} \right)^T \\ \sum_{i=1}^C \left(\sum_{v} \rho_{ij} x_{ij}^{(v)} (x_{ij}^{(v)})^T - \left(\frac{\rho_{ij} \rho_{ij}}{\rho_i} \right) m_{ij}^{(x)} (m_{ij}^{(x)})^T \right) \\ -\sum_{i=1}^C \left(\frac{\rho_{ij} \rho_{ir}}{\rho_i} \right) m_{ij}^{(x)} (m_{ir}^{(x)})^T \end{pmatrix}, \quad \begin{matrix} j = r \\ j \neq r, \end{matrix} \quad (7)$$

where $m_{iv}^{(x)} = (1/\rho_{iv}) \sum_{j=1}^{\rho_{iv}} x_{ij}^{(v)}$. And the discretized Laplacian regularization function is denoted as

$$J(\alpha) = \|\Lambda \cdot \alpha\|^2 = \alpha^T \Lambda^T \Lambda \alpha, \quad (8)$$

where $\Lambda = \text{diag}(\Lambda_1, \Lambda_2, \dots, \Lambda_n)$.

Finally, the objective function of (1) can be determined by solving the following generalized eigenvalue decomposition with its leading eigenvalues:

$$\mathbf{S}_B^x \alpha = \eta \left((1 - \lambda) \mathbf{S}_W^x + \lambda \Lambda^T \Lambda \right) \alpha. \quad (9)$$

3.2. *ELM-Based Classification*. In order to speed up the training phase of the classifier and to obtain a reasonable recognition performance, the recent popular extreme learning machine [25, 28] is employed in our FR framework. Based on a SLFN, the ELM classifier utilized in the proposed FR recognition system can be described as follows.

Assuming that the available training feature dataset is $\mathcal{A} = \{(\mathbf{x}_i, t_i)\}_{i=1}^N$, where \mathbf{x}_i , t_i , and N represent the feature vector of the i th face image, its corresponding category index, and the number of images, respectively, the SLFN with ζ nodes in the hidden layer can be expressed as

$$\mathbf{o}_i = \sum_{j=1}^{\zeta} \mathbf{w}_j g(\mathbf{a}_j, b_j, \mathbf{x}_i), \quad i = 1, 2, \dots, N, \quad (10)$$

where \mathbf{o}_i is the output obtained by the SLFN associated with the i th input protein sequence and $\mathbf{a}_j \in \mathbb{R}^d$ and $b_j \in \mathbb{R}$ ($j = 1, 2, \dots, \zeta$) are parameters of the j th hidden node, respectively. The variable $\mathbf{w}_j \in \mathbb{R}^m$ is the link connecting the j th hidden node to the output layer and $g(\cdot)$ is the hidden node activation function. With all training samples, (10) can be expressed in the compact form as

$$\mathcal{O} = \mathcal{H}\mathcal{W}, \quad (11)$$

where $\mathcal{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_\zeta)$ and \mathcal{O} are the output weight matrix and the network outputs, respectively. The variable \mathcal{H} denotes the hidden layer output matrix with the entry $\mathcal{H}_{ij} = g(\mathbf{a}_j, b_j, \mathbf{x}_i)$. To perform multiclass classification, the ELM classifier generally utilizes the one-against-all (OAA) method to transform the classification application to a multioutput model regression problem. That is, for a C -categories classification application, the output label t_i of the face image feature \mathbf{x}_i is encoded to a \mathcal{C} -dimensional vector $\mathbf{t}_i = (t_{i1}, t_{i2}, \dots, t_{i\mathcal{C}})^T$ with $t_{i\mathcal{C}} \in \{1, -1\}$ ($\mathcal{C} = 1, 2, \dots, \mathcal{C}$). If the category index of the face image \mathbf{x}_i is \mathcal{C} , then $t_{i\mathcal{C}}$ is set to be 1 while the rest of entries in \mathbf{t}_i are set to be -1 . Hence, the objective of training phase for the SLFN in (10) becomes finding the best network parameters set $\Delta = \{\{\mathbf{a}_j, b_j, \mathbf{w}_j\}\}_{j=1, \dots, \zeta}$ such that the following error cost function is minimized

$$\min_{\Delta} E = \min_{\Delta} \|\mathcal{O} - \mathcal{T}\|, \quad (12)$$

where $\mathcal{T} = (\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_N)$ is the target output matrix.

ELM theory claims that random hidden node parameters can be utilized for SLFNs and the hidden node parameters may not need to be tuned. In such case, the system (11) becomes a linear model and the network parameter matrix can be analytically solved by using the least-square method. That is,

$$\mathcal{W} = \mathcal{H}^\dagger \mathcal{T}, \quad (13)$$

where \mathcal{H}^\dagger is the Moore-Penrose generalized inverse of the hidden layer output matrix \mathcal{H} given by [35]. The universal approximation property of the ELM algorithm is also presented in [25].

4. Experiments

In this section, we investigate the performance of our approach for cross-modality 2D-3D face recognition on FRGC 2.0. The new approach is compared with some state-of-the-art cross-modality learning methods, such as PLS [36], CDFE [37], PCA + CCA [38], and MvDA [39] and some neural networks based methods. The description of the face database and the 3D face preprocessing are presented in the following subsection. Then, the results of the experiments are concluded, as well as the discussion and experimental analysis.

4.1. Database Description and Experimental Setting. We evaluate our experiments on the FRGC v2.0 [40] 2D versus 3D face database. In this experiment, FRGC v2.0, which

contains 4007 2D photos and 4007 3D faces of 466 persons respectively, is utilized to evaluate the performance of the new method. The images of FRGC are acquired with a Minolta Vivid 910 and Minolta 910 scanner utilizes triangulation with a laser stripe projector to build a 3D face model. FRGC face database consists of frontal views up above shoulder, facial expressions, male and female face modals. Some data has facial hair, but none of them is occupied by glasses. In FRGC v2.0, 57% are males and 43% are females. Our previous work [41] is utilized for 3D data preprocessing and the 2D photos are corresponding to their respective 3D faces.

(1) *3D Data Preprocessing [41].* The 3D data preprocessing consists of four main steps, face region detection, nose detection, face smoothing, and the generation of 2D and 3D face images.

Firstly, a 3×3 Gaussian filter is used to moving spikes and noise firstly, and then the range data are subsampled at a 1:4 ratio. Ada-boost face detecting method [42] is applied on 2D texture image to help 3D facial region extraction. Secondly, we calculate the central stripe to detect the nose region and the nose tip is supposed to be on the central stripe. ICP is utilized to align the stripe of Person A to the stripe of Person B. Thus, the nose tip lays on the highest point in a cropped sphere. Once the nose tip is confirmed, a region-of-interest, which is defined by a sphere radius of 90 mm centered at the nose tip, is cropped and used in the following experiments. Figure 3 shows how to find the nose tips from the central stripe.

(2) *Experimental Setting.* In order to evaluate the robustness of our method, we divide the database into two sets, the training set and the test set. We pick out 285 subjects with more than 6 samples and select 5 samples of each person for training and the rest for testing. All 2D photos and the 3D range images are scaled, transformed, and cropped in the same way to 100×100 size according to the eye position. The cropped examples of FRGC database are shown in Figure 4.

4.2. Experimental Results. In the following experiments, we use the five front images per person as the training set, and the remaining images are utilized as the testing set. In the testing phase, the 2D photos are utilized as the gallery set and their corresponding 3D range images are used as the probe set. Firstly, we compare a set of experiments of our method with some cross-modality learning methods with hidden nodes of ELM chosen to be 1000. The rank-1 recognition performance with different dimensions is reported in Figures 5 and 6. Figure 5 shows the experimental results of the MSDA based on ELM method compared with some cross-modality learning methods, such as PLS [36], CDFE [37], PCA + CCA [38], and MvDA [39]. Figure 6 shows the comparison results of the proposed method with some feature extraction based methods. It is clear to see that our method achieves the highest recognition rate (96.8%) compared to the related subspace learning methods. Furthermore, we still choose 1000 hidden nodes and compare the new method under

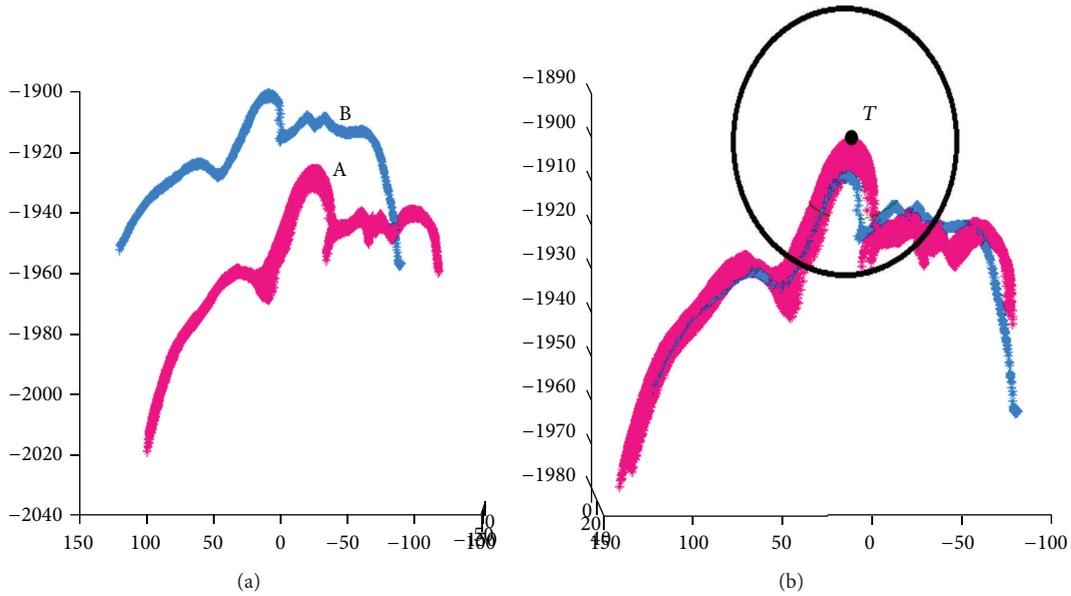


FIGURE 3: The illustration of ICP based nose-tip finding approach. (a) shows the two 3D face models, and (b) is the ICP based matching result.

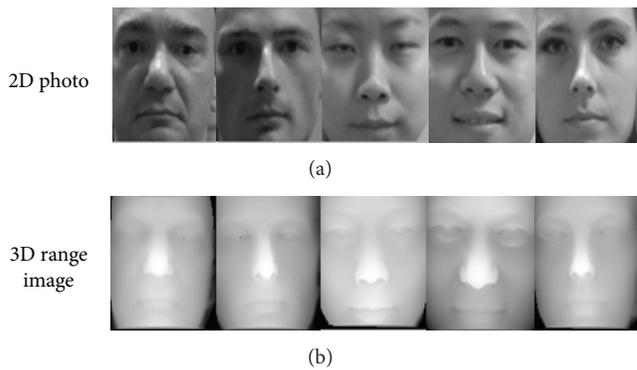


FIGURE 4: The cropped examples from FRGC database. The first row contains the examples of digital photo and the second row is the corresponding 3D range image.

different ELM activation functions, such as sigmoid, sine, and hardlim. Figure 7 shows the compared results. From Figure 7, we can conclude that ELM with sigmoid activation function gets the highest results in the three and the one with sine active function gets the worst.

Secondly, the recognition performance as well as the training time of the proposed approach is compared with BP neural networks [43]. The performance of the proposed method is compared with BP based method with a fixed dimension of 40. And the experimental results are obtained under different hidden nodes $\zeta = 100, 200, 500, 1000$. However, for huge computational cost of the BP neural networks, it is acceptable only when the choice of BP hidden nodes is less than 1000. The recognition results and training time are reported in Table 1. From Table 1, we can see that the ELM-based method achieves higher recognition accuracy with a much faster training speed than BP neural networks.

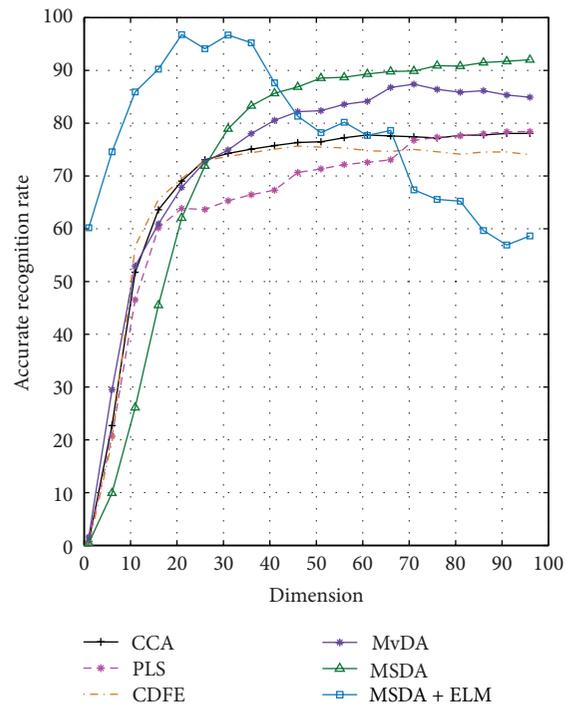


FIGURE 5: Comparison of the recognition rates with PLS, CDFE, PCA + CCA, and MvDA under different dimensions on the FRGC database.

Overall, the rank-1 recognition rate obtained by the new method is 96.8%, which is higher than that obtained using the other compared algorithms. It can be clearly seen that ELM-based MSDA method achieves better performance, and meanwhile, it takes a much faster training speed to get the good recognition result.

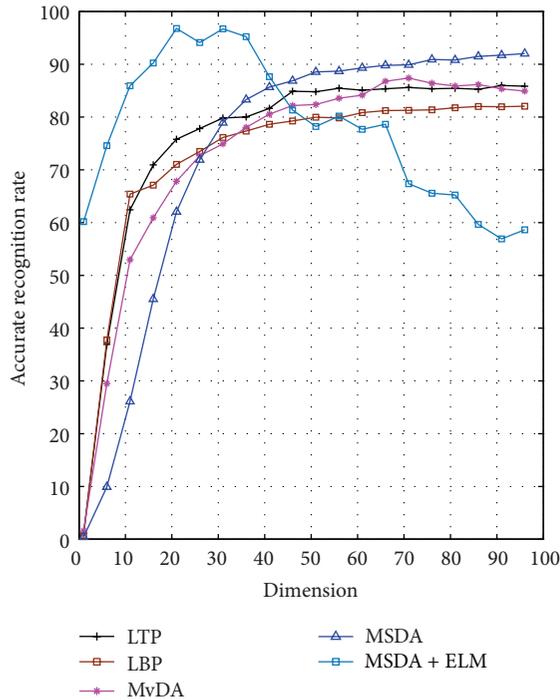


FIGURE 6: Comparison of the recognition rates with feature extraction methods under different dimensions on the FRGC database.

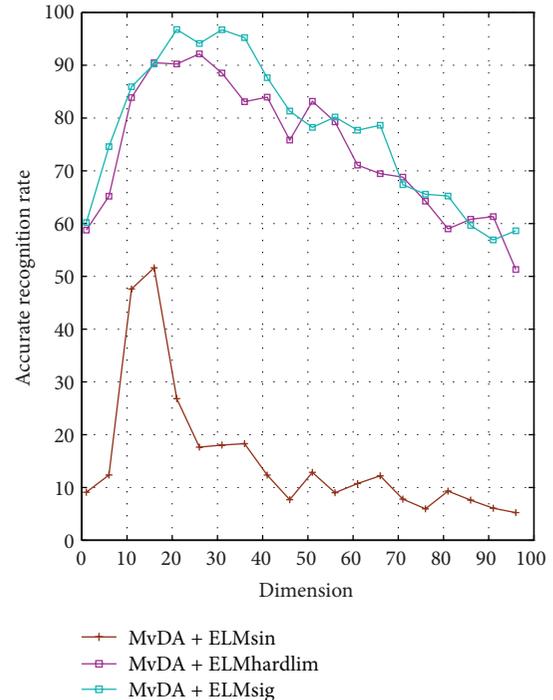


FIGURE 7: Comparison of the recognition rates with different ELM activation functions.

TABLE 1: Performance comparison of ELM and BP in terms of rank-1 Rec. Rate (in (%)) and training time (in (s)).

Nodes	Method			
	ELM		BP	
	R.R. (%)	Time (s)	R.R. (%)	Time (s)
100	42.5	0.0001	37.5	1214.6
200	70.0	0.0001	62.4	2209.8
500	85.5	0.0624	82.7	4055.6
1000	96.0	0.1872	94.0	6180.7

5. Conclusion

In this paper, a Multiview Smooth Discriminant Analysis based on ELM method is proposed for cross-modality 2D-3D face recognition. 2D-3D face recognition is an alternative and feasible approach to the traditional 3D FR systems, and it is much more convenient to acquire a 2D image than building a 3D face model. In this new approach, the Multiview Smooth Discriminant Analysis (MSDA) is firstly performed to get the cross-modality face features by using the Laplacian smoothing functional. The Laplacian penalized functional considers the image spatial relationship in the feature level and therefore obtains a much smoother linear projection subspace than those without smooth subspace learning. Furthermore, the ELM, which reduces the computational cost, is utilized for face feature classification. Experimental results show that the proposed method consistently outperforms the other cross-modality matching algorithms and achieves good recognition performance in both accuracy and speed.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by National Program on Key Basic Research Project (Grant no. 2012CB316304), the national Natural Science Foundation of China (Grant no. 61172128), Specialized Research Fund for the Doctoral Program of Higher Education (Grant no. 20120009120009), and the Fundamental Research Funds for the Central Universities (Grant no. 2012JBZ017).

References

- [1] S. Z. Li and A. K. Jain, *Encyclopedia of Biometrics: I-Z.*, vol. 2, Springer, New York, NY, USA, 2009.
- [2] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: a survey," *Pattern Recognition Letters*, vol. 28, no. 14, pp. 1885–1906, 2007.
- [3] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [4] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [5] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "An evaluation of multimodal 2D+3D face biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 619–624, 2005.

- [6] T. Maurer, D. Guigonis, I. Maslov et al., "Performance of geometrix activeid 3d face recognition engine on the frgc data," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, p. 154, San Diego, Calif, USA, 2005.
- [7] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "Multiple nose region matching for 3D face recognition under varying facial expression," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1695–1700, 2006.
- [8] C. S. Chua, F. Han, and Y. K. Ho, "3d human face recognition using point signature," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FG '00)*, pp. 233–238, 2000.
- [9] C. C. Queirolo, L. Silva, O. R. P. Bellon, and M. Pamplona Segundo, "3D face recognition using simulated annealing and the surface interpenetration measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, pp. 206–219, 2010.
- [10] S. Berretti, A. Del Bimbo, and P. Pala, "3D face recognition using isogeodesic stripes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2162–2177, 2010.
- [11] H. Tang, Y. Sun, B. Yin, and Y. Ge, "Expression-robust 3D face recognition using LBP representation," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '10)*, pp. 334–339, July 2010.
- [12] H. Tang, B. Yin, Y. Sun, and Y. Hu, "3d face recognition using local binary patterns," *Signal Processing*, vol. 93, no. 8, pp. 2190–2198, 2013.
- [13] Y. Wang and C.-S. Chua, "Face recognition from 2D and 3D images using 3D Gabor filters," *Image and Vision Computing*, vol. 23, no. 11, pp. 1018–1028, 2005.
- [14] L. Spreuwers, "Fast and accurate 3D face recognition: using registration to an intrinsic coordinate system and fusion of multiple region classifiers," *International Journal of Computer Vision*, vol. 93, no. 3, pp. 389–414, 2011.
- [15] F. B. ter Haar and R. C. Veltkamp, "A 3D face matching framework for facial curves," *Graphical Models*, vol. 71, no. 2, pp. 77–91, 2009.
- [16] S. Jahanbin, H. Choi, and A. C. Bovik, "Passive multimodal 2-D+3-D face recognition using gabor features and landmark distances," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 4, pp. 1287–1304, 2011.
- [17] Y. Ming, "Rigid-area orthogonal spectral regression for efficient 3d face recognition," *Neurocomputing*, vol. 129, pp. 445–457, 2013.
- [18] W. Yang, D. Yi, Z. Lei, J. Sang, and S. Z. Li, "2D-3D face matching using CCA," in *Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition (FG '08)*, pp. 1–6, September 2008.
- [19] D. Huang, M. Ardabilian, Y. Wang, and L. Chen, "Automatic asymmetric 3D-2D face recognition," in *Proceedings of the 20th International Conference on Pattern Recognition (ICPR '10)*, pp. 1225–1228, August 2010.
- [20] D. Jelsovka, R. Hudec, M. Breznan, and P. Kamencay, "2d-3d face recognition using shapes of facial curves based on modified cca method," in *Proceedings of the 22nd International Conference on Radioelektronika (RADIOELEKTRONIKA '12)*, pp. 1–4, 2012.
- [21] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep canonical correlation analysis," in *Proceedings of the 30th International Conference on Machine Learning*, pp. 1247–1255, 2013.
- [22] S. Wang, D. Zhang, Y. Liang, and Q. Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '12)*, pp. 2216–2223, 2012.
- [23] D.-A. Huang and Y.-C. F. Wang, "Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '13)*, December 2013.
- [24] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," in *Computer Vision (ECCV '12)*, vol. 7572 of *Lecture Notes in Computer Science*, pp. 808–821, 2012.
- [25] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.
- [26] G.-B. Huang, D. H. Wang, and Y. Lan, "Extreme learning machines: a survey," *International Journal of Machine Learning and Cybernetics*, vol. 2, no. 2, pp. 107–122, 2011.
- [27] J. Cao, Z. Lin, G.-B. Huang, and N. Liu, "Voting based extreme learning machine," *Information Sciences*, vol. 185, no. 1, pp. 66–77, 2012.
- [28] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 2, pp. 513–529, 2012.
- [29] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [30] G. Fung and O. L. Mangasarian, "Proximal support vector machine classifiers," in *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '01)*, pp. 77–86, August 2001.
- [31] L. L. C. Kasun, H. Zhou, G.-B. Huang, and C. M. Vong, "Representational learning with extreme learning machine for big data," *IEEE Intelligent Systems*. In press.
- [32] J. Jost and J. Jost, *Riemannian Geometry and Geometric Analysis*, vol. 2001, Springer, Berlin, Germany, 1995.
- [33] D. Cai, X. He, Y. Hu, J. Han, and T. Huang, "Learning a spatially smooth subspace for face recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–7, June 2007.
- [34] F. O'sullivan, "Discretized laplacian smoothing by fourier methods," *Journal of the American Statistical Association*, vol. 86, no. 415, pp. 634–642, 1991.
- [35] D. Serre, *Matrices: Theory and Applications*, Springer, New York, NY, USA, 2001.
- [36] A. Sharma and D. W. Jacobs, "Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 593–600, June 2011.
- [37] D. Lin and X. Tang, "Inter-modality face recognition," in *Proceedings of the European Conference on Computer Vision (ECCV '06)*, pp. 13–26, 2006.
- [38] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: an overview with application to learning methods," *Neural Computation*, vol. 16, no. 12, pp. 2639–2664, 2004.
- [39] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," in *Proceedings of the 12th European conference on Computer Vision (ECCV '12)*, pp. 808–821, 2012.

- [40] P. J. Phillips, P. J. Flynn, T. Scruggs et al., "Overview of the face recognition grand challenge," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 947–954, June 2005.
- [41] X. Wang, Q. Ruan, Y. Jin, and G. An, "Expression robust three-dimensional face recognition based on gaussian filter and dual-tree complex wavelet transform," *Journal of Intelligent and Fuzzy Systems*, vol. 26, no. 1, pp. 139–201, 2014.
- [42] P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision*, vol. 4, pp. 34–47, 2001.
- [43] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '89)*, pp. 593–605, June 1989.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

