

3D Face Recognition Using Spherical Vector Norms Map*

XUE-QIAO WANG^{1,2}, JIA-ZHENG YUAN^{1,3} AND QING LI^{1,2}

¹Beijing Key Laboratory of Information Service Engineering

²Computer Technology Institute

Beijing Union University

Beijing, 100101 P.R. China

³Beijing Advanced Innovation Center for Imaging Technology

Capital Normal University

Beijing, 100048 P.R. China

E-mail: {ldxueqiao; jiazheng; liqing10}@bnu.edu.cn

In this paper, we introduce a novel, automatic method for 3D face recognition. A new feature called a spherical vector norms map of a 3D face is created using the normal vector of each point. This feature contains more detailed information than the original depth image in regions such as the eyes and nose. For certain flat areas of 3D face, such as the forehead and cheeks, this map could increase the distinguishability of different points. In addition, this feature is robust to facial expression due to an adjustment that is made in the mouth region. Then, the facial representations, which are based on Histograms of Oriented Gradients, are extracted from the spherical vector norms map and the original depth image. A new partitioning strategy is proposed to produce the histogram of eight patches of a given image, in which all of the pixels are binned based on the magnitude and direction of their gradients. In this study, SVN map and depth image are represented compactly with two histograms of oriented gradients; this approach is completed by Linear Discriminant Analysis and a Nearest Neighbor classifier.

Keywords: spherical vector norms map, histograms of oriented gradients, 3D face recognition, linear discriminant analysis, face recognition grand challenge database

1. INTRODUCTION

With the rapidly decreasing costs of 3D capturing devices, many researchers are investigating 3D face recognition systems because it could overcome limitations illumination and make-up, but still bear limitations mostly due to facial expression. We summarize a smaller subset of expressive-robust methods below:

1. Deformable template-based approaches: Berretti *et al.* [1] proposed an approach that describes the geometric information of a 3D facial using a surface graph form, and the relevant information among the neighboring points could be encoded into a compact representation. 3DWWs (3D Weighted Walkthroughs) descriptors were proposed to demonstrate the mutual spatial displacement among pairwise arcs of points of the corresponding stripes. An 81.2% verification rate at a 0.1% FAR was achieved on the all vs. all experiment. The advantage of the method is the low computational complexity.

Kakadiaris *et al.* [2] mapped 3D geometry information onto a 2D regular grid using an elastically adapted deformable model. Then, advanced wavelet analysis was used for

Received January 11, 2017; revised March 27, 2017; accepted April 28, 2017.

Communicated by Chung-Lin Huang.

* This paper was supported by the Foundation of Beijing Union University (No. Zk10201703).

recognition and got good performance.

Drira *et al.* [3] used radial curves emanating from the nose tips which were already provided, and used elastic shape analysis of these curves to develop a Riemannian framework. Finally, they analyzed the shapes of full facial surfaces.

Amberg *et al.* [4] described an expression-invariant 3D Morphable Model for face recognition. The expression model was shown to significantly improve recognition. The strong prior knowledge of their model allows robust handling of noisy data.

2. Region fusion approaches: A regional-matching scheme was proposed by Faltemier *et al.* [19] firstly and the method is robust to facial expression. In their study, complete 3D face images were divided into 28 patches. The fusion results from independently matched regions could achieve good performance, but the authors did not report the computational time of their method.

Alyuz *et al.* [9] proposed an expression-resistant 3D face recognition method based on the region-based registration. A regional registration scheme was used to establish the relationship among all of the gallery samples in a single registration pass using common region models. An 85.64% verification rate at a 0.1% FAR was achieved on the ROC III experiment. Computational time of the method was not reported too.

Spreeuwiers *et al.* [23] presented a robust approach for 3D face registration. They proposed a 3D face classifier based on the fusion of many dependent overlapping face region classifiers. They used PCA-LDA for feature extraction. In addition, a voting approach was used and the decision was made by comparing the number of votes to a threshold. They obtain a 99.0% identification rate for first vs. rest experiment, while a verification rate of 94.6% at FAR=0.001 was obtained for the all vs. all verification test on the FRGC v2 data using fusion of 120 region classifiers. The strength of their method is that when the gallery only contains neutral faces, the method gets good performance.

3. Feature fusion approaches: Wang *et al.* [5] extracted Gabor, LBP, and Haar features from depth images, then the most discriminating local features were then selected optimally by boosting and were trained as weak classifiers to create three collective strong classifiers. The authors used three features for recognition and got good performances on FRGC database.

Mian *et al.* [6] extracted the Spherical Face Representation (SFR) feature using 3D facial data and the Scale Invariant Feature Transform (SIFT) descriptor using the corresponding 2D data to train a rejection classifier. The remaining faces were verified using a region-based matching method which is robust to facial expression. SFR is a simple, effective and expressive robust feature for 3D face recognition.

Mohammadzade *et al.* [7] presented a new Iterative Closest Normal Point (ICNP) method that could deal 3D faces with opened mouths. They performed experiments on FRGC database to prove that combining normal vectors and point coordinates could improve recognition performance. A verification rate of 99.6 percent at a FAR of 0.1 percent was achieved using that method for the all vs. all experiment. Their method got the best result in recent years.

Zhang *et al.* [8] found a novel, resolution-invariant, local feature for 3D face recognition. Six different scale invariant similarity measures were fused at the score level to finish recognition, which increased robustness against expression variation. The ad-

vantage is that their method is robust to poses.

Li *et al.* [22] proposed a novel highly discriminative 3D face feature named multi-scale and multi-component local normal patterns (MSMC-LNP) using normal vector of each point on 3D face then obtained a 96.3% recognition rate on the first vs. rest experiment. The strength of LNP histogram is that it includes both global and local cues of 3D face.

2. MOTIVATIONS AND MAIN CONTRIBUTIONS

As we know, facial expression variations are a critical issue in 3D face recognition, and the methods for overcoming expression variations are deformation approaches and fusion methods, which are presented in Section 1. However, the annotated deformable model method is quite computationally expensive.

Cloud data and normal vectors of points are original data of 3D face. Cloud data represents the 3D face which is created using points, while the normal vector describes the direction of each point. So, they are both important information of 3D face. Since 3D faces do not contain texture information, the distinctiveness of different 3D faces is small. Describing each 3D facial characteristic on a spherical surface and could maximize the between-class discriminatory. So, this paper presents a spherical vector norms map using the normal vectors of points on 3D cloud data to effectively describe the shape of the face surface. This feature map describes each 3D facial characteristic on a spherical surface and could thus maximize the between-class variability. In addition, inspired by the competitive performance and the computational efficiency of Histograms of Oriented Gradients (HOG) for behavior recognition, we propose to encode a 3D spherical vector norms map and depth image in the same way as the HOG operator for 2D image, generating histogram. Motivated by the two intuitions above, this paper proposes a novel automatic approach for 3D face recognition, which uses a new partitioning method of HOG (NP-HOG) by compacting the SVNS map into a histogram. The proposed method can be summarized as follows. First, the facial region is identified using an improved method described in an earlier study [10]. Secondly, each point's SVNs are obtained and then substitute for each point's z value. Subsequently, the spherical vector norm of each point on mouth region is adjusted within a small region. Then, we obtain the spherical vector norms map (SVNs map) by mapping SVNs onto the x - y plane. Histograms of Oriented Gradients (HOG) of the SVNs map and the depth image are then described using facial descriptors. The LDA method is used to train two feature subspaces using the FRGC v1 database. Then, the HOG of the SVNs map and the depth image are projected to the subspaces, respectively. The cosine distance is then used to measure the similarity of each pair of gallery and probe faces. The sum rule is used to fuse the two similarity matrices, and the Nearest Neighbor classifier is finally used to finish the recognition process.

In this study, we present a complete 3D face recognition system including preprocessing and recognition stages. Several comprehensive experiments are performed to prove the validity of the proposed method using BU_3DFE, BosphorusDB, and Face Recognition Grand Challenge (FRGC) databases. The primary contributions of this study can be summarized as follows.

The first contribution is a novel spherical vector norms map feature that is created using the normal vectors of 3D face cloud data. This feature could increase the distin-

guishability of similar regions in different 3D faces. In addition, the spherical vector norms map is robust to facial expression. Therefore, the proposed spherical vector norms map could increase the differences between different classes and reduce the dispersion within a given class. The second contribution is a new partitioning strategy of HOG descriptor, which is obtained by compacting the SVNS map into a histogram. This partitioning method maintains the completeness of the eyes and nose regions and could get better performance than traditional partitioning method. In addition, some state-of-the-art methods which perform excellent on first vs. all experiment but perform average on all vs. all experiment. Because for an expressional face, when the gallery only contains neutral faces, the method performs well, but when the gallery contains expressional faces of different people, the proposed method is unsatisfactory, while some approaches did the opposite. These two cases are both expression sensitive. Our approach achieved competitive scores on both experiments. It obtained a 96.5% recognition rate on the first vs. rest experiment and 98.76% recognition rate on the all vs. all experiment using FRGC v2 database. In addition, we got a 98.46% verification rate at 0.1% FAR on ROCIII experiment and 99.57% verification rate at 0.1% FAR on all vs. all experiment.

This paper is organized as follows. In Section 2, the 3D face data preprocessing methods are described. Then, the facial descriptor extraction method is presented in Section 3. In Section 4, the framework of the proposed 3D face recognition method is given. Experimental results are given in Section 5, and conclusions are drawn in Section 6.

3. FACIAL DESCRIPTOR EXTRACTING METHOD

3D faces contain hair, ears or shoulders data. To obtain uniform facial data and improve 3D face recognition, this study preprocesses the original 3D data before feature extraction. First, a 5×5 median filter is used to remove noise and spikes; then, the range data are subsampled at a ratio of 1:4. Then, the nose tip of the 3D face is located. In this study, each 3D face's nose tip is located using the method described in a previous study [10]. For reducing the computational cost, we change the following four parts:

- (1) Firstly, the method of getting the mirrored face is changed to: $\{x'_i = \max(X) + \min(X) - x_i; y'_i = y_i; z'_i = z_i\}$;
- (2) Secondly, before 3D face registration, the testing 3D face moves coarsely to the standard 3D face by the differential rent of the nose tips;
- (3) Thirdly, when the number of points of the testing face is above 20000, we downsample the number of the points into a half;
- (4) Finally, we crop a sphere (radius=80mm) centered at the nose tip as face region. Thus, the preprocessing step requires 5.5 seconds. Finally, the bilinear interpolation method is applied to normalize each face to be represented by three 384×384 matrices.

After preprocessing the 3D data, each 3D face's spherical vector norms map is created. Throughout the remainder of this paper, we refer to spherical vector norms maps as SVNs maps. Then, a new partitioning Histograms of Oriented Gradients of SVNs Map and depth image are obtained for recognition. Firstly, we define the spherical vector norm in Section 3.1.

3.1 Spherical Vector Norm

To better describe a face's surface shape and increase the discrimination of a 3D face, a novel feature is proposed for extracting detailed information from a 3D face; this feature is denoted as the spherical vector norm.

First, we define the spherical vector to be a vector that originates from a point on the 3D face and ends on a spherical surface. An example is shown in Fig. 1, where v_1 and v_2 are the spherical vectors of two different points. The center of the sphere is the nose tip and the radius of the sphere is 100 mm. The direction of spherical vector is the same as the direction of normal vector. Fig. 1 shows that using spherical vectors to establish 3D facial features could increase the distinguishability of certain smooth areas of a 3D face.

A spherical vector could be obtained in two steps: (1) combine the spherical equation, where the sphere is centered at the nose tip $p(a_1, b_1, c_1)$, and the linear equation, where the line crosses at p_1 and p_2 ; and (2) calculate the point of intersection $p_2(x, y, z)$ on the sphere such that:

$$\begin{cases} (x-a_1)^2 + (y-b_1)^2 + (z-c_1)^2 = R^2 \\ \frac{x-a_2}{u} = \frac{y-b_2}{v} = \frac{z-c_2}{w} \end{cases} \quad (1)$$

where $p(a_1, b_1, c_1)$ is the nose tip of the 3D face; (a_2, b_2, c_2) is point p_1 on the 3D face; (u, v, w) is the direction of v_1 ; and R is the radius of the sphere. The method of obtaining p_2 is presented in appendix I.

Second, the spherical vector v_1 is calculated using the following equation:

$$v_1 = p_2 - p_1 \quad (2)$$

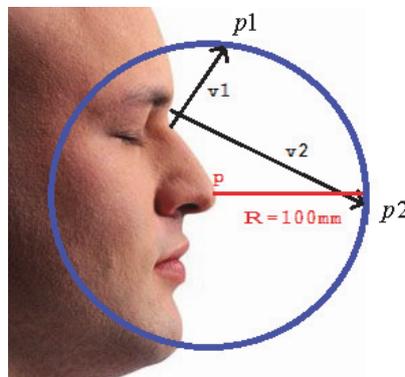


Fig. 1. v_1 and v_2 are spherical vectors.

The spherical vector norm (SVN) is the norm of the spherical vector such that:

$$n = \sqrt{u^2 t^2 + v^2 t^2 + w^2 t^2} \quad (3)$$

Using SVN's could increase the distinguishability of different local shape patterns, including convex patterns, concave patterns, *etc.*, as shown in Fig. 2. For instance, in Fig. 2 (a), p_1 and p_2 have the same z values but different SVN's. When each point's spherical vector norms substitutes for z value, the shape of the surface changes.

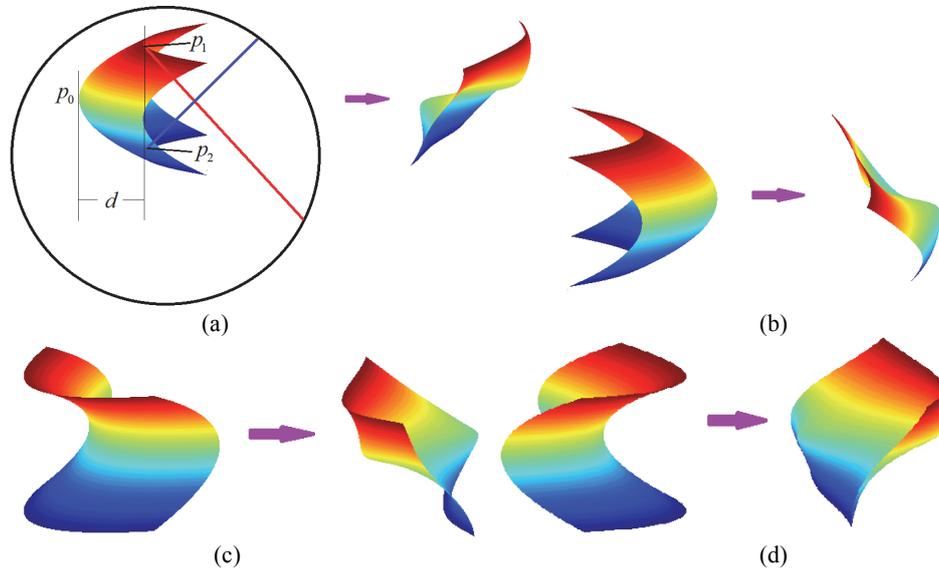


Fig. 2. SVN's increase the distinguishability of different local shape patterns, such as convex and concave patterns.

For instance, in Fig. 2 (a), let point $p_0 = (x_0, y_0, z_0)$ and the concave equation be $\frac{(x-x_0)^2}{a^2} + \frac{(y-y_0)^2}{b^2} = 2z$. Thus, $p_1 = (x_0, y_0 + b\sqrt{2d}, z_0 + d)$, while $p_2 = (x_0, y_0 - b\sqrt{2d}, z_0 + d)$. The normal vector of p_1 is $n_1 = (0, \frac{-\sqrt{2d}}{\sqrt{b^2+2d}}, \frac{b}{\sqrt{b^2+2d}})$, while $n_2 = (0, \frac{\sqrt{2d}}{\sqrt{b^2+2d}}, \frac{b}{\sqrt{b^2+2d}})$ is the normal vector of p_2 . The SVN of p_1 is $SVN_1 = \frac{\sqrt{2d}(y_0 + b\sqrt{2d}) - (z_0 + d)b}{\sqrt{b^2 + 2d}} + \sqrt{(\frac{\sqrt{2d}(y_0 + b\sqrt{2d}) - (z_0 + d)b}{\sqrt{b^2 + 2d}})^2 - [x_0^2 + (y_0 + b\sqrt{2d})^2 + (z_0 + d)^2 - R^2]}$, and the SVN of p_2 is $SVN_2 = \frac{\sqrt{2d}(b\sqrt{2d} - y_0) - (z_0 + d)b}{\sqrt{b^2 + 2d}} + \sqrt{(\frac{\sqrt{2d}(b\sqrt{2d} - y_0) - (z_0 + d)b}{\sqrt{b^2 + 2d}})^2 - [x_0^2 + (y_0 - b\sqrt{2d})^2 + (z_0 + d)^2 - R^2]}$. Let $f = SVN_1 - SVN_2$ thus, $f = SVN_1 - SVN_2 = \frac{2\sqrt{2d}y_0}{\sqrt{b^2 + 2d}} + \sqrt{(\frac{\sqrt{2d}(y_0 + b\sqrt{2d}) - (z_0 + d)b}{\sqrt{b^2 + 2d}})^2 - [x_0^2 + (y_0 + b\sqrt{2d})^2 + (z_0 + d)^2 - R^2]} - \sqrt{(\frac{\sqrt{2d}(b\sqrt{2d} - y_0) - (z_0 + d)b}{\sqrt{b^2 + 2d}})^2 - [x_0^2 + (y_0 - b\sqrt{2d})^2 + (z_0 + d)^2 - R^2]}$.

So, when $d = 0, f = 0$; when $y_0 = 0, f = 0$; If $SVN_2 > SVN_1, f < 0$, and if $SVN_2 < SVN_1, f < 0$. Thus, $\exists d = d'$ makes $f = 0$.

For the points p_1 and p_2 , which have the same z value, their SVNs have different values except for the three cases mentioned above. When all of the points' z values are substituted by SVNs, the new face is more discriminating than the original face. For example, in Fig. 3, the blue face is the original 3D face; then, each point's z value is substituted with its SVN to create the red face. As shown, certain regions were enhanced during this process.

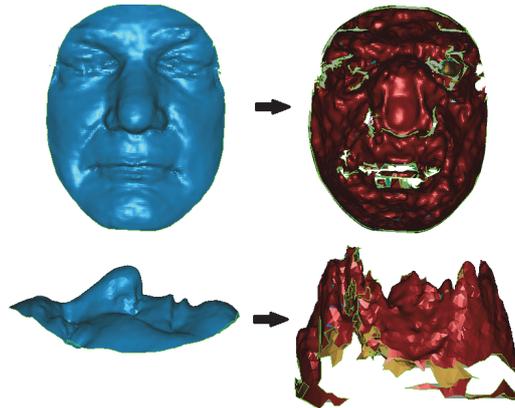


Fig. 3. SVNs increase the distinguishability of points on a 3D face.

3.2 Spherical Vector Norms Map

To reduce the within class dispersion which is mainly caused by facial expressions, an adjustment is made to each 3D face's mouth region. We use the same 3D face (the first 3D face in FRGC v1) in preprocessing section as reference. The mouth region is showed within the red ellipse in Fig. 4. Since the mouth is close on the reference, we find the corresponding relationship of the testing face and the reference on the mouth area for adjusting the opened mouth.



Fig. 4. The reference face's mouth area which is used for adjustment.

Since the testing face was aligned with the reference face in the preprocessing section, the positions of testing face and reference face are under the same coordinate sys-

tem, as shown in Fig. 5. First, we find the corresponding points between the reference mouth S and the input face. For each point s_i in reference mouth S , we find the closest point p_j in 3D face P using the Euclidean distance. Let $\{p_i^C\}_{i=1}^N$ denote the closest point set such that $p_i^C = p_k, k = \arg \min (\|s_i - p_j\|)$. Then, the spherical vector norm is located within a window of size 5×5 mm centered at p_i^C that has the smallest difference from the spherical vector norm n_i of the corresponding reference point. Let $\{n_i^C\}_{i=1}^N$ represent the set of the spherical vector norms such that:

$$n_i^C = n_k, k = \arg \min_{j \in \eta} (|n_i - n_j|) \quad (4)$$

where η denotes the set of indices of the spherical vector norms of the points that are located within a window of size 5×5 mm centered at the closest point, as shown in Fig. 5. Then, the z value of each point s_i is substituted with n_i^C , thus a new SVN's face F_{SVNs} is established. Finally, F_{SVNs} is projected onto the x - y plane to obtain the SVN's map.

Six neutral faces' SVN's maps are shown in Fig. 6 (a), while six other faces with expressions are shown in Fig. 6 (b). Fig. 6 (a) shows the eye regions are enhanced. Additionally, it is shown that the mouth is always closed in Fig. 6 (b). We expect that the success of the proposed method in enhancing the eyes region and managing opened mouths could significantly improve 3D face recognition.

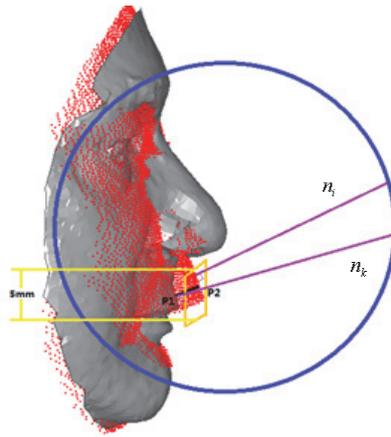
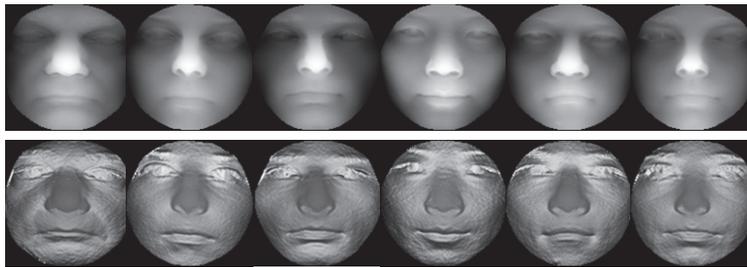
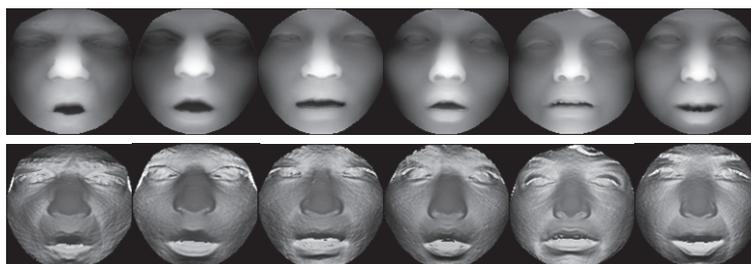


Fig. 5. Algorithm of adjusting the spherical vector norm.



(a) Six neutral faces' SVN's map.

Fig. 6. Examples of SVN's maps from the FRGC database.



(b) Six expressional faces' SVNs map.

Fig. 6. (Cont'd) Examples of SVNs maps from the FRGC database.

3.3 Histograms of Oriented Gradients of SVNs Map and Depth Image

Inspired by the competitive performance and the computational efficiency of Histograms of Oriented Gradients (HOG) for feature extraction, and for maintaining the completeness of the eyes and nose regions, this paper proposes a new partitioning method to generate a novel 3D geometric Histograms of Oriented Gradients for recognition.

In this section, we first recall the basics of Histograms of Oriented Gradients. Then, we present a new partitioning method to generate a novel 3D geometric facial description. The HOG of an SVNs map encodes the gradient direction and magnitude of the image variances in a fixed length feature vector. The information contained in the SVNs map and depth image can therefore be represented by an HOG histogram.

However, the original HOG is not as discriminating as expected for 3D facial representation because it cannot correctly distinguish similar local surfaces. To address this problem, two solutions are considered. First, using the HOG of SVNs map could distinguish similar local surfaces. Second, a new partitioning strategy is introduced to represent local surfaces to different extents, which are then combined to create a comprehensive description.

Histograms of Oriented Gradients (HOG) [16] have been successfully used as a texture descriptor in many applications related to human behavior recognition and other computer vision applications. A HOG descriptor produces a histogram of a given image in which all of the pixels are binned based on the magnitude and direction of their gradients.

Specifically, the first step of this calculation is the computation of each point's gradient value and gradient direction such that:

$$G(x, y) = \sqrt{G_x(x, y)^2 + G_y(x, y)^2} \quad (5)$$

$$\alpha(x, y) = \tan^{-1}\left(\frac{G_y(x, y)}{G_x(x, y)}\right) \quad (6)$$

where $G_x(x, y) = H(x+1, y) - H(x-1, y)$ and $G_y(x, y) = H(x, y+1) - H(x, y-1)$, $H(x, y)$ is the pixel value of point (x, y) , $G(x, y)$ is the gradient value of point (x, y) , and $\alpha(x, y)$ is the gradient direction of point (x, y) .

The HOG descriptor is then identified as the concatenated vector of the components of the normalized cell histograms from all of the block regions. In this study, 2×2 cells

or 3×3 cells of each block are used and each cell's oriented gradients are normalized into 30 bins or 18 bins, as shown in Fig. 7.

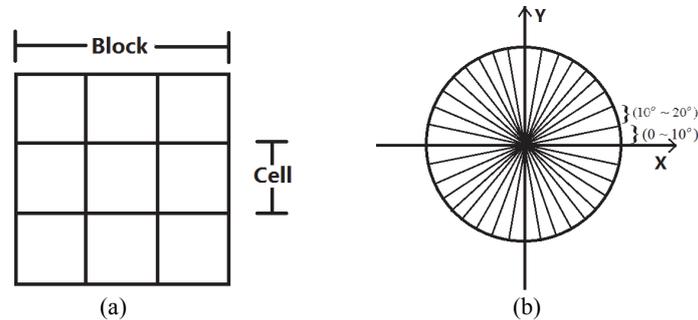


Fig. 7. (a) Block and cell of HOG; and (b) Eighteen bins of each cell.

Unfortunately, this direct application of HOG to depict the shapes of 3D surfaces also leads to unexpected confusion when distinguishing similar yet different local shapes. This lack of descriptive power is problematic when one needs to derive a facial description to enhance distinctiveness for face recognition. Thus, we proposed a new partitioning strategy, which uses SVN maps to overcome this weakness of HOG.

In the proposed algorithm, HOG is applied to the SVN maps, which are 384×384 pixels and are divided into eight image patches. The partitioning method used is shown in Fig. 8, and the positions of the top left corner and the size of each patch are shown in Table 1. Then, the HOG descriptors of the eight patches are combined into one feature vector. We use overlapping block descriptors for each patch, and the size of each block is 50×50 , while the size of the overlap is 25 pixels. A total of 16 blocks with 30 bins in each cell are used in Figs. 8 (b) through (g); a total of 14 blocks with 9 bins in each cell are used in Fig. 8 (a); and a total of 56 blocks with 9 bins in each cell is used in Fig. 8 (h). The dimension of the HOG descriptor of each block of (b) to (g) is 120, while (a) and (h) is 81; thus, the dimension of the proposed HOG descriptor is 17190. In addition, different weights are used into eight patches. Five times weights are applied to patch (c) and (f), while two times are applied to patch (b), (d), (e) and (g). Throughout the rest of this paper, we refer to the new partitioning method of HOG simply as the NP-HOG.

The proposed partitioning strategy, which maintains the eye and nose regions of the original 3D data, could produce improved performance. To determine the effectiveness of the proposed method, an experiment is performed for comparison with the other partitioning methods detailed in Section 5.1. This HOG partitioning method contains significantly detail information of local shape variations.

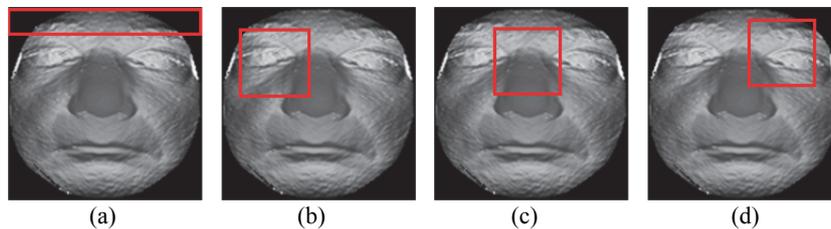


Fig. 8. Eight image patches which NP-HOG extracting method applied.

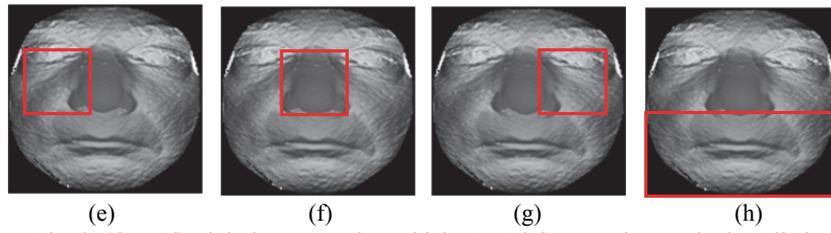


Fig. 8. (Cont'd) Eight image patches which NP-HOG extracting method applied.

Table 1. Position of the top left corner and the size of each patch.

Patch	Blocks	Cells	The position of the top left corner of each patch	Size of patch	The dimension of feature vector of each patch
(a)	14	3×3	(1, 1)	50×384	1134
(b)	16	2×2	(48, 38)	125×125	1920
(c)	16	2×2	(48, 129)	125×125	1920
(d)	16	2×2	(48, 219)	125×125	1920
(e)	16	2×2	(129, 38)	125×125	1920
(f)	16	2×2	(100, 129)	125×125	1920
(g)	16	2×2	(129, 193)	125×125	1920
(h)	56	3×3	(257, 1)	125×384	4563

4. FACE RECOGNITION METHOD

After extracting the new partitioning Histograms of Oriented Gradients (NP-HOG) of SVNs Map and depth image, we fuse the two descriptors for 3D face recognition. The flow-process diagram of the proposed 3D face recognition method is shown in Fig. 9.

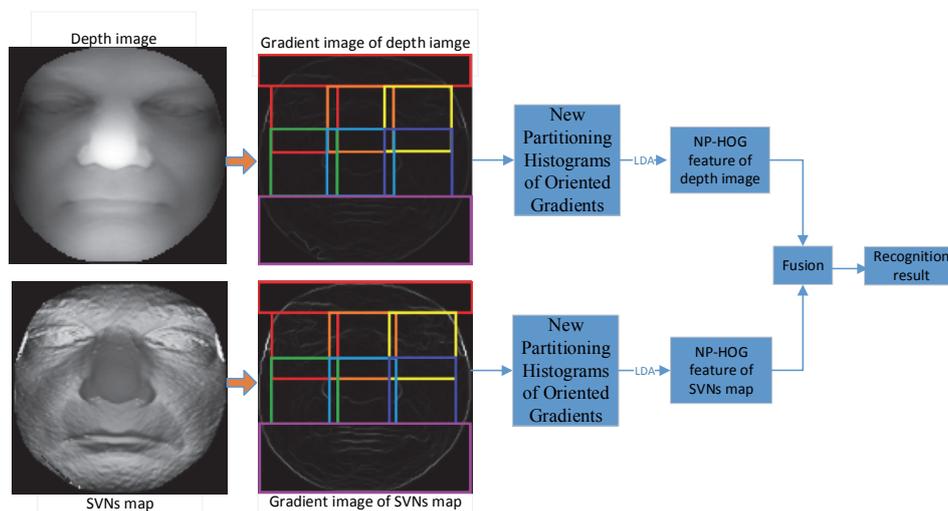


Fig. 9. Flow-process diagram of the proposed method.

The training and testing steps are presented as follows:

During training, we use all 943 faces in the FRGC 1.0 dataset for training. First, we obtain the HOG features of the SVNs map and depth image of each training face using the proposed partitioning strategy. Then, we use LDA [5] to identify two discriminant subspaces. Because FRGC 1.0 dataset contains 275 persons, the best LDA projecting dimension is 274. And then we record the transformation matrices. Finally, we obtain the feature vectors of the gallery faces by projecting their HOG feature vectors onto their corresponding LDA spaces.

During testing, we obtain all of the probe faces' features using the same method. The cosine distance is used to establish two similarity matrices between the gallery features and probe features; the sum rule is used to fuse the two similarity matrices, and the Nearest Neighbor classifier is used to finish the recognition process.

5. RESULTS AND DISCUSSION

We perform the proposed experiments using the BU-3DFE database [17], Bosphorus database [21] and FRGC [22] 3D face database. The BU-3DFE database [17] consists of 100 subjects in various expressions. Each person is represented by 25 scans that contain one neutral face and 24 different expressional faces. Some samples of the original depth images and SVNs maps of BU-3DFE database are shown in Fig. 10.

The Bosphorus database consists of 105 subjects in various poses, expressions and occlusion conditions. Eighteen subjects have a beard or moustache, while 15 subjects have short facial hair. The majority of the subjects are between 25 and 35 years old. Twenty-seven professional actors or actresses are incorporated in the database. Up to 54 facial scans are available per subject, but 34 of these subjects have 31 scans. The number of total face scans in Bosphorus database is 4666. In this study, we use all faces for the experiment. Some samples of the original depth images and SVNs maps of the Bosphorus database are shown in Fig. 11.

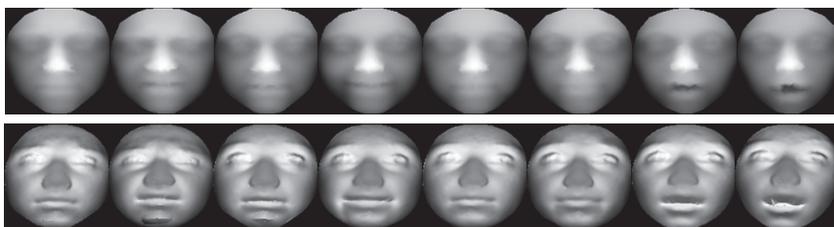


Fig. 10. Original depth images and SVNs maps of same samples in BU_3DFE database.

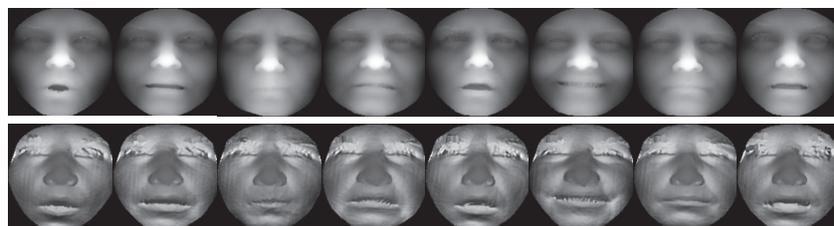


Fig. 11. Original depth images and SVNs maps of same samples in Bosphorus database.

FRGC v1 contained 943 3D faces of 275 persons, while FRGC v2 contained 4007 faces of 466 persons. The images were acquired with a Minolta Vivid 910. The 3D faces are available in the form of four matrices, each of which contains 640×480 pixels and consist of frontal views. Some subjects have facial hair, but none are wearing eyeglasses. The 2D faces correspond to their respective 3D face. In FRGC v2, 57% are male and 43% are female. The database was collected during the years of 2003 and 2004. To evaluate the robustness of the proposed method against expression variations, we classified 1648 faces with expression as a non-neutral dataset of 411 persons and 2359 neutral faces as a neutral dataset of 422 persons. The numbers of faces in the neutral and non-neutral data sets are not equal because some people in FRGC v2 contained only one face.

We finished four experiments which contained neutral vs. neutral experiment, neutral vs. expression experiment, all vs. all experiment, and ROCIII experiment. In the all vs. all experiment, every image in FRGC v2 database is matched with all of the other remaining images, resulting in 16,052,042 combinations. Similarly, in neutral vs. neutral experiment, every image of the neutral database is matched with all of the other remaining images, resulting in 5,562,522 combinations. In neutral vs. expression experiment, the gallery images originated from the neutral data set, and the probe entries originated from the expression data set. In ROCIII experiment, the gallery images originated from the Fall 2003 semester, while the probe entries originated from the Spring 2004 semester.

5.1 Comparisons With Other HOG Partitioning Strategies

The proposed method is now compared with other HOG partitioning strategies, and the all vs. all experiment with the SVNs map is completed in this section. LDA is used to establish a subspace of each partitioning strategy, and the rank-one recognition rates of different partitioning strategies are shown in Table 2. The dimensions of the HOG descriptor of each partitioning method are also shown. As shown in the table, the proposed strategy achieved the best result.

Table 2. Rank-one recognition rates of different partitioning strategies.

Partitioning strategy	Dimension of HOG descriptor	Rank-one recognition rate
4 × 4	1296	85.62%
8 × 8	5184	89.34%
12 × 12	11664	90.48%
16 × 16	20736	90.21%
20 × 20	32400	89.6%
Our strategy	21627	98.25%

5.2 Comparisons With SVNs Map Without Mouth Region Adjustment

For decreasing the influence of expression for recognition, we make an adjustment on mouth region in this paper. In this part, we compare SVNs map with mouth region adjustment and SVNs map without mouth region adjustment using neutral vs. expression experiment on three 3D databases.

As shown in the Table 3, SVNs maps with mouth region adjustment perform better, so the adjustment method to the mouth area is efficient.

Table 3. Comparing Rank-one recognition rates to SVN's map without mouth region adjustment.

3D face database	With mouth region adjustment	Without mouth region adjustment
FRGC v2	95.28%	93.54%
Bosphorus	87.93%	85.94%
BU_3DFE	91.63 %	88.75%

5.3 Comparing the Fusion Method with SVN's Map + LDA and Depth Image + LDA

In this section, we combine the SVN's map with the original 3D depth images for 3D face recognition. HOG features of the proposed partitioning strategy are extracted from the SVN's map and original depth image, respectively; we then use the simple sum rule to fuse these two modalities. Thus, the final matching score for a pair of gallery-probe faces is the summation of the cosine distances from the feature vectors of the two modalities. We perform four experiments in this section, and the training and testing set configurations of the four experiments are presented in Table 4.

Using a simple fusion rule, we showed that combining the SVN's map and the depth image can improve recognition performance. The CMC and ROC curves of the four experiments are shown in Figs. 12 and 14. As shown, combining the SVN's map with the original 3D depth faces can improve recognition performance.

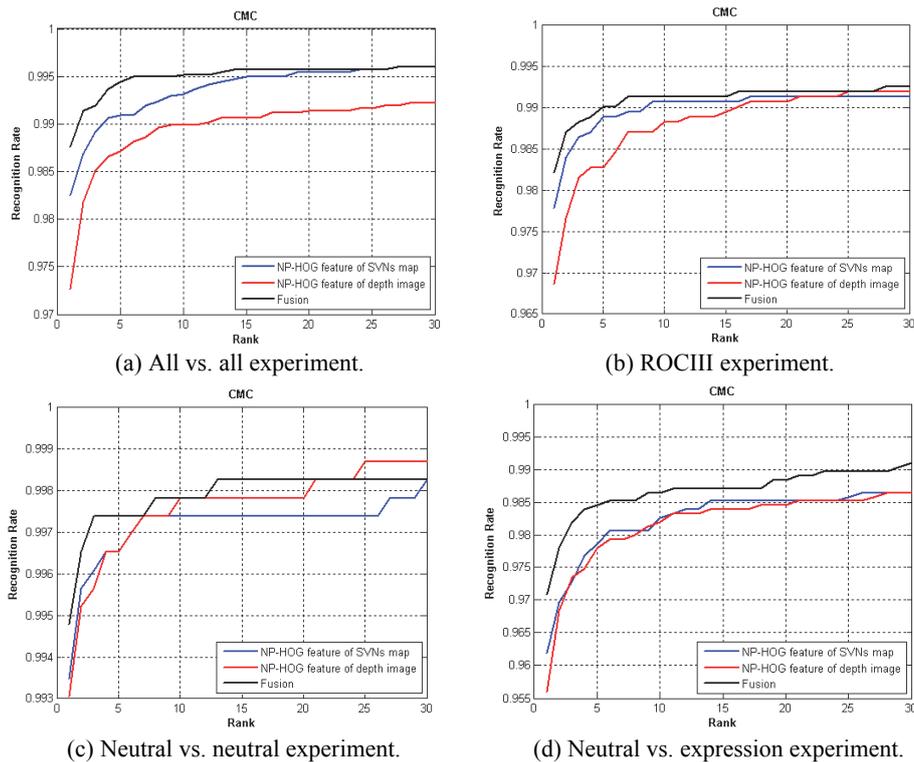


Fig. 12. CMC curves of the four experiments.

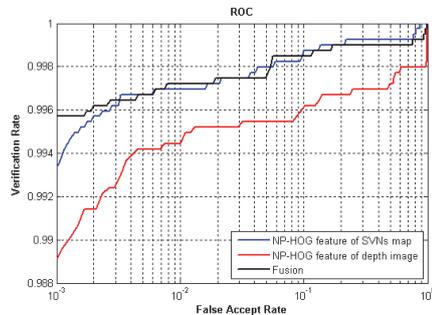
Table 4. Training and testing set configurations of the four experiments.

Experiments	Gallery	Probe
all vs. all	4007 faces of FRGC v2	4007 faces of FRGC v2
neutral vs. neutral	2359 neutral faces	2359 neutral faces
neutral vs. expression	2359 neutral faces	1648 expressional faces
ROCIII	1893 faces from the fall of 2003	2114 faces from the spring of 2004

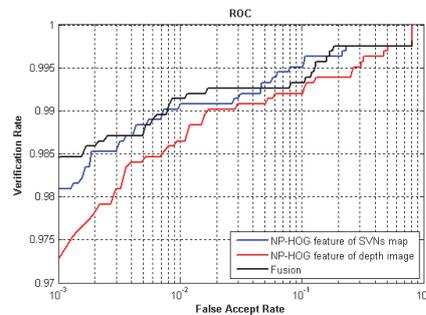
Some failure cases of fusion result are shown in Fig. 13. From the figure, we can find that hair and inaccurate nose tip detection cause failure recognition result.



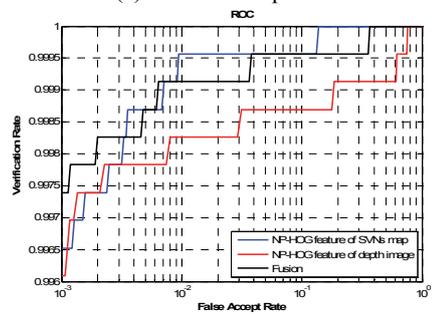
Fig. 13. Some failure cases of the fusion method.



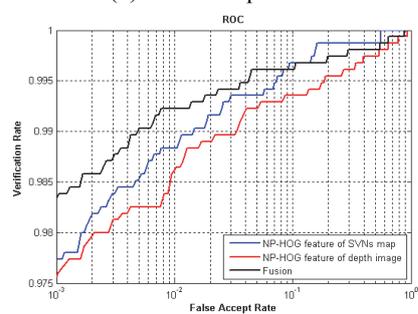
(a) All vs. all experiment.



(b) ROCIII experiment.



(c) Neutral vs. neutral experiment.



(d) Neutral vs. expression experiment.

Fig. 14. ROC curves of the four experiments.

5.4 Comparison to Li

Because the paper of Li *et al.* [22] used MSMC-LNP into 3D face recognition, which is established by the normal of 3D face as we do, we compared our method with their approach in this section. Since they do not use their method on the standard masks

for the verification experiment of FRGC database (as the other state-of-the-art methods), we did two experiments which apply the same gallery and probe to them. The first one uses FRGC v2.0 database. The gallery is the first face of each person of FRGC v2, while the probes are the neutral faces and expression faces. The second is for the BU-3DFE database, we divide all the probes into four subsets according to the expression intensity levels as the same method to Li *et al.* [22]. Each subset consists of 600 probes with the same level of expression intensity but different expression types.

From Table 5, we could find that our method got better result than Li *et al.* [22] using FRGC database. Still, as shown in Table 6, our approach got better result when using BU_3DFE database for experiment, especially as the elevating of the level of expression.

Table 5. Comparison with Li *et al.* of the performances due to facial expression variations on the FRGC v 2.0 database.

Methods	Train set	First vs. neutral	First vs. expression
MSMC-LNP	Bosphorus	98%	94.2%
Our method	FRGC v1	98.35%	94.65%

Table 6. Comparison with Li *et al.* of the performances due to facial expression variations on the BU 3DFE database.

Methods	Train set	Level 1	Level 2	Level 3	Level 4
MSMC-LNP	Bosphorus	97.3%	95%	92.7%	83.8%
Our method	FRGC v1	98.33%	97%	95%	91.17%

5.5 Comparison to State of the Art: Verification Rate and Processing Speed

The proposed method is now compared with other state-of-the-art methods. Table 7 shows the verification results for the state-of-the-art methods using the FRGC database, as reported in the literature; note that only the results of their 3D engines are shown. In neutral vs. neutral experiment and neutral vs. expression experiment, our method gets the best performances. For all vs. all experiment and ROCIII experiment, the performances of the proposed experiments are shown to be marginally lower than the best performer.

Table 8 lists the computation time of our method and a number of approaches testing on the FRGC v2.0 database using first vs. rest experiment. For each method, we report the time required for preprocessing and recognition of a single probe using a gallery of 466 subjects of the FRGCv2.0 database. Also, we report the rank-one recognition rates of different methods. From the table, we could find that, our method achieves comparable accuracies on the first vs. rest experiment. The main reason comes from that, when the FRGC v1 is used for training, the large differences of the distributions of the samples between the training and testing databases.

The computation time of our approach was tested on a PC with Intel Core i5 CPU of 3.2GHz and running in Matlab. In our case, it takes 5.5s for preprocessing and 1.46s to match a probe feature and all the 466 gallery features. Thus, totally, our approach only takes 6.96 s to identify a single probe in a gallery of 466 peoples. From the table, we can find that our approach runs much faster than all the approaches except [22, 23]. Spreeuwers *et al.* [23] got the best performance on first vs. rest experiment, but their method

seems did not work very well on large gallery as all vs. all experiment. Li *et al.* [22] used W-SRC for recognition. Both methods use Bosphorus database for training, we know that Bosphorus database contains expressive facial scans, but FRGC v1.0 database which we used for training only contains neutral facial scans.

From Tables 7 and 8, we also find that, some methods which perform excellent on first vs. rest experiment, but perform average on all vs. all experiment, while some approaches did opposite. Our approach gets competitive scores on both experiments.

Table 9 shows the rank-one scores for the state-of-the-art methods using the Bosphorus and BU_3DFE database, as reported in the literature. Overall, our approach achieves very competitive recognition rates on the Bosphorus, and BU_3DFE databases. We plan to investigate pose correction method which could work on big pose and learned fusion weights, and expect further improvement of the recognition scores.

Table 7. Comparison of verification rates including state-of-the-art methods at FAR=0.001.

Approaches	neutral vs. neutral	neutral vs. expression	all vs. all	ROCI
Maurer <i>et al.</i> [14]	97.8%	–	87%	–
Cook <i>et al.</i> [15]	–	–	92.31%	92.01%
Faltemier <i>et al.</i> [19]	–	–	93.2%	94.8%
Mian <i>et al.</i> [17]	–	–	86.6%	–
Alyuz <i>et al.</i> [9]	–	–	–	85.64%
Zhang <i>et al.</i> [8]	98.3%	89.5%	–	–
Wang <i>et al.</i> [5]	–	–	98.13%	98.04%
Berretti <i>et al.</i> [1]	97.7%	91.4%	81.2%	–
Passalis <i>et al.</i> [20]	–	79.4%	–	–
Husken <i>et al.</i> [21]	–	–	–	86.9%
Kakadiaris <i>et al.</i> [13]	–	–	97.2%	97%
Queirolo <i>et al.</i> [18]	–	–	–	96.6%
Mohammadzade <i>et al.</i> [7]	–	–	99.6%	99.2%
Spreeuwens <i>et al.</i> [23]	–	–	94.6%	94.6%
Ocegueda <i>et al.</i> [24]	–	–	97.1%	96.8%
Smeets <i>et al.</i> [26]	–	–	78.97%	77.2%
Our method	99.7%	98.38%	99.57%	98.46%

Table 8. Computation time for preprocessing and identification of a single probe using a gallery of 466 subjects of the FRGCv2 database.

Approaches	Training set	preprocessing	Total times	Rank-one
Queirolo <i>et al.</i> [18]	–	–	1864	98.4%
Faltemier <i>et al.</i> [19]	FRGC v1	7.5	1312	97.2%
Alyuz <i>et al.</i> [9]	FRGC v1	131	131	97.5%
Mian <i>et al.</i> [17]	FRGC v1	4	50.6	96.5%
Kakadiaris <i>et al.</i> [13]	–	15	15.5	97%
Li <i>et al.</i> [22]	Bosphorus	3.05	3.55	96.3%
Spreeuwens <i>et al.</i> [23]	Bosphorus and 3DFace	2.5	2.5	99%
Our method	FRGC v1	5.5	6.96	96.5 %

Table 9. Comparisons of the rank-one scores on Bosphorus databases and BU-3DFE databases.

Approaches	Training set	Bosphorus	BU_3DFE(2400/100)
Ocegueda <i>et al.</i> [25]	FRGC v1	96.4%(2797/105)	98.3%
Alyuz <i>et al.</i> [9]	Bosphorus	98.2%(2814/105)	–
Smeets <i>et al.</i> [26]	–	93.7%(4561/105)	–
Li <i>et al.</i> [27]	–	94.1%(4561/105)	–
Li <i>et al.</i> [22]	BU_3DFE/Bosphorus	95.4%(2797/105)	92.21%
Our approach	FRGC v1	91.1%(4561/105)	95.38%

6. CONCLUSIONS

We presented a fully automatic 3D face recognition algorithm and demonstrated its performance on the BU_3DFE, Bosphorus, and FRGC v2.0 databases. In this study, we introduced the spherical vector norms map (SVNs map), which can maximize between-class variability while minimizing within-class variability of 3D face effectively.

We also showed that a successful application of a new partitioning method of Histograms of Oriented Gradients to a spherical vector norms map, which could approve the 3D face recognition rate. The proposed partitioning, which maintains complete regions of the eyes and nose, could produce better performance than other partitioning strategies.

In this paper, 3D cloud data are used to establish a depth image, while the norm vectors of the cloud data are used to create an SVNs map. Then, HOG descriptors of the two images are fused for 3D face recognition. We observed that using the proposed fusion method which combines the depth image and the SVNs map provides significant improvement to 3D face recognition.

REFERENCES

1. S. Berretti, A. D. Bimbo, and P. Pala, “3D face recognition using iso-geodesic stripes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, 2010, pp. 2162-2177.
2. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, “Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, 2007, pp. 640-649.
3. H. Drira, B. B. Amor, A. Srivastava, M. Daoudi, and R. Slama, “3D face recognition under expressions, occlusions and pose variations,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, 2013, pp. 2270-2283.
4. B. Amberg, R. Knothe, and T. Vetter, “Expression invariant 3D face recognition with a morphable model,” in *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, 2008, pp. 1-6.
5. Y. Wang, J. Liu, and X. Tang, “Robust 3D face recognition by local shape difference boosting,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, 2010, pp. 1858-1870.
6. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. fisherface:

- recognition using class special linear projection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, 1997, pp. 711-720.
7. H. Mohammadzade and D. Hatzinakos, “Iterative closest normal point for 3D face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, 2013, pp. 381-397.
 8. G. Zhang and Y. Wang, “Robust 3D face recognition based in resolution invariant features,” *Pattern Recognition Letters*, Vol. 32, 2011, pp. 1009-1019.
 9. N. Alyuz, B. Gokberk, and L. Akarun, “Regional registration for expression resistant 3-D face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 5, 2010, pp. 425-440.
 10. X. Wang, Q. Ruan, Y. Jin, and G. An, “Three dimensional face recognition under expression variation,” *EURASIP Journal on Image and Video Processing*, Vol. 2014, 2014, pp. 1-11.
 11. P. J. Phillips, P. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, “Overview of the face recognition grand challenge,” in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2005, pp. 947-954.
 12. C. Yang and G. Medioni, “Object modeling by registration of multiple range images,” *Image and Vision Computing*, Vol. 10, 1992, pp. 145-155.
 13. I. A. Kakadiaris, G. Passalis, and G. Toderici, “Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 29, 2007, pp. 640-649.
 14. T. Maurer, D. Guigonis, and I. Maslov, “Performance of Geometrix ActiveID TM 3D face recognition engine on the FRGC data,” in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2005, p. 154.
 15. J. Cook, M. Cox, and V. Chandran, “Robust 3D face recognition from expression categorisation,” in *Proceedings of IEEE International Conference on Advances in Biometrics*, 2007, pp. 271-280.
 16. N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886-893.
 17. A. Mian, M. Bennamoun, and R. Owens, “An efficient multimodal 2D-3D hybrid approach to automatic face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol 29, 2007, pp. 1927-1943.
 18. C. C. Queirolo, L. Silva, O. R. P. Bellon, and M. P. Segundo, “3D face recognition using simulated annealing and the surface interpenetration measure,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, 2010, pp. 206-219.
 19. T. C. Faltmier, K. W. Bowyer, and P. L. Flynn, “A region ensemble for 3-D face recognition,” *IEEE Transactions on Information Forensics Security*, Vol. 3, 2008, pp. 62-73.
 20. G. Passalis, I. Kakadiaris, and T. Theoharis, “Evaluation of 3D face recognition in the presence of facial expressions: an annotated deformable model approach,” in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2005, pp. 171-203.

21. M. Husken, M. Brauckmann, and S. Gehlen, "Strategies and benefits of fusion of 2D and 3D face recognition," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2005, p. 174.
22. H. Li, D. Huang, J. Morvan, L. Chen, and Y. Wang, "Expression-robust 3D face recognition via weighted sparse representation of multi-scale and multi-component local normal patterns," *Neurocomputing*, Vol. 14, 2014, pp. 179-193.
23. L. Spreeuwers, "Fast and accurate 3D face recognition," *International Journal of Computer Vision*, Vol. 3, 2011, pp. 389-414.
24. O. Ocegueda, S. K. Shan, and I. A. Kakadiaris, "Which parts of the face give out your identity?" in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2011, pp. 641-648.
25. O. Ocegueda, G. Passalis, and T. Theoharis, "UR3D-C: Linear dimensionality reduction for efficient 3D face recognition," in *Proceedings of IEEE International Conference on Biometrics*, 2011, pp. 1-6.
26. D. Smeets, J. Keustermans, D. Vandermeulen, and P. Suetens, "Meshsift: local surface features for 3d face recognition under expression variations and partial data," *Computer Vision and Image Understanding*, Vol. 117, 2012, pp. 158-169.
27. H. Li, D. Huang, P. Lemaire, J. Morvan, and L. Chen, "Expression robust 3D face recognition via mesh-based histograms of multiple order surface differential quantities," in *Proceedings of IEEE International Conference on Image Processing*, 2011, pp. 3053-3056.



Xue-Qiao Wang (王雪娇) received the Ph.D. degree from Beijing Key Laboratory of Advanced Information Science and Network Technology at Beijing Jiaotong University, Beijing, China, in 2016. She now works at Beijing Union University. Her research interests are pattern recognition and computer vision.



Jia-Zheng Yuan (袁家政) received his Ph.D. degree in Computer Science Department from Beijing Jiaotong University in 2007. He is now a Professor, Software Engineering of Beijing Union University in Beijing of China, his current research interests include graph and image processing, machine learning and artificial intelligence.



Qing Li (李青) received the Ph.D. degree from the State Key Laboratory of Virtual Reality Technology and Systems at Beihang University, Beijing, China, in 2014. She now works at Beijing Union University. Her research interests are computer vision and visual computing.