Learning Weighted Sparse Representation of Encoded Facial Normal Information for Expression-Robust 3D Face Recognition

Huibin Li[1,2], Di Huang[1,2], Jean-Marie Morvan[1,3,4], Liming Chen[1,2]

[1]Université de Lyon, CNRS, [2]Ecole Centrale de Lyon, LIRIS UMR5205, F-69134, Lyon, France

[3]Université Lyon 1, Institut Camille Jordan, 43 blvd. du 11 Nov. 1918, F-69622 Villeurbanne - Cedex, France

[4]King Abdullah University of Science and Technology, GMSV Research Center, Bldg 1, Thuwal 23955-6900, Saudi Arabia

{huibin.li,di.huang,liming.chen}@ec-lyon.fr

morvan@math.univ-lyon1.fr

Abstract

This paper proposes a novel approach for 3D face recognition by learning weighted sparse representation of encoded facial normal information. To comprehensively describe 3D facial surface, three components, in X, Y, and Z-plane respectively, of normal vector are encoded locally to their corresponding normal pattern histograms. They are finally fed to a sparse representation classifier enhanced by learning based spatial weights. Experimental results achieved on the FRGC v2.0 database prove that the proposed encoded normal information is much more discriminative than original normal information. Moreover, the patch based weights learned using the FRGC v1.0 and Bosphorus datasets also demonstrate the importance of each facial physical component for 3D face recognition.

1. Introduction

In recent years, 3D face recognition technologies have achieved considerable progress mainly in: i) automatic and accurate facial landmark localization under expression, pose and occlusion variation [25]; ii) efficient face registration algorithm for pose normalization [34]; iii) discriminative 3D facial shape descriptors [15] [14]; iv) robust models or learning strategies to reduce the influences caused by facial expression variations [9] [19] [32]; v) high performance and low computational cost [29]. However, all these issues need to be improved to meet the requirements in the real world.

Focusing on issues iii and iv, this paper explores a discriminative facial shape descriptor and investigates an expression-robust algorithm by learning strategy. To address issue iii, we propose a novel facial shape descriptor, named multi-scale local normal patterns (MS-LNPs), to represent local shape variations by encoding their normal

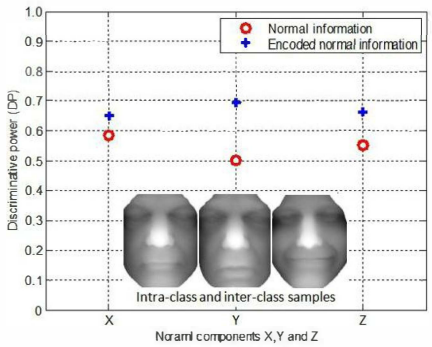

Figure 1. A simple example illustrating that the encoded normal information is more discriminative than the original normal information. The three scans (rang images), from left to right: a gallery face of subject A, a probe face of subject A, a gallery face of subject B.

information. As we known, curvatures and shape index [15] are widely used for facial surface characterization. However, surface normal, which determines (at each point) the orientation of the facial surface, has not been well discussed. To the best of our knowledge, Abate et al. [2] [3] [4] [1] introduced normal map to describe facial surface but this direct use of normal information in the holistic way did not achieve a satisfying result compared with the state of the art. Kakadiaris et al. extracted wavelet coefficients from normal and geometry maps, and reported a rank one recognition rate of 97% on the FRGC v2.0 database, however, the wavelet transform is computationally expensive [16]. Inspired by the competitive performance and computational efficiency of local descriptors, such as LBP [22] [5], SIFT [18], and more recently DAISY [31], we propose to encode original normal information, X,Y and Z components, in a local manner to generate histograms of Local Normal Patterns (LNPs), similar to the way LBP does for texture description. The idea behind it is that different shapes can

978-1-4577-1359-0/11/$26.00 ©2011 IEEE

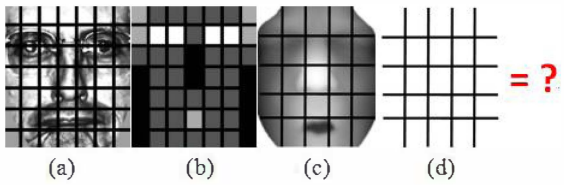

Figure 2. (a) and (b), a 2D facial image divided into patches, and the patch weights for recognition [5], (c) and (d), a 3D facial surface divided into patches, and the patch weights for recognition are unknown.

be described by different LNPs. As illustrated by a simple example in Fig. 1, the discriminative power is defined as: DP = similarity(intra-class)/similarity(inter-class); a bigger value of DP indicates a stronger discriminative power. The proposed feature (marked in blue cross) increases DP and thereby improves the discriminative power of original normal information (marked in red circle).

To pursue expression-robust 3D face recognition, some works dedicated to choose rigid facial regions, such as nose region [10], since they are expected to change less when facial expression appears. However, there has been no study to show a large subset of the face that is perfectly shape invariant across all facial expressions, meanwhile, the useful information conveyed in non-rigid facial areas is ignored. Some methods tried to model a virtual face to improve the discrimination of non-rigid regions by distorting the shape of entire face region, but it also changed the rigid parts, leading to the loss in discriminative power [9]. All the above facts demonstrate that it is not so straightforward to segment rigid facial parts from non-rigid ones, and non-rigid areas still contain the information which is important for identification. In our view, a better alternative to solve this problem is finding the average quantification weights for all facial regions or facial physical component such as eyes, nose, mouth, etc. according to their importance in 3D face recognition. As shown in Fig.2 a) and b), the quantification weights of local patches for 2D face recognition has been investigated [5], and in this paper, we present their corresponding weights in the viewpoint of machine learning and demonstrate their effects for expression-robust 3D face recognition.

Our contributions are: i) propose a novel compact facial surface descriptor named MS-LNPs (multi-scale local normal patterns); ii) Find out the average quantification weights of local patches for expression-robust 3D face recognition.

The rest of the paper is organized as follows. In section 2, we present the construction of the proposed LNPs (local normal patterns) descriptor, including normal estimation and encoding. Section 3 introduces the weighted sparse representation classifier. Section 4 shows the experimental settings and experimental results. Finally, we conclude the paper in section 5.

2. Local normal patterns facial descriptor

2.1. Normal estimation

Existing normal estimation methods can be roughly classified into optimization-based methods and averaging methods [17]. Base on the format of face data used in this paper, we employ optimization-based local plane fitting method. A 3D facial surface can be represented by a set of n points written as a $n \times 3$ matrix: $P = [p_1, p_2, ..., p_n]^T$, where $p_i = [p_{ix}, p_{iy}, p_{iz}]^T, i = 1, 2, \ldots, n$; denotes the 3D coordinates of measured points. For each point p_i, its normal vector $n_i = [n_{ix}, n_{iy}, n_{iz}]^T$ can be estimated from its neighborhood consisting of m points $Q_i = \{q_{i1}, q_{i2}, ..., q_{im}\}, q_{ij} \in P, q_{ij} \neq p_i$. We refer to the $m \times 3$ matrix $\mathbf{Q}_i = [q_{i1}, q_{i2}, ..., q_{im}]^T$ and the $(m+1) \times 3$ matrix $\bar{\mathbf{Q}}_i = [p_i, q_{i1}, q_{i2}, ..., q_{im}]^T$ represent the neighborhood points Q_i and its extended neighborhood points $\bar{\mathbf{Q}}_i$ containing p_i. Estimate normal vector n_i can be modeled by solving the optimization problem:

$$\min_{n_i} J(p_i, Q_i, n_i). \tag{1}$$

Here we choose the cost function $J(p_i, Q_i, n_i)$ as the distance of neighborhood points Q_i to a local plane $S_i = n_{ix}x + n_{iy}y + n_{iz}z + d$, i.e. solve

$$\min_{n_i, d} ||\bar{\mathbf{Q}}_i n_i - d\mathbf{1}_{m+1}||_2, \tag{2}$$

where $\mathbf{1}_{m+1}$ represents an $(m+1) \times 1$ vector of ones. This optimization problem can be solved by linear least square method. Since current face models are usually saved as range images, it is easy to choose the neighborhood points and in this paper, we utilize the neighborhood points in a 5×5 window as did in [13], i.e. $Q_i = \{q_{i1}, q_{i2}, \ldots, q_{i24}\}$. The estimated normal vectors are then normalized to unit vectors.

2.2. Normal encoding with local normal patterns

Inspired by the discriminative power and computational simplicity of Local Binary Patterns (LBP) to describe 2D texture, we encode each normal component, X, Y and Z respectively as LNPs (local normal patterns). Given a scan face model represented by an $m \times n \times 3$ matrix as follows,

$$\mathbf{P} = \{p_{ij}(x, y, z)\}_{m \times n} = \{p_{ijk}\}_{m \times n \times \{x,y,z\}}, \tag{3}$$

where $p_{ij}(x, y, z) = [p_{ijx}, p_{ijy}, p_{ijz}]^T$ represents the 3D coordinates of the point p_{ij}. Its unit normal vector matrix $(m \times n \times 3)$ is

$$\mathbf{N}(\mathbf{P}) = \{n(p_{ij}(x, y, z))\}_{m \times n} = \{n_{ijk}\}_{m \times n \times \{x,y,z\}}, \tag{4}$$

where $n(p_{ij}(x, y, z)) = [n_{ijx}, n_{ijy}, n_{ijz}]^T$ denotes the unit normal vector of p_{ij}. Further, each normal component can

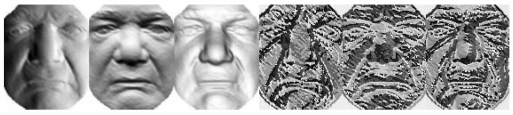

Figure 3. Left three, images of normal components X, Y and Z; right three, their corresponding encoded normal information represented as images (using the neighborhood $Q_{1,8}$).

be represented by an $m \times n$ matrix:

$$\mathbf{N}(\mathbf{P}) = \begin{cases} \mathbf{N}(\mathbf{X}) = \{n_{ijx}\}_{m\times n}, \\ \mathbf{N}(\mathbf{Y}) = \{n_{ijy}\}_{m\times n}, \\ \mathbf{N}(\mathbf{Z}) = \{n_{ijz}\}_{m\times n}. \end{cases} \quad (5)$$

where $-1 \leq n_{ijk} \leq 1, k \in \{x, y, z\}$.

This kind of matrix form, which is convenient for us to locate the neighborhood of each normal component of any point p_{ij} for the following encoding step, and the neighborhood of 3D point $Q(p_{ij})$ can be in the same way as pixels in 2D images. Specifically, the value of every point in each normal component i.e. X, Y, and Z, is compared with its neighbors in a pre-defined neighborhood. A local neighborhood is defined as a set of sampling points evenly spaced on a circle which is centered at the pixel to be labeled, and the sampling points that do not fall within the pixels are interpolated using bilinear interpolation, thus allowing for any radius and any number of sampling points in the neighborhood. Fig.4 shows two examples of the neighborhood of LNPs, where the notation $Q_{n,m}$ denotes a neighborhood of m sampling points on a circle of radius of n. After subtracting the center pixel value, the resulting strictly negative values are encoded with 0 and the others with 1; a binary number is thus obtained by concatenating all these binary codes in a clockwise direction starting from the top-left one and its corresponding decimal value is used for labeling. The derived binary numbers are referred to as Local Normal Patterns. Formally, given a point p_{ij}, its normal component noted as $n_{ijk}(0)$, the derived LNPs decimal value is:

$$LNPs(Q_{n,m}(p_{ij})) = \sum_{q=1}^{m-1} t(n_{ijk}(q) - n_{ijk}(0))2^q, \quad (6)$$

where $t(x) = 1$, if $x \geq 0$ and $t(x) = 0$, if $x < 0$.

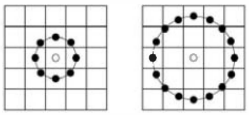

Figure 4. Examples of the neighborhood of LNPs: $Q_{1,8}$ and $Q_{2,16}$.

$LNPs(Q_{n,m})$ encodes local normal variations of each normal component as decimal value, noted by $e(\{n_{ijk}\}_{m\times n}), k \in \{x, y, z\}$. See Fig.3 for an example of $LNPs(Q_{1,8})$ on three facial normal components of same subject. It extracts the differential structure at point level. In order to describe a local shape region, histogram statistic is introduced as facial feature vector. For a given normal component $k \in \{x, y, z\}$, the histogram of encoded normal component $e(\{n_{ijk}\}_{m\times n})$ can be defined as:

$$H = \sum_{i,j} I\{e(\{n_{ijk}\}_{m\times n}) = r\}, r = 0, \ldots, R-1, \quad (7)$$

where R is the encoded decimal number, for $Q_{1,8}, R = 2^8 = 256$. $I\{A\} = 1$, if A is ture, else $I\{A\} = 0$. This histogram contains the local micro-patterns of normal component over the whole face model.

2.3. LNPs Based Facial Representation

To utilize spatial information of facial shape, each facial normal component, X, Y and Z, can be further divided into several patches, from which local normal patterns histograms H are extracted; then concatenated by facial configuration to form a global histogram G to represent the facial normal (see Fig.5). Finally, the original facial surface is described by three global feature histograms G_X, G_Y, and G_Z.

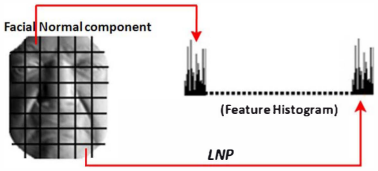

Figure 5. LNPs based facial representation.

3. Weighted sparse representation classifier

Based on the model of face subspace claiming that a well-aligned frontal face image under different lighting conditions and various facial expressions lies close to a special low-dimensional linear subspace spanned by sufficient training samples from the same subject, J.Wright et.al [33] modeled the face recognition problem by solving the l_1 minimization sparse representation and proposed sparse representation-base classification (SRC) classifier.

Analogously, we assume that a well-aligned frontal test 3D face model represented by a feature vector under different facial expressions approximately lies in a linear subspace spanned by the 3D faces in the training set (represented by the same type of facial features) associated with the same subject.

That is, given n_i training samples of i-th subject, $A_i = [v_{i,1}, v_{i,2}, \ldots, v_{i,n_i}] \in \mathbb{R}^{m\times n_i}$, any test sample $y \in \mathbb{R}^m$ from the same subject can be represented by:

$$y = \alpha_{i,1}v_{i,1} + \alpha_{i,2}v_{i,2} + \ldots + \alpha_{i,n_i}v_{i,n_i}, \quad (8)$$

where $\alpha_{i,j} \in \mathbb{R}, j = 1, 2, \ldots, n_i$.

However, unlike 2D face images with different illumination conditions, the only difference between two well-aligned 3D face models from the same subject is the local shape distortion caused by expression variations. To fairly compare the result with the state of the art, SRC is modeled with one training model ($n_i = 1$) from each subject by introducing an additional error term $\varepsilon \in \mathbb{R}^m$. Model (8) can be modified as:

$$y \approx \alpha_i v_i = \alpha_i v_i + \varepsilon, \tag{9}$$

where $y \in \mathbb{R}^m$, $v_i \in \mathbb{R}^m$ and $\alpha_i \in \mathbb{R}$ represent a probe face, a gallery face from the same subject and their linear scalar factor respectively.

Recall that in this paper face models are divided into K different regions, the feature vector v_i hence can be rewritten as $v_i = [v_{11}; v_{12}; \ldots; v_{1K}]$, where $v_{ik} \in \mathbb{R}^{(m/K)\times 1}, k = 1, 2, \ldots, K$, according to the MATLAB convention:

$$[x_1; x_2] \doteq \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

To reduce the model error ε in (9), we learned different weights $w_i, (i = 1, 2, \ldots, K)$ for different regions based on their rank-one recognition rates. Then, we have:

$$W(v_i) = [w_1 v_{11}; w_2 v_{12}; \ldots; w_K v_{1K}], \tag{10}$$

and

$$W(y) = [w_1 y_1; w_2 y_2; \ldots; w_K y_K]. \tag{11}$$

(9) can be rewrite as

$$W(y) \approx \alpha_i W(v_i) = \alpha_i W(v_i) + \tilde{\varepsilon}, (||\tilde{\varepsilon}||_2 \leq ||\varepsilon||_2). \tag{12}$$

Considering the whole gallery set with n 3D faces, each of which belongs to one subject, $W(A) \doteq [W(v_1), W(v_2), \ldots, W(v_n)] \in \mathbb{R}^{m\times n}$ and any probe $W(y) \in \mathbb{R}^m$, (12) can be rewritten as

$$W(y) = W(A)\mathbf{x} + \tilde{\varepsilon}, \tag{13}$$

where $\mathbf{x} = [0, \ldots, 0, \alpha_i, 0, \ldots, 0]^T \in \mathbb{R}^n$ is the coefficient vector whose entries are zero except the one associated with the i-subject. we named (13) as weighted sparse representation which is equivalent to solve the following l_1 minimization problem:

$$\hat{x}_1 = argmin_{\mathbf{x}} ||\mathbf{x}||_1 \text{s.t.} ||W(A)\mathbf{x} - W(y)||_2 \leq ||\tilde{\varepsilon}||_2, \tag{14}$$

We employ the OMP [23] algorithm to solve (14) and compute the residuals:

$$r_i(W(y)) = ||W(y) - W(A)\delta_i(\hat{x}_1)||_2, i = 1, 2, \ldots, n. \tag{15}$$

where δ_i is a characteristic function which selects coefficient associated with i-th gallery. Finally, the index of minimal $r_i(W(y))$ corresponding to identity of y. we call this sparse representation classifier enhanced by introducing spatial weights as W-SRC in the subsequent.

4. Experimental results

4.1. Databases and preprocessing

In our experiments, three datasets are used: FRGC v1.0 [26] and Bosphorus [28] for learning weights and FRGC v2.0 [26] for evaluation. The FRGC v1.0 dataset (Spring2003) consists of 943 textured 3D face models of 275 subjects with neutral expression. The FRGC v2.0 dataset (Fall2003 and Spring2004) is made up of 4007 textured 3D face models of 466 subjects with different facial expressions. The Bosphorus dataset contains 4666 textured 3D face models of 105 subjects in various facial expressions, pose and occlusion conditions. The 3D face models in all these three datasets are displayed in the form of range images, with a resolution of 640×480 for FRGC v1.0 and FRGC v2.0, and 1600×1200 for Bosphorus. The x, y, and z-coordinates of each 3D face model are containing in three matrices respectively. All the face models were preprocessed using the tool developed by P. Szeptycki [30], containing removing spikes and noise, filling holes, 3D nose tip detection, face cropping (for FRGC datasets). Then an ICP [34] based fine registration was employed to correct pose variations.

4.2. Experimental Settings

To evaluate the proposed approach, five experiments were designed for face recognition on FRGC v 2.0 dataset, including the effectiveness of facial features, classifier performance, the importance of different local patches, the comparison with the state of the art, as well as the robustness analysis to facial expression variations. The first scan from each subject was used to make a gallery of 466. The remaining 3D face scans were treated as probes. Before encoding the normal information, three normal components n_x, n_y and n_z matrices are resized as 120×96 respectively. Each matrix is divided into 10×8, 6×6 and 3×3 windows corresponding to sizes of 12×12, 20×16 and 40×32 local patches for the operators $Q_{1,8}$, $Q_{2,16}$ and $Q_{3,24}$ respectively, and their similarity measurements were finally combined to achieve a multi-scale based accuracy. Similar to LBP, in order to reduce the dimensionality of final facial features, the uniform pattern strategy [22] was adopted to decrease the number of bins in each local patch.

To highlight the importance of different patches for identification, 838 face modes of 267 subjects in the FRGC v1.0 dataset were selected to learn the weights of patches. At the same time, 2909 face models of 105 subjects without occlusions and rotations were chosen from the Bosphorus dataset for comparison since it contains informative expression variations while FRGC v 1.0 does not. To solve (14), OMP algorithm with sparse number L = 30 was used.

4.3. Experimental results

4.3.1 Experiment I: Comparison of facial features

To test the effectiveness of the proposed LNPs based facial feature, we compared it with two kinds of facial features: i) the original normal information based facial features NX, NY and NZ, which were achieved simply by stacking the columns of each normal component matrices n_{ijx}, n_{ijy} and n_{ijz} respectively, and their fusion NXYZ. ii) Local shape feature achieved by using LBP operator, $LBP_{2,16}$. For a fair comparison, LNPs descriptor used the same parameter with LBP to extract feature vector on each normal component, noted as LNPX, LNPY and LNPZ, and their fusion LNPXYZ. All features were finally fed to SRC classifier [33]. When considering that the three components X, Y and Z have the same weights for face recognition, in this paper, score-level fusion using a simple sum rule was employed, other fusion rules such as learning the weights of the scores can also been used. The results are shown in Table 1, we can see that our descriptor performs much better (about 20% higher) than the original un-encoded normal feature; This given a statistical illustration for the idea shown in Fig.1. On the other side, without normal information, the result based on the range images directly encoded by LBP operator is about 5% lower than that of each encoded normal component and 10% lower than their fusion LNPXYZ($Q_{2,16}$).

Approches	Rank-one Scores
(1) NX + SRC	67.83%
(2) NY + SRC	65.62%
(3) NZ +SRC	71.63%
(4) NXYZ + SRC	73.19%
(5) $LBP_{2,16}$ + SRC	82.07%
(6) LNPX($Q_{2,16}$) + SRC	87.01%
(7) LNPY($Q_{2,16}$) + SRC	86.13%
(8) LNPZ($Q_{2,16}$) + SRC	88.43%
(9) LNPXYZ($Q_{2,16}$) + SRC	92.60%

Table 1. Comparison of rank-one recognition rate using different features on the FRGC v2.0 database

4.3.2 Experiment II: Comparison of classifiers

Normally, Chi-Square distance based classifier is more popular and efficient than other classifier especially for dealing with histogram based feature [24]. Table 2 illustrates that using our proposed features, the sparse representation-based classifier (SRC) always surpasses about 8.5% higher than the Chi-Square distance based classifier.

Approches	Rank-one Scores
(1) LNPX($Q_{2,16}$) + Chi-Square	77.36%
(2) LNPX($Q_{2,16}$) + SRC	87.01%
(3) LNPY($Q_{2,16}$) + Chi-Square	77.87%
(4) LNPY($Q_{2,16}$) + SRC	86.13%
(5) LNPZ($Q_{2,16}$) + Chi-Square	81.33%
(6) LNPZ($Q_{2,16}$) + SRC	88.43%
(7) LNPXYZ($Q_{2,16}$) + Chi-Square	82.64%
(8) LNPXYZ($Q_{2,16}$) + SRC	92.60%

Table 2. Comparison of rank-one recognition rates using different classifiers on the FRGC v2.0 database

4.3.3 Experiment III: Comparison of weights of patches associating to different learning databases

Learned from the FRGC v 1.0 and Bosphorus datasets, the weights of patches were achieved by following the steps below: i) divide each normal component into local patches; ii) extract feature vector of each patch by LNPs; iii) compute recognition rate for each patch using SRC classifier. iv) recognition rates are normalized as patch weights.

Fig. 6 shows the average quantification weights of local patches of each normal component X, Y and Z learned from Bosphorus dataset. Each normal component matrix was divided into 6×6 windows for LNPs with $Q_{2,16}$. The weights are encoded by gray values where darker ones indicate lower weights while the brighter ones indicate higher weights. For example, the rigid regions including nose, eyes and forehead (patches circled by red lines) are assigned to higher weights and they totally possess about 56% importance of the whole face area for identification. While the mouth region has only about 2.8% importance. It is worth noting that facial cheek regions (in two sides), which are usually considered as non-rigid regions, own about more than 20% importance, showing that there also exists much information which is critical to recognition conveyed in non-rigid facial regions. In addition, the weights of 3D facial regions are quite different from those of 2D based ones, especially in the nose region as compared with Fig.2(b). This kind of differences may be caused by the different data properties between 2D and 3D.

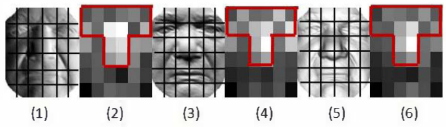

Figure 6. Images (1, 3, 5) of normal components X, Y and Z divided by 6×6 patches, images (2, 4, 6) of their corresponding weights of patches learned from Bosphorus database (LNPs $Q_{2,16}$). Darker patches indicate lower weights, while brighter ones indicate higher weights.

To evaluate the effectiveness of patch-weights, we compared the performance of un-weighted sparse representation classifier (SRC) and weighted sparse representation classifier (W-SRC) (Table 3). Table 3 presents that both F-W-SRC (learned by FRGC v1.0) and B-W-SRC (learned by Bosphorus) performance slightly better than SRC except LNPX($Q_{2,16}$) + F-W-SRC. For the final fusion score, the improvement was about 1% and 2% by F-W-SRC and B-W-SRC respectively. These results demonstrate W-SRC is efficient. On the other hand, since Bosphorus contains of informative models with different expressions it is thus more helpful than FRGC v1.0 to identify the expression models of FRGC v2.0.

Approches	Rank-one scores
(1) LNPX($Q_{2,16}$)+ SRC	87.01%
(2) LNPX($Q_{2,16}$) + F-W-SRC	86.63%
(3) LNPX($Q_{2,16}$) + B-W-SRC	88.62%
(4) LNPY($Q_{2,16}$)+ SRC	86.13%
(5) LNPY($Q_{2,16}$) + F-W-SRC	88.40%
(6) LNPY($Q_{2,16}$) + B-W-SRC	88.88%
(7) LNPZ($Q_{2,16}$) + SRC	88.43%
(8) LNPZ($Q_{2,16}$) + F-W-SRC	88.65%
(9) LNPZ($Q_{2,16}$) + B-W-SRC	90.41%
(10) LNPXYZ($Q_{2,16}$) + SRC	92.60%
(11) LNPXYZ($Q_{2,16}$) + F-W-SRC	93.59%
(12) LNPXYZ($Q_{2,16}$) + B-W-SRC	94.61%

Table 3. Comparison of rank-one recognition improvements on the FRGC v 2.0 database using patch weights learned from FRGC v1.0 (F-W-SRC) and Bosphorus (B-W-SRC) databases respectively.

4.3.4 Experiment IV: Comparison to the State-of-art

In Table 4, MS-LNPs was achieved by combining the results of three scales: $Q_{1,8}$, $Q_{2,16}$ and $Q_{3,24}$. The performance of each multi-scale normal component (13), (14) and (15) (Table 4) is better than their corresponding single scale (3), (6) and (9) (Table 3). The fusion of three multi-scale normal component (16) performances better than each single normal component (13), (14) and (15). These results prove that multi-scale LNP achieved by fusing results of different single scale is a promising way for improved performance.

Approches	Rank-one scores
(1) Abate et al. (1024 2D+3D) [1]	92.2%
(2) Chang et al. [8]	91.9%
(3) Cook et al. [11]	92.9%
(4) Mian et al. [21]	93.5%
(5) Mian et al. [20]	96.2%
(6) Mian et al. [6]	93.78%
(7) Huang et al. [15]	96.1%
(8) Huang et al. [14]	97.2%
(9) Kakadiaris et al. [16]	97.0%
(10) Faltemier et al. [12]	97.2%
(11) Alyuz et al. [7]	97.5%
(12) Queirolo et al. [27]	98.4%
(13) MS-LNPX + B-W-SRC	92.0%
(14) MS-LNPY + B-W-SRC	94.3%
(15) MS-LNPZ + B-W-SRC	94.2%
(16) MS-LNPXYZ + B-W-SRC	96.3%

Table 4. Rank-one recognition rates on the FRGC v2.0 database.

Compared with the works in the literature, the rank-one recognition rate of the proposed method (MS-LNPXYZ + B-W-SRC) outperforms (1)-(7), and it is slightly below (8)-(12). It should be emphasized that compared with the tasks (1) and (9), which made use of normal information, (1) used the difference normal maps as similarity measurement yet did not provide competitive results on the complete FRGC v2.0 dataset; while (9) used the wavelet coefficients as similarity measurement on both normal map and geometry map, and the wavelet filter lead to high computational cost. Moreover, there was no reported results only based on normal map.

4.3.5 Experiment V: Comparison of degradation influenced by facial expression

According to the experimental protocol used in [21] and [14] [15], we divided all the probe faces into two subsets based on their original labels of expressions. The first subset consists of only neutral faces, while the second is made up of only non-neutral faces. We made a fair comparison with two works in the literature, i.e. Mian et al. [21] and Huang et al. [14] [15], and the performance degradation, reflected by the difference between accuracies of subset I and II , is utilized to analyze the robustness to facial expression variations. When using the proposed normal based descriptor along with the original SRC, it achieves a 6.6% drop, which is lower than the 12.3% in [21] and as good as that in [15], showing that the proposed facial feature has a good tolerance to facial expressions. This result is further improved by the spatial weights learned from the FRGC v1.0 and Bosphorus datasets, and the weights based on the Bosphorus dataset achieves the lowest degradation of 3.8%, highlighting that the proposed weighting strategy improves the robustness to the variations of facial expression as well.

	Subset I	Subset II	Degradation
(1) Mian et.al [21]	99.0%	86.7%	12.3%
(2) Huang et.al [15]	99.1%	92.5%	6.6%
(3) Huang et.al [14]	99.0%	94.9%	4.1%
(4) MS-LNPXYZ + SRC	97.1%	90.5%	6.6%
(5) MS-LNPXYZ + F-W-SRC	97.8%	92.5%	5.3%
(6) MS-LNPXYZ + B-W-SRC	98.0%	94.2%	**3.8%**

Table 5. Comparing the degradations of rank-one score influenced by facial expression on the FRGC v 2.0 database. (Subset I: neutral probes, Subset II: non-neutral probes.)

5. Conclusions and future work

This paper presented an effective approach for 3D face recognition. We first proposed a novel 3D facial surface shape descriptor, named the Multi-Scale Local Normal Patterns (MS-LNPs), which encodes three normal components n_x, n_y and n_z as local pattern histograms at different scales. The effectiveness of the proposed shape descriptor was demonstrated by our experimental results showing that: 1) The feature extracted by LBP from original range faces are not so discriminative as local normal patterns (LNPs), which means normal information is more distinctive than range information after local encoding; 2) Original normal information is not so discriminative either, which highlight the necessity of the encoding method; 3) Sparse representation-based classifier (SRC) is more effective than the Chi-square distance based method (Experiments 2); 4) The average quantification learning based weights of different local patches are helpful to improve the final accuracy; 5) Evaluated by the complete FRGC v2.0, our method displays a rank-one recognition rate of 96.3% which is comparable to the best performance in the literature. 6) The

proposed approach is very insensitive to facial expression variations.

In our further work, we will continue to investigate other methods to encode normal information, and to explore the possible contribution of the proposed method to be combined with other existing features. Furthermore, more experimental results will be demonstrated to show the usefulness of the proposed approach in face verification.

6. Acknowledgment

This work was in part supported by the French research agency ANR through the 3D Face Analyzer project under the contract ANR 2010 INTB 0301 01 and the joint international lab LIA 2MCSI through the 3D Face Interpreter project.

References

[1] A. F. Abate, M. D. Marsico, S. Ricciardi, and D. Riccio. Normal maps vs. visible images: Comparing classifiers and combining modalities. *VLC*, 20(3):156–168, 2009. 1, 6

[2] A. F. Abate, M. Nappi, S. Ricciardi, and G. Sabatino. Fast 3d face recognition based on normal map. In *ICIP*, 2005. 1

[3] A. F. Abate, M. Nappi, S. Ricciardi, and G. Sabatino. Multimodal face recognition by means of augmented normal map and pca. In *ICIP*, 2006. 1

[4] A. F. Abate, M. Nappi, S. Ricciardi, and G. Sabatino. Fast 3d face alignment and improved recognition through pyramidal normal map metric. In *ICIP*, 2007. 1

[5] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. In *ECCV*, 2004. 1, 2

[6] F. R. Al-Osaimi, M. Bennamoun, and A. S. Mian. Integration of local and global geometrical cues for 3d face recognition. *Pattern Recognition*, 41(3):1030–1040, 2008. 6

[7] N. Alyüz, B. Gökberk, and L. Akarun. Regional registration for expression resistant 3-d face recognition. *IEEE TIFS*, 5(3):425–440, 2010. 6

[8] K. W. Bowyer, K. Chang, and P. Flynn. Adaptive rigid multi-region selection for handling expression variation in 3d face recognition. In *CVPR*, 2005. 6

[9] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Expression-invariant representations of faces. *IEEE TIP*, 16(1):188–197, 2007. 1, 2

[10] K. I. Chang, K. W. Bowyer, and P. J. Flynn. Multiple nose region matching for 3d face recognition under varying facial expression. *IEEE TPAMI*, 28(10):1695–1700, 2006. 2

[11] J. Cook, V. Chandran, and C. Fookes. 3d face recognition using log-gabor templates. In *BMVC*, 2006. 6

[12] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn. A region ensemble for 3d face recognition. *IEEE TIFS*, 3(1):62–73, 2008. 6

[13] R. Hoffman and A. K. Jain. Segmentation and classification of range images. *IEEE TPAMI*, 9(5):608–620, 1987. 2

[14] D. Huang, M. Ardabilian, Y. Wang, and L. Chen. A novel geometric facial representation based on multi-scale extended local binary patterns. In *FG*, 2011. 1, 6

[15] D. Huang, G. Zhang, M. Ardabilian, Y. Wang, and L. Chen. 3d face recognition using distinctiveness enhanced facial representations and local feature hybrid matching. In *BTAS*, 2010. 1, 6

[16] I. 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 TPAMI*, 29(4):640–649, 2007. 1, 6

[17] K. Klasing, D. Althoff, D. Wollherr, and M. Buss. Comparison of surface normal estimation methods for range sensing applications. In *ICRA*, 2009. 2

[18] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004. 1

[19] X. Lu and A. K. Jain. Deformation modeling for robust 3d face matching. *IEEE TPAMI*, 30(8):1346–1357, 2008. 1

[20] A. S. Mian, M. Bennamoun, and R. A. Owens. An efficient multimodal 2d-3d hybrid approach to automatic face recognition. *IEEE TPAMI*, 29(11):1927–1943, 2007. 6

[21] A. S. Mian, M. Bennamoun, and R. A. Owens. Keypoint detection and local feature matching for textured 3d face recognition. *IJCV*, 79(1):1–12, 2008. 6

[22] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE TPAMI*, 24(7):971–987, 2002. 1, 4

[23] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *ACSSC*, 1993. 4

[24] O. Pele and M. Werman. The quadratic-chi histogram distance family. In *ECCV*, 2010. 5

[25] P. Perakis. 3d facial landark detection and face registration. *University of Athens, Tech. Rep.*, 2011. 1

[26] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *CVPR*, 2005. 4

[27] C. Queirolo, L. Silva, O. Bellon, and M. Segundo. 3d face recognition using simulated annealing and the surface interpenetration measure. *IEEE TPAMI*, 32(2):206–219, 2010. 6

[28] A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun. Bosphorus database for 3d face analysis. In *BIOID*, 2008. 4

[29] L. Spreeuwers. Fast and accurate 3d face recognition using registration to an intrinsic coordinate system and fusion of multiple region classifiers. *IJCV*, 93(3):389–414, 2011. 1

[30] P. Szeptycki, M. Ardabilian, and L. Chen. A coarse-to-fine curvature analysis-based rotation invariant 3d face landmarking. In *BTAS*, 2009. 4

[31] E. Tola, V. Lepetit, and P. Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE TPAMI*, 32:815–830, 2010. 1

[32] Y. Wang, J. Liu, and X. Tang. Robust 3d face recognition by local shape difference boosting. *IEEE TPAMI*, 32(10):1858–1870, 2010. 1

[33] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE TPAMI*, 31(2):210–227, 2009. 3, 5

[34] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *IJCV*, 13(2):119–152, 1994. 1, 4