Expression-Robust 3D Face Recognition via Weighted Sparse Representation of Multi-Scale and Multi-Component Local Normal Patterns

Huibin Li[a,b], Di Huang[c,*], Jean-Marie Morvan[a,d,e], Liming Chen[a,b], Yunhong Wang[c]

[a] *Université de Lyon, CNRS*
[b] *Ecole Centrale Lyon, LIRIS UMR5205, F-69134, Lyon, France*
[c] *Laboratory of Intelligent Recognition and Image Processing, Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China*
[d] *Université Lyon 1, Institut Camille Jordan, 43 blvd du 11 Novembre 1918, F-69622 Villeurbanne-Cedex, France*
[e] *King Abdullah University of Science and Technology, GMSV Research Center, Bldg 1, Thuwal 23955-6900, Saudi Arabia*

Abstract

In the theory of differential geometry, surface normal, as a first order surface differential quantity, determines the orientation of a surface at each point and contains informative local surface shape information. To fully exploit this kind of information for 3D face identification, this paper proposes a novel highly discriminative facial shape descriptor, namely Multi-Scale and Multi-Component Local Normal Patterns (MSMC-LNP). Given a registered facial range image, three components of normal vectors are first estimated, leading to three normal images. Then, each normal image is encoded locally to Local Normal binary Patterns (LNP) at different scales. To utilize spatial information of facial shape, each normal image is divided into several patches, and their LNP histograms are computed and concatenated according to facial configuration. Finally, each original facial surface is represented by a set of LNP histograms including both global and local cues. Moreover, to make the proposed solution robust to the variations of facial expressions, being subtle, prototypical or exaggerated, we propose to learn the weight of each local patch under a given encoding scale and normal component. Based on the learned weights and the weighted LNP histograms, we formulate a Weighted Sparse Representation-based Classifier (W-SRC). In contrast to the overwhelming of 3D FR algorithms which were only benchmarked on the FRGC v2.0 dataset, we carried out extensive experiments on the FRGC v2.0, Bosphorus, BU-3DFE and 3D-TEC databases, thus enclosing 3D face data captured under different scenarios through various sensors and depicting in particular different challenges with respect to facial expressions. The experimental results show that the proposed approach consistently achieves competitive rank-one identification rates over those datasets despite their heterogeneous nature, and demonstrates thereby its effectiveness and its generalization skills.

Keywords: facial surface normal, local normal patterns (LNP), weighted sparse representation, 3D face recognition, identical twins

1. Introduction

Biometric systems are dedicated for identifying human beings from their own unique hard or soft physiological attributes such as iris, face, fingerprint, hand palm, hand vessel, gait, gender, *etc*. Among these attributes, face has proved to be one of the most popular and promising biometric modalities mainly due to the nature of human perception and the non-intrusiveness of face data acquisition. Although intensity image based 2D face recognition (FR) systems have provided solutions to achieve high performance under constrained conditions, the variations, especially caused by illumination and pose, which typically occur under uncontrolled environment, are still its big block [1]. With the advent of 3D sensors, it is widely expected that 3D face scans, in providing geometrical information of facial surfaces, open a new avenue to handle those unsolved issues in 2D. As such, 3D face recognition (FR) has attracted increasing attention in recent years [2, 3].

*Corresponding author. Tel: (+86)13810015904
Email addresses: huibin.li@ec-lyon.fr (Huibin Li), dhuang@buaa.edu.cn (Di Huang), morvan@math.univ-lyon1.fr (Jean-Marie Morvan), liming.chen@ec-lyon.fr (Liming Chen), yhwang@buaa.edu.cn (Yunhong Wang)

1.1. Related work

A typical 3D FR algorithm comprises the following major components although they are strongly interwoven each other [4]: 3D face landmarking, 3D face registration, the extraction of facial features along with the design of a matching scheme which closely depends upon the chosen facial features. Automatic 3D face landmarking is to automatically locate some key facial fiducial points, *e.g.*, nose tip, inner eye corners, *etc.*, which are instrumental for face cropping, face alignment and pose normalization. The most challenging issue of automatic landmarking is to tolerate the disturbance caused by arbitrary variations of facial expression, pose, or occlusion [5], and existing landmarking techniques are mainly based on the analysis of facial surface curvatures, shape index values, the facial symmetry central profile or depth information [6, 7, 8]. 3D face registration is to align 3D face scans on a common coordinate system so that the matching of facial features can be carried out in a consistent way. Popular methods for the registration of 3D face scans are ICP-based which consists of minimizing in an iterative way the distance of two 3D point clouds [9, 10] although they are reputed to be computationally expensive. The extraction of facial features is to generate a discriminating facial representation which should comprehensively describe each 3D face scan for the latter stage of matching. As all human faces are similar to each other in terms of configuration whereas a 3D face scan accurately captures the geometrical shape of the underlying 3D facial surface, thereby making it likely more sensitive to facial expressions in comparison with 2D facial images, the design of a discriminating facial feature which stays robust to facial expressions is a critical issue in 3D FR. A number of approaches has been proposed in the literature, including facial curves [11], geometry and normal maps [12], tensor based representations [13], iso-geodesic stripes [14], Multi-Scale Local Binary Pattern (MS-LBP) Depth maps and Shape Index (SI) maps [15], Multi-Scale extended Local Binary Pattern (MS-eLBP) maps [16] *etc.* Other essays try to explicitly account for facial expression variations. An original tentative was made by Bronstein *et al.* [17] who assumed that facial expressions can be modeled as isometries of the facial surface and proposed a facial expression invariant canonical form. However, their assumption proves to be inexact, especially in the presence of exaggerated facial expression [18]. A far more popular approach observes that facial expressions introduce facial distortions but there are still relatively stable facial regions, *e.g.*, forehead, nose region, from which expression robust features can be extracted [19, 20]. Chang *et al.* [21] selected three regions around the nose for 3D face matching whereas Faltemier *et al.* [9] extended the later number to 28 small regions on the face. However, automatic detection and segmentation of facial surface into rigid and mimic regions is still problematic [8, 20].

The overwhelming majority of 3D FR algorithms proposed thus far in the literature is evaluated on the FRGC v2.0 dataset [2] which has become de facto the standard benchmark for 3D FR algorithms. Very high performance, up to 99% rank-one recognition rate [22], was reported on that dataset. However, although FRGC v2.0, with its 4007 3D face scans from 466 subjects, is the largest public 3D face dataset so far known in the literature, all its scans were captured in frontal pose position under controlled lighting conditions, and less than half of them depict only a limited number of facial expressions, *e.g.*, happiness, surprise, *etc.* 3D face scans captured from uncooperative subjects in real-life applications can feature other challenges, *i.e.* missing data due to arbitrary pose, external occlusions, and uncountable other types of facial expressions, being subtle, prototypical or exaggerated. As a result, 3D FR algorithms with high performance on FRGC v2.0 can vastly degrade under other settings as revealed the recent studies on the 3D Twins Expression Challenge (3D-TEC) database [23, 24]. 3D-TEC stages a scenario of distinguishing 107 sets of identical twins through 3D face scans, each subject depicting a neutral and a smiling facial expressions. This is a very challenging scenario for 3D FR systems because of the strong similarities between the 3D facial surfaces of twins in addition to the traditional interference factors like facial expression variations. Vijayan et al. [23] evaluated the performance of four state-of-the-art 3D FR algorithms on the 3D-TEC dataset. They found that some algorithms performed very well on FRGC v2.0 but vastly degraded on 3D-TEC, especially in the joint presence of disturbing factors, *i.e.* strong inter-class facial similarities and intra-class variations of facial expressions. Their results suggest that benchmarking 3D FR algorithms on FRGC v2.0 is certainly necessary but not sufficient to ensure the same performance and robustness with respect to the challenges of real-life applications, including in particular uncountable facial expressions, leading to subtle, moderate and exaggerated facial surface deformations.

1.2. Motivations and the proposed approach

In this paper, we target the challenge of facial expressions in 3D FR and propose a discriminative facial surface representation and an expression-robust method to handle expression variations. Specifically, we propose a novel

facial shape descriptor, namely Multi-Scale and Multi-Component Local Normal Patterns (MSMC-LNP), which represents the local facial shape by encoding their three normal components: x, y, and z as binary patterns in a multi-scale way, respectively. An input facial surface can then be represented as a certain number of local normal patterns based maps or histograms of Local Normal Patterns (LNP).

As we know, surface curvatures [7, 25, 26] and shape index values [15, 27, 28] have been widely investigated for facial surface representation and characterization. However, the surface normal, which determines (at each point) the orientation of a facial surface, has not been well explored in terms of 3D face representation [1]. Abate *et al.* [31, 32, 33, 34] introduced normal maps to describe facial surfaces. But this direct use of normal information in the holistic way did not achieve satisfying results. Gokbert *et al.* [35] used surface normal variance at each pixel location as a distance measure between face images and reported a rank-one score of 87.8% on the whole FRGC v2.0 database, while this performance vastly degraded on the 3D-TEC database [23]. Kakadiaris *et al.* [12] proposed to extract wavelet coefficients from normal and geometry maps for the computation of similarity, and reported a rank one recognition rate of 97% on the FRGC v2.0 database; however, the wavelet transform along with the fitting of the annotated deformable model is quite computationally expensive. Inspired by the competitive performance and computational efficiency of local binary patterns (LBP) for texture classification and 2D FR [36, 37, 38], we propose to encode surface normal information, namely x, y, and z component normal images, in a local manner to generate histograms of LNP, in a similar way that LBP does for texture image description. The idea behind it lies in that different facial shapes can be described by different LNP under given encoding scales and normal components, which makes LNP a very discriminative descriptor to recognize 3D faces and even to distinguish identical twins.

To pursue expression-robust 3D face recognition, we have seen that the popular method consists of choosing rigid facial regions, *e.g.*, the nose and forehead regions [9, 21], for the purpose of 3D face matching. However, the segmentation of a 3D facial surface into relatively rigid and elastic regions is problematic. Furthermore, such an approach also tends to ignore the elastic facial regions which also bear significant discriminating information. In this paper, we consider another alternative and propose to find the average quantification weights for all facial regions or facial physical components, *e.g.*, eyes, nose, mouth, *etc.*, according to their discriminating power in 3D face recognition. This kind of quantification weights of local patches has been investigated in 2D FR and several works [37, 39, 40] demonstrated its effectiveness (see Fig. 1 (a) and (b) for an example). Interesting enough, to the best of our knowledge, the corresponding strategy in 3D using the weighting of facial components has not been studied yet for the purpose of expression-robust 3D FR . As shown in Fig. 1 (c) and (d), in this paper, we will show that the weights associated with 3D face components are largely different from those of their 2D counterparts, especially in the regions of nose and mouth. These weights can be learned from a given training set in the training phase (see Fig. 2). The learned patch weights then can be used to build a weighted sparse representation model and compute the weighted reconstruction errors, leading to a Weighted Sparse Representation-based Classifier (W-SRC).

The main contributions of this paper can be summarized as follows:

1. We introduce a new 3D facial representation based on Multi-Scale and Multi-Component Local Normal Patterns (MSMC-LNP) for facial shape description. MSMC-LNP describes the micro-structure of facial normal information in multiple scales and multiple normal component channels. The extensive experiments show that the proposed LNP based facial representation is more discriminative than both the raw normal information and the encoded range image, *i.e.*, Local Shape binary Patterns (LSP) [41] (see Tab. 3). We also show that the fusion of both multiple scales and multiple components is a helpful way to improve the final performance and demonstrate its competency on 3D face identification as well as the challenging issue of recognizing expressive identical twins.

2. A learning-based strategy is proposed to find out the quantification weights of local patches of 3D facial surfaces in terms of discriminating power. Given a training database, the patch weights associated with different facial regions, encoding scales, and normal components are learned by normalizing the patch matching scores. Those scores are computed by running the sparse representation-based classifier (SRC) over MSMC-LNP features of local patches. The experimental results show that the weights associated with these local patches in 3D are quite different from those of their 2D counterparts, especially in the nose region, thereby highlight the fact that 3D and 2D face data bear different facial information which can be further explored in a multimodal face recognition scenario.

[1] Note that, recently, normal constraint based surface registration for 3D FR methods such as [29, 30] have achieved very high performance on FRGC v2.0. In this paper, we focus on extracting normal based facial descriptor for 3D FR.

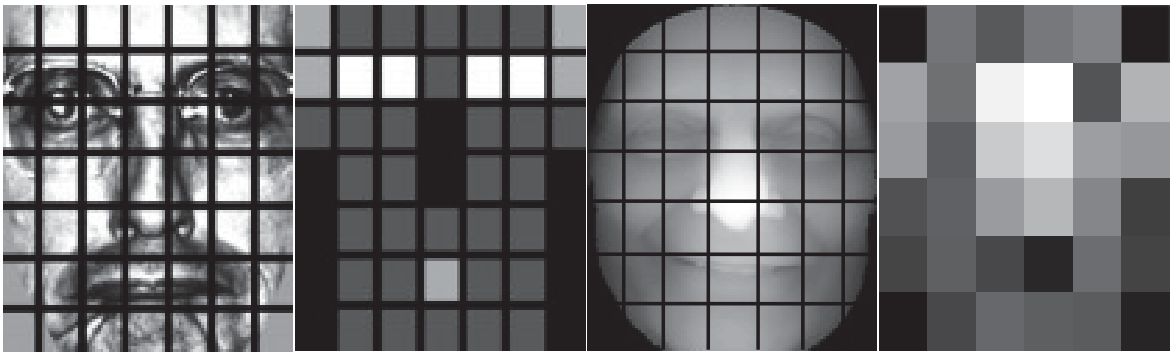


Figure 1: Illustration of patch weights for 2D and 3D face recognitions: (a-b) a 2D face image and its corresponding patch weights [37]; (c-d) a 3D face depth image and its corresponding patch weights learned by our method. All images are split to 6×6 local patches. Darker patches indicate lower weights, while brighter ones indicate higher weights.

3. Using the learned patch weights along with the proposed MSMC-LNP feature, a weighted sparse representation-based classifier (W-SRC) is formulated to account for the different sensitivities of facial components to facial expressions, making the proposed 3D FR algorithm both discriminating enough for distinguishing identical twins and robust enough to tolerate various facial expressions. In particular, we formally establish that a weighted combination of SRCs amounts to directly weighting the corresponding feature vectors of those SRCs.

4. Extensive experiments were carried out using various 3D face datasets, including FRGC v2.0, BU-3DFE, Bosphorus and 3D-TEC, to benchmark the effectiveness and the generalization skills of the proposed approach with respect to 3D face scans captured under different scenarios and conditions with different 3D sensors, depicting in particular different challenges in terms of facial expressions. This is in clear contrast with the overwhelming majority of 3D FR algorithms so far proposed in the literature which are only evaluated on FRGC v2.0 with 3D face scans depicting a limited number of facial expressions. The experimental results demonstrate the effectiveness and the generalization skills of the proposed approach which consistently displays competitive rank-1 recognition rates over those datasets of different nature.

This paper integrates our preliminary work in [42] but significantly extends that work. First, we carefully reformulate here the weighted sparse representation-based classifier (W-SRC) and formally establish that a weighted combination of SRCs amounts to directly weighting the feature vectors of those SRCs. Then, we extensively evaluate the robustness of the proposed system with respect to a rich set of facial expressions, using in particular the BU-3DFE database which encloses expressive 3D face scans depicting the six prototypical expressions in different intensities, as well as the Bosphorus database which features both the six prototypical expressions and subtle facial expressions through action units. This is in clear contrast with the overwhelming majority of 3D FR algorithms which were only benchmarked over FRGC v2.0 depicting merely a limited number of facial expressions. Finally, we validate the discriminative power of the proposed feature and the robustness of the proposed approach to facial expressions in the challenging issue of distinguishing expressive identical twins using the 3D-TEC database.

The reminder of the paper is organized as follows. The framework overview of the proposed system is presented in Section 2. Section 3 introduces the proposed Local Normal Patterns (LNP) based facial descriptor. Section 4 describes the weighted sparse representation-based classifier. In section 5, we present the experimental and algorithmic settings and discuss the experimental results. Section 6 concludes the paper.

2. Overview of the Proposed Approach

As illustrated in Fig. 2, the framework of the proposed approach consists of two phases: *i.e.* a training phase and a testing phase. Before the training and testing phases, each raw 3D face scan is preprocessed, *e.g.*, spike and noise removing, holes filling, nose tip localization, face cropping and alignment, to generate a range image with a predefined size. The training process is carried out to learn on a predefined training set the quantitative weights of

facial physical components in terms of discriminating power and robustness to facial expression variations. It includes three procedures: feature extraction, identification and score normalization within different patches. They are shortly summarized as follows:

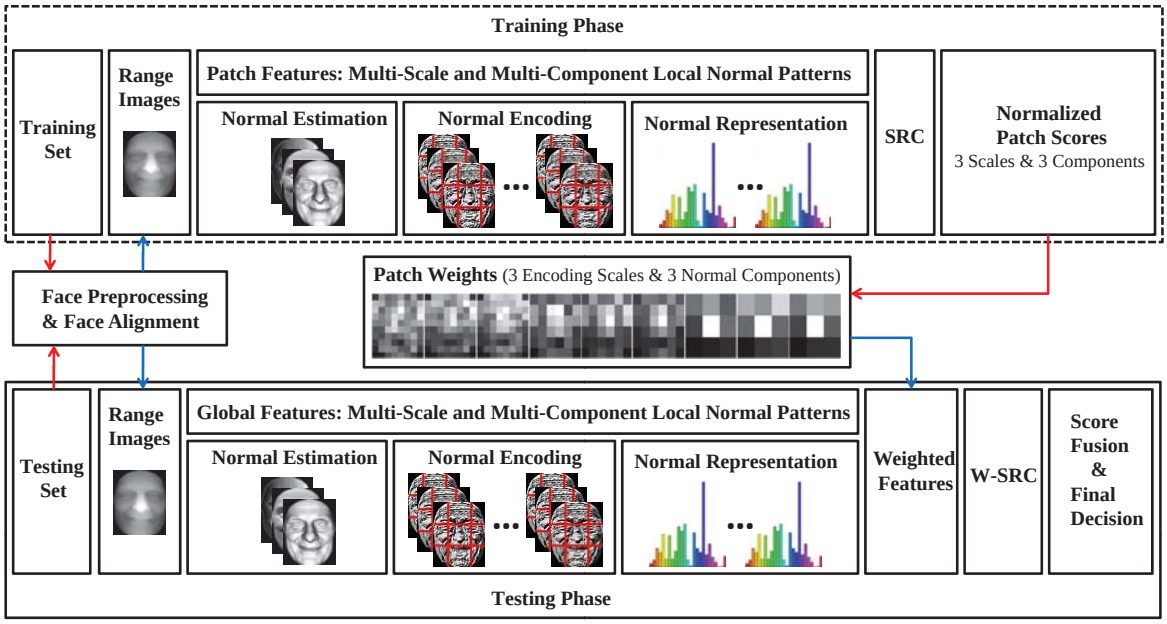


Figure 2: Overview of the proposed approach.

(1) Patch feature extraction. This procedure consists of three steps: (a) facial normal estimation; (b) facial normal encoding; (c) facial normal representation. Specifically, given a raw 3D face scan, we first launch the preprocessing pipeline (see Sec. 5.2) to normalize the range image to an $m \times n \times 3$ matrix (*i.e.*, x, y and z coordinates). Based on the range image, we estimate its three normal components (x, y, and z) by the local plane fitting method (see Sec. 3.1). Each normal component map is coarsely split into several local patches (*e.g.* 3×3); then each of these local patches is encoded as LNP with multiple scales, giving birth to multi-scale and multi-component local normal patterns (MSMC-LNP) to comprehensively describe the shape of each patch.

(2) Patch-based identification. Given a patch, an encoding scale, and a normal component, the corresponding LNP is extracted and fed into the sparse representation-based classifier (SRC) to generate a rank-one recognition rate using the training set.

(3) Patch score normalization. The patch scores, *i.e.* rank-one recognition rates, of different encoding scales and normal components are further normalized as the corresponding patch weights. The importance of facial physical regions in terms of discriminating power and robustness to facial expression variations can thus be measured by those quantitative patch weights.

During the testing phase, given a preprocessed range image in the testing set, we first compute the MSMC-LNP features over all the patches as in procedure (1) in the training phase. The global MSMC-LNP features are then obtained by simply stacking all these patch based MSMC-LNP features according to the holistic configuration of facial surfaces (see Sec. 3.3). Based on the patch weights learned in the training phase, weighted sparse representation (W-SRC) is formulated as seeking the sparse solution of the sum of the weighted patch based sparse representation (see (13) in Sec. 4). Then, W-SRC carries out face identification by finding the minimal weighted reconstruction residuals. (see (14) in Sec. 4). The final similarity measurement of MSMC-LNP for decision making is computed using the score level fusion which combines the matching scores of three encoding scales and three normal components through simple sum rule.

3. Local Normal Patterns (LNP) based Facial Descriptor

3.1. Facial Normal Estimation

To highlight local variations of facial surfaces, recall that we make use of their normal information instead of the original point-cloud or range images. Existing normal estimation methods can be roughly classified into optimization based methods (*i.e.*, local fitting methods) and averaging methods [43].

The basic idea of optimization based methods is as follows: 1) the normal vector on a given point can be calculated as the normal vector of a plane or quadratic surface which it belongs to; 2) the underlying plane or surface can be estimated by fitting it to the local neighboring points around that given point; 3) the fitting problem then can be solved by minimizing a cost functional penalizing a certain criterion, *e.g.*, the distance of the neighboring points to the searched local plane (see Fig. 3 (a)). The averaging methods estimate the normal vector of a given point by computing a weighted average of the normal vectors of the triangles in its one-ring neighbors, the weights being the inverse ratios of the areas or the surrounding angles of the triangles in its one-ring neighbors (see Fig. 3 (b)).

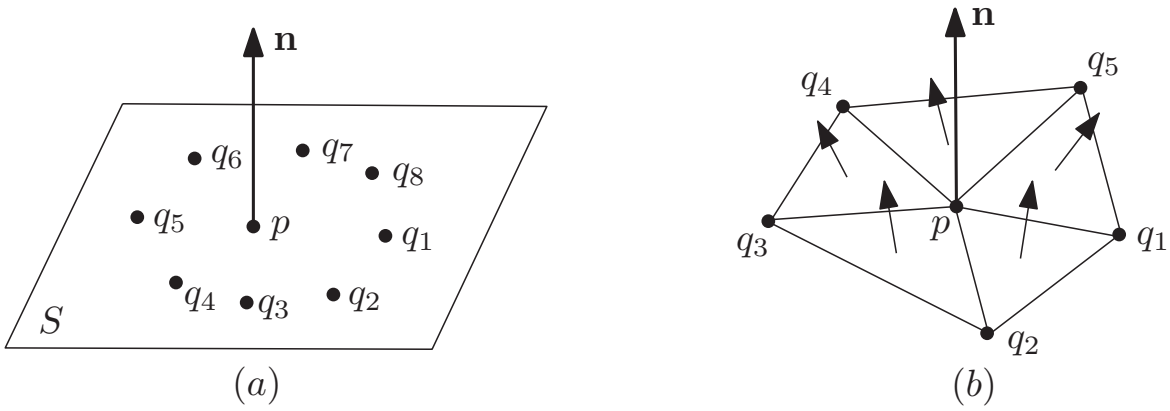


Figure 3: Illustration of two approaches for normal estimation: (a) a plane is fitted to a vertex p and its neighbors; (b) the normal vectors of triangles in one-ring of p are averaged.

The optimization-based methods can be applied to 3D point-clouds and triangular meshes while the averaging methods can only work on triangular meshes. Both types of methods are competent for normal calculation. In this paper, 3D face scans from various datasets were captured under different conditions and scenarios with divers 3D sensors, resulting in a diversity of data formats. As a result, optimization-based method is adopted for the estimation of normals.

Specifically, given a facial range image $\mathbf{P}$ represented by an $m \times n \times 3$ matrix:

$$\mathbf{P} = [p_{ij}(x, y, z)]_{m\times n} = [p_{ijk}]_{m\times n\times\{x,y,z\}}, \tag{1}$$

where $p_{ij}(x, y, z) = (p_{ijx}, p_{ijy}, p_{ijz})^T, (1 \le i \le m, 1 \le j \le n, i, j \in \mathbb{Z})$ represents the 3D coordinates of the point p_{ij}. Let its unit normal vector matrix $(m \times n \times 3)$ be

$$\mathbf{N}(\mathbf{P}) = [n(p_{ij}(x, y, z))]_{m\times n} = [n_{ijk}]_{m\times n\times\{x,y,z\}}, \tag{2}$$

where $n(p_{ij}(x, y, z)) = (n_{ijx}, n_{ijy}, n_{ijz})^T, (1 \le i \le m, 1 \le j \le n, i, j \in \mathbb{Z})$ denotes the unit normal vector of p_{ij}. As described in [44], the normal vector **N(P)** of range image **P** can be estimated using the local plane fitting method. That is to say, for each point $p_{ij} \in \mathbf{P}$, its normal vector $n(p_{ij})$ can be estimated as the normal vector of the following local fitted plane:

$$S_{ij} : n_{ijx}q_{ijx} + n_{ijy}q_{ijy} + n_{ijz}q_{ijz} = d, \tag{3}$$

where $(q_{ijx}, q_{ijy}, q_{ijz})^T$ represents any point within the local neighborhood of point p_{ij} and $d = n_{ijx}p_{ijx} + n_{ijy}p_{ijy} + n_{ijz}p_{ijz}$. In this work, a neighborhood of 5×5 window is used. To simplify, each normal component in equation (2)

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

$$\mathbf{N}(\mathbf{P}) = \begin{cases} \mathbf{N}(\mathbf{X}) = [n_{ij}^{x}]_{m\times n}, \\ \mathbf{N}(\mathbf{Y}) = [n_{ij}^{y}]_{m\times n}, \\ \mathbf{N}(\mathbf{Z}) = [n_{ij}^{z}]_{m\times n}. \end{cases} \tag{4}$$

where $\|(n_{ij}^{x}, n_{ij}^{y}, n_{ij}^{z})^{T}\|_2 = 1$.

Fig. 4 shows a sample range image extracted from the 3D-TEC database and its estimated three normal component matrices (images). As we can see from that figure, the normal component images contain more informative geometric information than their corresponding range image which is quite smooth. In particular, one can see that the geometric shape details around the eyes, mouth and forehead regions are quite well highlighted.

(a) (b) (c) (d)

Figure 4: Illustration of facial normal estimation: (a) the original range image, (b-d) its normal images of component x, y and z (the sample comes from the 3D-TEC dataset).

3.2. Facial Normal Encoding

Inspired by the discriminative power and computational simplicity of LBP for 2D texture description, we propose to encode each normal component, x, y, and z, respectively, as local normal patterns (LNP) to further highlight local shape variations. Thanks to the matrix form of these normal components as in equation (4), the encoding of these local shape variations can be carried out in a similar way as LBP encodes 2D texture images. Specifically, the value of each point in a normal component 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. 5 shows two examples of neighborhood of LNP, where the notation $Q_{n,m}$ denotes a neighborhood of m sampling points on a circle of radius of n.

After subtracting the central 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 decimal numbers which result from such a process are referred to as local normal patterns (LNP). Formally, given a point p_{ij}, its normal component noted as $n_{ij}^{k}(0)$, the derived LNP decimal value is:

$$LNP(Q_{n,m}(p_{ij})) = \sum_{q=1}^{m-1} t(n_{ij}^{k}(q) - n_{ij}^{k}(0))2^{q}, \tag{5}$$

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

Given a normal component, $LNP(Q_{n,m})$ thus encodes local normal variations around each point as a decimal value, denoted as $e([n_{ij}^{k}]_{m\times n}), k \in \{x, y, z\}$. Fig. 6 gives an example of $LNP(Q_{1,8})$ on the three facial normal components of the same subject.

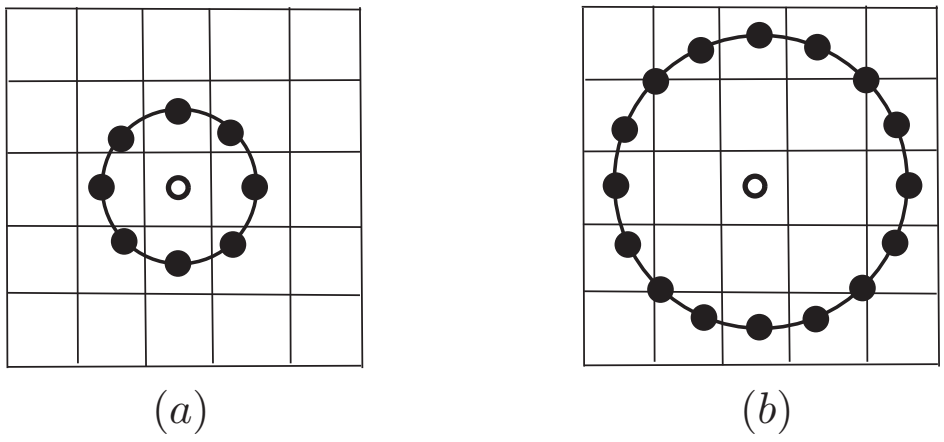


Figure 5: Examples of the neighborhood of LNP: (a) $Q_{1,8}$ and (b) $Q_{2,16}$.

Figure 6: Illustration of facial normal encoding: (a) to (c), normal images of component x, y and z; (d) to (e), their corresponding LNP maps using the neighborhood $Q_{1,8}$.

LNP encodes the local shape variations at point level. In order to characterize the shape of a local region, histogram-based statistics is computed and used as facial feature vector. For a given normal component $k \in \{x, y, z\}$, the histogram of the encoded normal component $e([n_{ij}^k]_{m\times n})$ can be defined as:

$$H = \sum_{i,j} I\{e([n_{ij}^k]_{m\times n}) = r\}, r = 0, \ldots, R-1, \tag{6}$$

where r is the encoded decimal number, and $R-1$ is the maximum value of local normal patterns (LNPs) given a neighborhood $Q_{n,m}$, *e.g.*, for $Q_{1,8}, R = 2^8 = 256$. $I\{A\} = 1$, if A is true, else $I\{A\} = 0$. This histogram describes the local micro-patterns of a given normal component over the whole face model.

3.3. Facial Normal Representation

To utilize the spatial information of 3D face scans, each facial normal component, x, y, and z, can be further divided into several patches, from which LNP histograms H are extracted and then concatenated to form a global histogram G which thus captures the facial configuration of the encoded facial normal feature (see Fig. 7). Finally, the original facial surface is described by three global feature histograms G_x, G_y, and G_z at a given encoding scale.

4. Weighted Sparse Representation-based Classifier

Once 3D face scans described by their respective facial normal representations, a proper classifier is needed for the purpose of 3D face identification. In this work, we introduce a Weighted Sparse Representation-based Classifier (W-

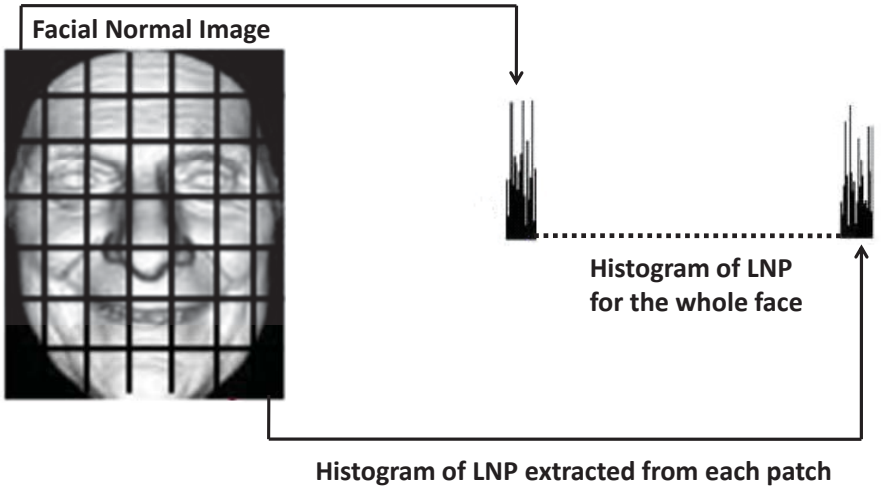


Figure 7: Illustration of facial normal representation: histogram of LNP.

SRC) which is described in this section. We first introduce 3D face subspace based sparse representation model and its corresponding SRC. Then, we theoretically formulate a weighted sparse representation model and its corresponding W-SRC to account for the patch weights which are learned using a training set to quantify the relative discriminating power of each facial component. We demonstrate in particular that solving a weighted SRC amounts to computing a single SRC with global feature vectors in stacking weighted feature vectors of each patch.

4.1. 3D Face Subspace and Sparse Representation-based Classifier

A common assumption in 2D FR is the 2D face subspace model which assumes that well-aligned frontal face images under different lighting conditions and various facial expressions, lie close to a special low-dimensional linear subspace spanned by sufficient training samples from the same subject. Using such an assumption, Wright *et al.* [45] proposed a sparse representation model and its corresponding SRC for robust 2D FR. In this work, we make an extension of this assumption to the 3D case, and assume that well-aligned frontal 3D face scans under different facial expressions approximately lie close to a special low-dimensional linear subspace spanned by sufficient training 3D face scans from the same subject. We call this assumption as 3D face subspace model. Formally, it can be formulated by the following equation:

$$y \approx \alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_n v_n. \tag{7}$$

That is, given n_i training samples of i-th subject, $[v_{i,1}, v_{i,2}, \ldots, v_{i,n_i}] \in \mathbb{R}^{m \times n_i}$,, any test sample $y \in \mathbb{R}^m$ of the i-th subject can be represented, according to (7), as:

$$y_i \approx \alpha_{i,1} v_{i,1} + \alpha_{i,2} v_{i,2} + ... + \alpha_{i,n_i} v_{i,n_i}, \tag{8}$$

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

Note that the most common experimental setting in 3D FR is that there is only one training sample per each subject in the the gallery. Therefore, without occlusion, the only difference between two well-aligned frontal 3D face scans from the same subject is the local shape distortion caused by expression variations. This problem of insufficient training samples along with the shape distortion caused by expression variations introduces a new model error term, denoted as $\varepsilon_i \in \mathbb{R}^m$. Thus, model (8) can be modified as:

$$y_i \approx \alpha_{i,1} v_{i,1} = \alpha_{i,1} v_{i,1} + \varepsilon_i, \tag{9}$$

where $y_i \in \mathbb{R}^m$, $v_{i,1} \in \mathbb{R}^m$ and $\alpha_{i,1} \in \mathbb{R}$ represent a probe face, a gallery face from the same subject and their linear scalar factor respectively.

Using (9), sparse representation model and its corresponding SRC for 3D FR can be formulated as follows. Given a gallery set with n 3D face scans, each of which belongs to one subject, we define the dictionary as $D \doteq [v_1, v_2, \ldots, v_n] \in \mathbb{R}^{m\times n}$. Then for any probe $y \in \mathbb{R}^m$, we have

$$y = Dx + \varepsilon, \tag{10}$$

where $x = [0, \ldots, 0, \alpha_j, 0, \ldots, 0]^T \in \mathbb{R}^n$ is the coefficient vector whose entries are zero except the one associated with the j-subject. Sparse coefficients x in (10) can be solved by the following l_0 minimization problem:

$$\hat{x} = \arg\min_x \|x\|_0 \text{ s.t. } \|y - Dx\|_2^2 \leq T, \tag{11}$$

where $T = \|\varepsilon\|_2^2$.

In practice, we employ Orthogonal Matching Pursuit (OMP) [46] algorithm to solve (11) and compute the reconstruction residuals:

$$r_i(y) = \|y - D\delta_i(\hat{x})\|_2^2, i = 1, 2, \ldots, n. \tag{12}$$

where δ_i is a characteristic function which selects the coefficient associated with the i-th gallery. Finally, the index of minimal $r_i(y)$ delivers the identity of the probe y.

4.2. Weighted Sparse Representation-based Classifier

Assume now that each face scan is divided into K different patches. Denote w_k as the learned weight for patch k. Using the MATLAB convention:

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

the feature vector v_i can be rewritten as

$$v_i = [v_{i1}; v_{i2}; \ldots; v_{ik}; \ldots; v_{iK}],$$

where $v_{ik} \in \mathbb{R}^{(m/K)\times 1}$, and the dictionary D can be denoted as

$$D = [D_1; D_2; \ldots; D_k; \ldots; D_K],$$

where $D_k = [v_{1,k}, v_{2,k}, \ldots, v_{i,k}, \ldots, v_{n,k}]$, and a probe y can be denoted as

$$y = [y_1; y_2; \ldots; y_k; \ldots; y_K],$$

where $y_k \in \mathbb{R}^{(m/K)\times 1}, k = 1, 2, \ldots, K$.

Eq. (11) can then be rewritten as the following weighted sparse representation model:

$$\hat{x} = \arg\min_x \|x\|_0 \text{ s.t. } \sum_{k=1}^{K} w_k \|y_k - D_k x\|_2^2 \leq T, \tag{13}$$

and the corresponding weighted reconstruction residuals is

$$r_i(y) = \sum_{k=1}^{K} w_k \|y_k - D_k\delta_i(\hat{x})\|_2^2, i = 1, 2, \ldots, n. \tag{14}$$

To solve eq. (13), we notice that it equals to solve

$$\hat{x} = \arg\min_x \|x\|_0 \text{ s.t. } \sum_{k=1}^{K} \|w_k y_k - w_k D_k x\|_2^2 \leq T. \tag{15}$$

We denote

$$W(D) = [w_1 D_1; w_2 D_2; \ldots; w_K D_K],$$

and

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

Then eq. (15) equals to

$$\hat{x} = \arg\min_x \|x\|_0 \text{ s.t. } \|W(y) - W(D)x\|_2^2 \leq T. \tag{16}$$

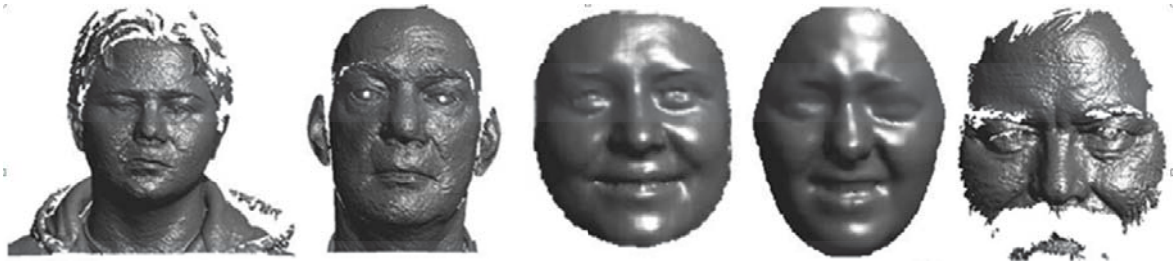
Eq. (16) means that the weighted sparse representation model as expressed in Eq. (13) amounts to solving a single SRC with global feature vectors in simply stacking weighted features of the corresponding patches. Eq. (16) can be solved by the OMP algorithm [46]. Once determined the sparse representation coefficient $\hat{x}$ of Eq. (16), weighted reconstruction residuals in Eq. (14) can be computed. Then the minimal $r_i(y)$ can be used to determine the identity of y. We call this sparse representation-based classifier enhanced by spatial weights as Weighted Sparse Representation-based Classifier (W-SRC) in the subsequent.

5. Experiments

Recent studies [23] show that 3D FR algorithms with very high performance on the de facto standard benchmark, namely FRGC v2.0, can vastly degrade when evaluated on expressive 3D face scans captured under different scenario. As a result, we decide to evaluate the effectiveness and the generalization skills of the proposed approach using various datasets, namely FRGC, BU-3DFE, Bosphorus and 3D-TEC, with 3D face scans depicting a rich set of facial expressions, being subtle, prototypical and exaggerated. Depending on the underlying dataset, those expressive 3D face scans were captured under different scenarios and conditions, *e.g.*, 3D sensors, lighting conditions, and thus offer different levels of challenge with respect to facial expression variations. In this section, we first introduce the different datasets and describe their particularity, then present the different experimental settings aiming to highlight the various facets of the proposed approach and finally discuss the experimental results.

5.1. Databases and Preprocessing

In our experiments, three databases, namely FRGC v1.0 [2], BU-3DFE [47] and Bosphorus [48], are used as training sets to learn the patch weights respectively, while four databases, the BU-3DFE, Bosphorus, FRGC v2.0 [2] and 3D-TEC [24] are used as testing sets for cross database validation and evaluation. Fig. 8 plots a raw 3D face scan from each of these databases. They are briefly introduced as follows:



(a) (b) (c) (d) (e)

Figure 8: Illustrate of the raw samples of the five databases: (a) FRGC v1.0, (b) FRGC v2.0, (c) Bosphorus, (d) BU-3DFE, (e) 3D-TEC.

- ***FRGC v1.0*:** The FRGC v1.0 database (Spring2003) consists of 943 textured 3D face models of 275 subjects with the neutral expression. The hardware used to acquire these range images is a Minolta Vivid 900 (MV 900) laser range scanner, with a resolution of 640×480.

- ***FRGC v2.0*:** The FRGC v2.0 database (Fall2003 and Spring2004) is made up of 4007 textured 3D face models of 466 subjects with different facial expressions. The same hardware as in FRGCv1.0 is used for data acquisition, and the resolution of each range image is also 640×480. FRGC v2.0 is the largest public 3D face dataset which encloses 1642 expressive frontal 3D face scans captured under controlled lighting conditions. *Although*

FRGC v2.0 has become de facto the standard benchmark for evaluating 3D FR algorithms, in particular because of its size, it only contains a very limited of facial expressions, including happiness, surprise, disgust, sadness and others. This is to be compared with uncountable number of facial expressions which can occur in real-life applications under uncontrolled conditions. As a result, a 3D FR algorithm with high performance on FRGC v2.0 does not guarantee a same level of performance on other 3D face scans captured under different conditions and scenarios. The generalization skills of a claimed robustness to facial expression variations needs to be checked on representative expressive 3D face scans.

- ***BU-3DFE*:** The BU-3DFE database contains 100 subjects (56 females and 44 males), ranging age from 18 to 70 years old, with a variety of ethnic ancestries. Each subject performs seven expressions. Except neutral, each of the six prototypic expressions (happiness, disgust, fear, angry, surprise, and sadness) includes four levels of intensity. Therefore, there are 25 instant models for each subject, resulting in a total of 2,500 3D facial models. The 3D models are captured with a 3D face imaging system named 3DMD digitizer. Each model is saved as a polygonal mesh with a resolution ranging from 20,000 to 35,000 polygons. *This dataset thus permits to benchmark 3D FR algorithms over 3D face scans depicting prototypical facial expressions of different intensities, ranging from very subtle to exaggerated, and quantitatively measures the robustness of a given 3D FR technique in varying the intensities of facial expressions.*

- ***Bosphorus*:** The Bosphorus database contains 4666 textured 3D face models of 105 subjects in various facial expressions, action units, poses and occlusions. The 3D models are acquired with a device named Inspeck Mega Capturor II (IMC II). Each model is saved as a range image with a resolution of $1,600 \times 1,200$. *This dataset thus permit to benchmarking 3D FR techniques under in controlled environment in close real-life conditions. In this work, we are interested in Bosphorus for its challenge of facial expressions which contain not only prototypical expressions but also action units.*

- ***3D-TEC*:** The 3D-TEC database consists of 106 pairs of identical twins and a set of triplets, totalizing 214 subjects. Each subject contains two scans: one neutral scan and one smile scan. More details can be found in [23]. *This dataset thus enbales the evaluation of 3D FR algorithms under hard conditions, i.e. the joint presence of strong inter-class similarities and intra-class expression variations.*

All scans of FRGC v1.0, FRGC v2.0, and 3D-TEC databases are preprocessed by using the *3D Face Models Preprocessing Tool* [2] developed by Szeptycki *et al.* [6]. The preprocessing pipeline contains: spike and noise removing, holes filling, nose tip localization and face cropping. As introduced in [6], a decision-based median filtering technique is used to remove spikes, and the holes are detected by searching vertexes having less than 8 neighbors, and filled by fitting square surfaces. Nose tip is located through a curvature analysis-based coarse grained search and generic face model-based fine grained search. As an exception, for 3D-TEC, the manually labeled nose tips provided by the database are used in this work. Finally, each scan is cropped by a sphere centering at nose tip and with a radius of 90 mm. The polygon surface scans in BU-3DFE are first preprocessed as discrete manifold triangular meshes and then projected as range images by an interpolation algorithm. Then, nose tip localization and face cropping are carried out for all the scans of BU-3DFE and Bosphrous databases by using the same aforementioned preprocessing tool. Regarding the registration, we select a face scan with neutral expression in frontal pose for each of the five databases as a reference model, all the other face scans are then aligned to the reference model using the Iterative Closest Point (ICP) [49] algorithm. Fig. 9 illustrates several examples of preprocessed face models.

5.2. Experimental Settings

To comprehensively evaluate the proposed approach and highlight its various facets, six experiments are designed. 1) The discriminative power of the proposed LNP descriptor. This experiment aims to compare the proposed LNP with the raw normal maps and the direct encoding of range images to highlight the effectiveness of the proposed encoding to normal maps; 2) The effectiveness of SRC. This experiment aims to highlight the advantage of using sparse representation-based classifier (SRC) in comparison with the popular Chi-square distance when histogram-based features (*e.g.*,LNP) are involved; 3) The patch weights learning and the effectiveness of W-SRC. This experiment is to

[2] http://pszeptycki.com/tool.html

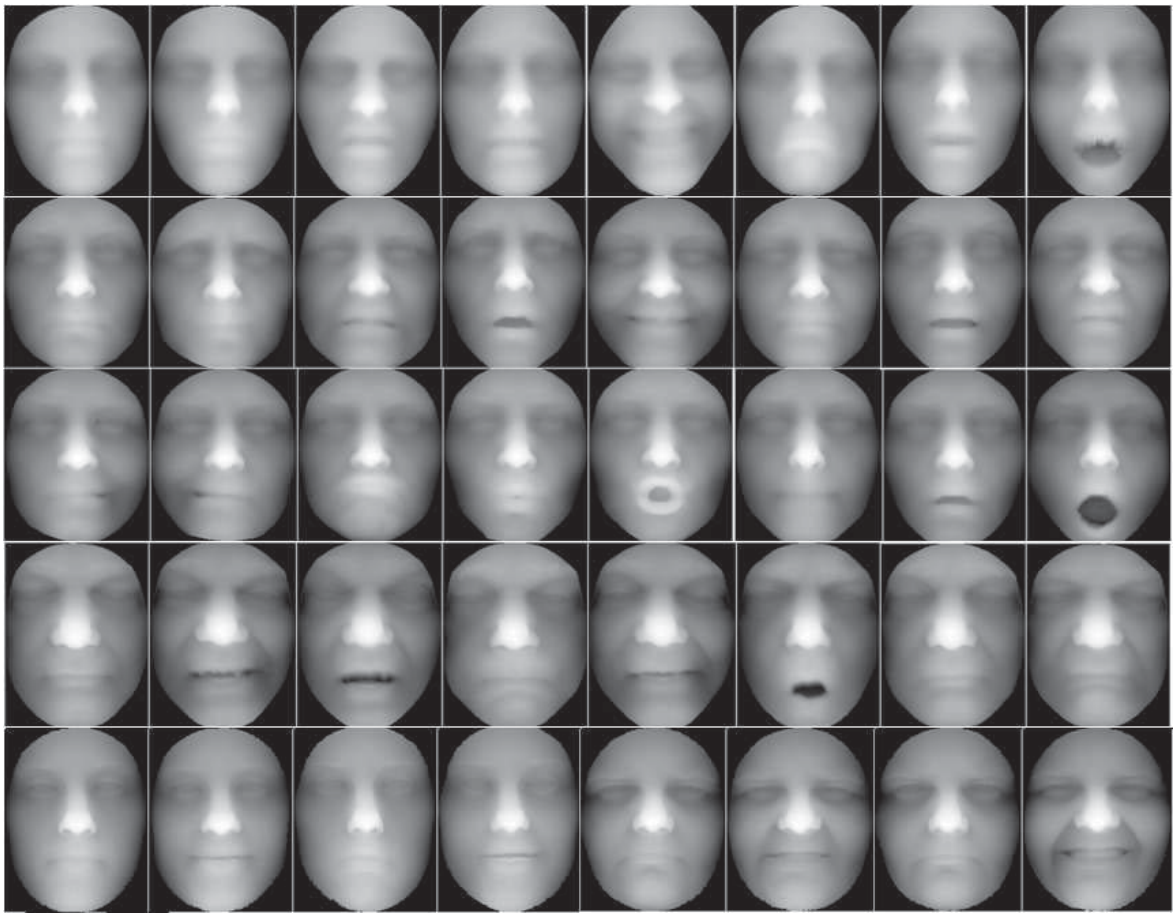


Figure 9: Illustration of several examples of preprocessed face models: first row: models of one subject with different facial expressions (BU-3DFE); second and third rows: models of one subject with different facial expressions and action units (Bosphorus); forth row: models of one subject with different facial expressions (FRGC v2.0); last row: two pairs of identical twin models with neutral and simile expressions (3D-TEC).

show the learning process of patch weights, to bring out the difference of patch weights between 3D and 2D face data, and to highlight the effectiveness of the proposed W-SRC in terms of performance improvement; 4) The robustness analysis of the proposed approach with respect to various facial expression variations. This experiment aims to highlight the behavior of the proposed approach with respect to facial expressions as well as their intensities; 5) The performance of distinguishing expressive identical twins. This experiment aims to benchmark through 3D-TEC the proposed 3D algorithm in hard conditions, *i.e.* the recognition of expressive identical twins with strong inter-class similarities in presence of intra-class variations introduced by facial expressions. 6) The comparison with the state-of-the-art. This study compares the proposed approach with the state of the art using the four datasets whenever state of the art results are available.

The experimental settings are as follows: for the FRGC v1.0 and FRGC v2.0 databases, the first scans of each subject are used to make a gallery set and the remaining 3D face scans are treated as probe; for the BU-3DFE database, the neutral scans are used to make a gallery set and the remaining scans are treated as the probe set. For the Bosphorus database, we focus on the challenge of facial expression variations and select the first neutral scans to make the gallery set whereas the remaining scans with frontal pose and without occlusions are treated as probe, thus excluding those scans with pose variations and occlusions which depict different challenges. Table 1 summarizes these protocols, and presents the sizes of both the gallery sets and probe sets for those four databases.

Table 1: Experimental settings of FRGC v1.0, BU-3DFE, Bosphorus, and FRGC v2.0 databases (O/R means Occlusion and Rotation).

Database	**Gallery**	**Probe**
FRGC v1.0	first scans (267)	remaining (571)
FRGC v2.0	first scans (466)	remaining (3541)
BU-3DFE	neutral scans (100)	remaining (2400)
Bosphorus	first neutral scans (105)	without O/R (1797)

The gallery and probe scans used for 3D-TEC database is based on the standard protocol shown in Table 2 [23]. One person in each pair of twins is arbitrarily labeled as Twin A and the other as Twin B, and four Cases are considered. In Case I, all the images in the gallery set possess a smiling expression while all the images in the probe set have a neutral expression. Case II reverses these roles of Case I. In Case III, Twin A smiling and Twin B neutral make up of the gallery set; while Twin A neutral and Twin B smiling as probe compose the probe set. Case IV reverses these roles of Case III. As pointed out in [23], theoretically the main challenge would be to distinguish between the probe image and the image of his/her twin in the gallery. Case III and IV are more difficult than Cases I and II since the expression of the probe face is different from his/her image in the gallery but is the same as the image of his/her twin in the gallery.

Table 2: Experimental setting of 3D-TEC database: "A Smile, B Neutral" means that the set contains all images with Twin A smiling and Twin B neutral [23].

No.	**Gallery**	**Probe**
I	A Smile, B Smile	A Neutral, B Neutral
II	A Neutral, B Neutral	A Smile, B Smile
III	A Smile, B Neutral	A Neutral, B Smile
IV	A Neutral, B Smile	A Smile, B Neutral

Before encoding the normal information, three normal component matrices or images $[n_{ij}^x]_{m\times n}$, $[n_{ij}^y]_{m\times n}$ and $[n_{ij}^z]_{m\times n}$ are resized into 120×96 matrices, respectively. Each normal component matrix is divided into 10×8, 6×6 and 3×3 windows corresponding to local patches with sizes of 12×12, 20×16 and 40×32, respectively. Then, three different scales are considered with all kinds of local patches. This amounts to performing encoding operators $Q_{1,8}$, $Q_{2,16}$, and $Q_{3,24}$ on local patches with sizes of 12×12, 20×16 and 40×32, respectively. As a result, for each normal component, we encode it with three different scales, achieving three histograms of local normal patterns (LNP). In order to reduce the dimensionality of final facial features, the uniform pattern strategy [36] as in LBP is adopted to decrease the number of bins in each local patch. Finally, from one original 3D face scan, we generate 9

histograms of local normal patterns (3 normal components and 3 encoding scales) involving both local patch based and global features. Each histogram representation of the whole face is fed into the classifier to achieve one similarity score matrix. All the 9 similarity score matrices are then fused through a simple sum rule to compute the final accuracy of MSMC-LNP. To solve (11) and (16), the Orthogonal Matching Pursuit (OMP) [46] algorithm with the sparse number $L = 30$ of the sparse representation coefficient $\hat{x}$ is used in all the experiments.

5.3. Experimental Results

5.3.1. Experiment I: The discriminative power of local normal patterns

To highlight the discriminating power of the proposed LNP based facial feature, we compare it with other two kinds of facial features: i) The raw normal information based facial features Nx, Ny and Nz, by simply stacking the columns of each normal component matrices n_{ijx}, n_{ijy} and n_{ijz} respectively, and their fusion Nxyz. ii) Local Shape binary Patterns (LSP), *i.e.* LBP histograms extracted directly from range images. For a fair comparison, LNP descriptor used the same encoding parameter (*i.e.* $Q_{2,16}$) with LSP to extract the feature vector on each normal component, respectively noted as LNPx, LNPy and LNPz, and their fusion, *i.e.* Multi-Component Local Normal Patterns (MC-LNP). All features were finally fed into a SRC classifier. Once again, the score-level fusion through a simple sum rule was employed for combining different normal components and encoding scales.

Table 3: Comparison of rank-one scores: original normal, LSP and LNP on the whole FRGC v2.0 database.

Approaches	**Rank-one Scores**
(1) Nx + SRC	67.83%
(2) Ny + SRC	65.62%
(3) Nz + SRC	71.63%
(4) Nxyz + SRC	73.19%
(5) $LSP_{2,16}$ + 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) MC-LNP($Q_{2,16}$) + SRC	**92.60%**

Table 3 reports the rank-one recognition rates on the whole FRGC v2.0 database. We can see that LNP performs much better (about 20% higher) than the original normal feature. On the other side, without normal information, the result based on LSP is about 5% lower than that of each encoded normal component and 10% lower than their fusion, *i.e.* MC-LNP($Q_{2,16}$). This experiment indicates that the encoded normal information (LNP) is more discriminative not only than the original normal information, but also than the encoded depth information (LSP).

5.3.2. Experiment II: The effectiveness of sparse representation-based classifier

For histogram based feature vector (*e.g.*, LNP), Chi-Square distance is the preferred similarity measurement in the literature [50]. Table 4 compares the rank-one recognition rates achieved by SRC and Chi-Square distance based classifiers on the whole FRGC v2.0 database. All the results are achieved using LNP feature with the same encoding scale (i.e. $Q_{2,16}$).

As it can be seen from Table 4, the rank-one scores of SRC using LNPx, LNPy and LNPz as well as their fusion MC-LNP, with an average gain of 8 points, consistently outperform those of Chi-square distance-based classifier using the same feature vectors. These results highlight the effectiveness of SRC when using local normal patterns (LNP) based facial representation.

5.3.3. Experiment III: The patch weight learning and the effectiveness of W-SRC

In this experiment, firstly, we describe the way to learn patch weights and analyze the relative importance of facial physical components for face identification. Then, we compare the performance of W-SRC and SRC on FRGC v2.0, Bosphorus, and BU-3DFE respectively. Three databases are used for learning the patch weights: FRGC v1.0, BU-3DFE, and Bosphorus. The experimental protocol listed in Table 1 is used, and according to the proposed framework,

Table 4: Comparison of rank-one scores: Chi-Square vs. SRC on the whole FRGC v2.0 database.

Approaches	**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) MC-LNP($Q_{2,16}$) + Chi-Square	82.64%
(8) MC-LNP($Q_{2,16}$) + SRC	**92.60%**

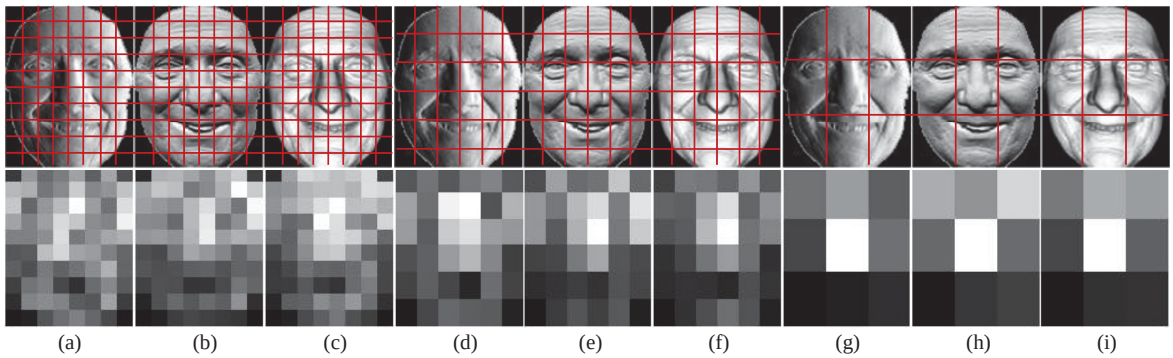


Figure 10: Illustration of the patch weights learned from the Bosphorus database. Columns (a-c), normal images x, y and z and their patch weights (10×8 patches); columns (d-f), normal images x, y and z and their patch weights (6×6 patches); columns (g-i), normal images x, y and z and their patch weights (3×3 patches). Darker patches indicate lower weights, while brighter ones indicate higher weights.

the patch weights are achieved by the following four steps: 1) divide each normal component into local patches (10 $\times$ 8, 6 $\times$ 6, and 3 $\times$3 windows); 2) extract patch based MSMC-LNP features, three normal components and three encoding scales ($Q_{1,8}$, $Q_{2,16}$, and $Q_{3,24}$); 3) compute patch based rank-one scores using SRC classifier on a given training database. 4) compute patch based weights by normalizing the patch based scores.

Fig. 10 shows the results of such a learning process using Bosphorus as training database in displaying the patch weights of three normal component images x, y and z with three binary encoding scales $Q_{1,8}$, $Q_{2,16}$ and $Q_{3,24}$. The patch number with respect to those three encoding scales is 10×8, 6×6 and 3×3, respectively. The weights are marked by gray values where darker ones indicate lower weights while the brighter ones indicate higher weights. We can see that the weight distribution patterns are quiet different to each other among different normal components and different encoding scales but we can observe some similar trends, in particular with the largest weights near the nose regions, and larger weights near the eyes, while smallest weights near the mouth regions and the boundary parts. For more detail, consider column (e) in Fig. 10 as an example. The rigid regions including nose, eyes and forehead totally possess about 56% importance of the whole face. 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 identity related information in those non-rigid facial regions. Note that this kind of weight distribution patterns are quite different from those of 2D face, especially in the nose region in comparison with Fig. 1 (b). This difference can be explained by the different data nature between 2D and 3D faces; for example, the nose region in a 2D image is easily influenced by the variations of illumination whilst the one of the 3D face remains stable under expression variations.

To evaluate the effectiveness of W-SRC to facial expression variations, we compare the performance of SRC and the W-SRC on FRGC v2.0, Bosphorus, and BU-3DFE respectively (see Table 5, 6, and 7). The weights learned from FRGC v1.0, BU-3DFE, and Bosphorus are denoted as F-W-SRC, BU-W-SRC, and BO-W-SRC respectively. The local normal encoding operator $Q_{2,16}$ is used in all the three Tables. Table 5 presents the rank-one scores on the FRGC v2.0 database using SRC, F-W-SRC, BU-W-SRC, and BO-W-SRC. The results using the single normal component

Table 5: Comparison of rank-one score improvements on the FRGC v 2.0 database: patch weights are learned using FRGC v1.0, BU-3DFE, and Bosphorus respectively.

	LNPx($Q_{2,16}$)	LNPy($Q_{2,16}$)	LNPz($Q_{2,16}$)	MC-LNP($Q_{2,16}$)
SRC	87.01%	86.13%	88.43%	92.60%
F-W-SRC	86.63%	88.40%	88.65%	93.59%
BU-W-SRC	**88.85%**	88.54%	**90.58%**	94.50%
BO-W-SRC	88.62%	**88.88%**	90.41%	**94.61%**

Table 6: Comparison of rank-one score improvements on the BU-3DFE database: patch weights are learned from Bosphorus.

	LNPx($Q_{2,16}$)	LNPy($Q_{2,16}$)	LNPz($Q_{2,16}$)	MC-LNP($Q_{2,16}$)
SRC	**78.92%**	80.92%	84.08%	88.25%
BO-W-SRC	78.83%	**84.29%**	**86.21%**	**90.71%**

LNPx($Q_{2,16}$), LNPy($Q_{2,16}$), and LNPz($Q_{2,16}$) and the one of their fusion MC-LNP($Q_{2,16}$) are reported. We can see from Table 5 that the performance of F-W-SRC is slightly better than SRC except LNPx($Q_{2,16}$). The results of BU-W-SRC and BO-W-SRC are similar and both are improved by 1.5% to 2% in comparison with SRC. These results suggest W-SRC along with the weight learning strategy does provide more robustness to facial expression variations than SRC.

Table 6 presents the rank-one scores on the BU-3DFE database using SRC and BO-W-SRC. We can see that the performance improvements based on BO-W-SRC are largely different in the three normal components, with -0.09%, 3.37% and 2.13% for LNPx, LNPy and LNPz respectively. These results indicate that the facial surface deformations caused by facial expression variations are likely to decompose into different quantities over different normal components. The improvement of the fusion result using MC-LNP is about 2.5% which also proves the effectiveness of W-SRC handling facial expression variations.

Table 7 presents the rank-one scores on the Bosphorus database using SRC and BU-W-SRC. We can see that BU-W-SRC improves the performance for all three normal components, with 3.76%, 2.03% and 4.19% for LNPx, LNPy, and LNPz respectively.

Table 7: Comparison of rank-one score improvements on the Bosphorus database: patch weights are learned from BU-3DFE.

	LNPx($Q_{2,16}$)	LNPy($Q_{2,16}$)	LNPz($Q_{2,16}$)	MC-LNP($Q_{2,16}$)
SRC	83.12%	86.24%	84.91%	90.92%
BU-W-SRC	**86.88%**	**88.27%**	**89.10%**	**93.21%**

5.3.4. Experiment V: Comparison of the performance degradation due to facial expression variations

We first evaluate the degradation influenced by facial expression variations on the FRGC v2.0 database. According to the experimental protocol used in [13], [15] and [51], we split all probe faces into two subsets based on their original expression labels. The first subset consists of only neutral faces, while the second one is only made up of non-neutral faces. The performance degradation, reflected by the difference between the accuracies of subset I and II reported in Table 8, is utilized to analyze the robustness to facial expression variations. We can see from that table that 6.6% drop is achieved based on the proposed MSMC-LNP descriptor and SRC, and 3.8% drop is obtained by using Bosphorus database as training set for W-SRC. Note that our performance on subset I is a little bit worse than [13, 15, 51], while the degradations are competitive in comparison with their results.

Table 8: Comparing the degradations of rank-one scores influenced by facial expression changes on the FRGC v 2.0 database (Subset I: neutral probes; Subset II: non-neutral probes).

	Sub. I	**Sub. II**	**Degradation**
(1) Mian et al. [13]	99.0%	86.7%	12.3%
(2) Huang et al. [15]	99.1%	92.5%	6.6%
(3) Huang et al. [51]	99.0%	94.9%	4.1%
(4) MSMC-LNP + SRC	97.1%	90.5%	6.6%
(5) MSMC-LNP + BO-W-SRC	98.0%	94.2%	**3.8%**

Since all the 2,400 non-neutral probe faces in the BU-3DFE database have labels of expression intensity levels (increasing from level 1 to level 4), it is thus possible to quantify the performance drop as the the intensity of facial expressions increases. For this purpose, all the probe faces are divided into four subsets according to their labels of expression intensity. Subset I, II, III, and IV are made up of the probe faces with the expression intensity of level 1, level 2, level 3, and level 4 respectively, and each subset consists of 600 probe faces with six prototypical expressions.

Table 9: the degradations of rank-one scores of the proposed approach as the intensity of facial expressions increases on the BU-3DFE database.

	Sub. I	**Sub. II**	**Sub. III**	**Sub. IV**
MSMC-LNP + SRC	97.0%	94.0%	90.5%	80.5%
MSMC-LNP + BO-W-SRC	**97.3%**	**95.0%**	**92.7%**	**83.8%**

The performance is shown in Table 9. We can find out that the degradation from the lower level to higher level expression intensity becomes larger and larger especially from Subset III to Subset IV. By using SRC without weight learning, the degradations are 3.0% from Subset I to Subset II, 3.5% from Subset II to Subset III, and 10.0% from Subset III to Subset IV. By using BO-W-SRC, the degradations are 2.3% from Subset I to Subset II, 2.3% from Subset II to Subset III, and 8.9% from Subset III to Subset IV. As we can see, all the three degradations using BO-W-SRC are smaller than those using SRC, suggesting that the weight learning using different facial components is an effective way for an improved robustness to expression changes.

5.3.5. The performance of distinguishing identical twins across expression variations

In this experiment, we evaluate the performance of our system to distinguish identical twins with a smile expression. We regard the SRC based recognition rate as the baseline and compare it with W-SRC, where the patch weights are learned from different training sets. Given the fact that there are only neutral and smile scans in the 3D-TEC dataset, a specific training set is designed based on the subset of Bosphrous dataset, *i.e.* 105 first neutral scans as gallery and 105 happy scans as probe, the corresponding W-SRC is denoted as BOS-W-SRC. All the rank-one scores achieved by using MSMC-LNP feature as well as SRC, F-W-SRC, BU-W-SRC, BO-W-SRC, and BOS-W-SRC classifiers are shown in Table 10.

Table 10: Comparison of the rank-one scores on 3D-TEC by using different training sets.

Algorithm	**Rank-one scores**			
	I	II	III	IV
MSMC-LNP + SRC	**94.9%**	96.3%	89.3%	88.3%
MSMC-LNP + F-W-SRC	93.5%	94.4%	88.8%	88.3%
MSMC-LNP + BU-W-SRC	93.9%	96.3%	90.7%	91.6%
MSMC-LNP + BO-W-SRC	94.4%	**96.7%**	**90.7%**	**92.5%**
MSMC-LNP + BOS-W-SRC	**95.8%**	**96.7%**	**95.3%**	**95.3%**

From Table 10, we can see that the performance improvements are very limited for F-W-SRC, BU-W-SRC, and BO-W-SRC. The main reason is the asymmetry of the training and testing data. The 3D-TEC dataset contains identical twin samples only with neutral and smile expressions, while the FRGC v1.0 database only includes neutral expression scans; BU-3DFE and Bosphrous databases consist of the scans with different expression types. The performance of BOS-W-SRC confirms this analysis, *i.e.* when the sample distributions of the training and testing sets are more similar to each other, W-SRC will be more efficient, with 6% and 7% improvements for Cases III and IV.

5.3.6. Experiment IV: Comparison with the state-of-the-art

To further evaluate the performance of the proposed method, we display, in Table 11, a comprehensive comparison of the rank-one recognition rates on the FRGC v2.0, Bosphrous, BU-3DFE, and 3D-TEC databases with the state of the art. In that table, our best results are highlighted as well as those of the state of the art better than ours. From that table, we can observe that:

(i) The vast majority of 3D FR algorithms only report their performance on FRGC v2.0 which is de facto standard benchmark in 3D FR. Table 11 only lists some of them. However, because of different nature of 3D face data captured under different conditions and scenarios, a 3D FR technique with high performance on FRGC v2.0 does not guarantee the same level of performance over other 3D face data and can even vastly degrade (see (5-b) and (5-c) on 3D TEC). This means that it is important to benchmark 3D FR techniques over other datasets to check their generalization skills. To the best of our knowledge, except our method, only (8-a) and (8-b) report their results on all the four databases.

(ii) The proposed approach, *i.e.* Multi-Scale and Multi-Component Local Normal Patterns (MSMC-LNP) along with the weighted SRC for an improved robustness to facial expression variations, consistently display competitive rank-one recognition rates over these various datasets, and even the best results in the hard scenarios, namely case III and case IV with 3D-TEC, and thereby demonstrates its effectiveness and its generalization skills in particular with respect to various types and intensities of facial expressions. The results achieved by the proposed approach on 3D-TEC tend to suggest that the proposed MSMC-LNP facial feature is discriminative enough to distinguish even identical twins despite of their strong similarities whereas W-SRC, in learning the discriminative weights of different facial regions, makes the proposed solution robust enough to facial expression variations as stated in hard case III and IV on 3D-TEC.

(iii) Facial normal information is also used as 3D facial representation for the extraction of facial features, *e.g.*, (4), (7), (8-a) and (8-b). In (4), difference of normal maps is used as similarity measurement, while a rank-one score of 92.2% is reported on a subset of FRGC v2.0 (1024 samples) database; In (7), Gokbert *et al.* use surface normal variance at each pixel location as a distance measure between face images and report a rank-one score of 87.8% on the whole FRGC v2.0 database, while this reasonable performance vastly degrades on the 3D-TEC database, achieving only around 60% rank-one recognition rate. In (8-a) and (8-b), wavelet coefficients are used as similarity measurement on both normal and geometry maps. Note that this method is the one of the best 3D FR method in the literature with a very good generalization skill. Kakadiaris *et al.* in (8-a) and (8-b) use a very sophisticated face registration (spin images and ICP) and fitting (Annotated Face Model) techniques, and a Linear Discriminant Analysis (LDA) based feature selection techniques in their following works [58, 59]. Compared with the proposed method, they make use of additional information of 3D face scans, *e.g.*,geometry image, and achieve better results on all the four databases except hard Case III and Case IV over the 3D-TEC database.

6. Conclusion and Future Work

In this paper, we presented an expression-robust 3D face identification approach based on a novel 3D facial surface descriptor, namely Multi-Scale and Multi-Component Local Normal Patterns (MSMC-LNP), along with a weighted Sparse Representation-based Classifier (W-SRC). The extensive experiments that we carried out using 4 different 3D face datasets indicate that: 1) LNP is much more discriminative than both the original normal information and LSP. 2) Both multi-scale and multi-component are efficient manners to improve the performance of LNP. 3) SRC is more efficient than the Chi-square distance based classifier. 4) The importance of facial physical component for 3D face identification is quite different from the one of 2D based, especially in the nose region. 5) Patch-weight based W-SRC is very robust to facial expression variations, even for identical twins with expression changes, and large improvement can be achieved if the distributions of training and testing sets are similar to each other. 6) The proposed system (i.e.

Table 11: Comparison of the Rank-one recognition rates on the FRGC v2.0, Bosphrous, BU-3DFE and 3D-TEC databases.

Approaches	FRGC v2.0	Bosphrous	BU-3DFE	3D TEC			
				Case I	Case II	Case III	Case IV
(1) Chang et al. [52]	91.9%	-	-	-	-	-	-
(2) Cook et al. [53]	92.9%	-	-	-	-	-	-
(3-a) Mian et al. [13]	93.5%	-	-	-	-	-	-
(3-b) Mian et al. [10]	96.2%	-	-	-	-	-	-
(3-c) Mian et al. [54]	93.8%	-	-	-	-	-	-
(3-d) Mian et al. [55]	**96.5%**	-	-	-	-	-	-
(4) Abate et al. (Normal) [34]	92.2%	-	-	-	-	-	-
(5-a) Faltemier et al. (ICP) [9]	**97.2%**	-	-	-	-	-	-
(5-b) Faltemier et al. (ICP) [a] [9] [56]	**98.0%**	-	-	93.5%	93.0%	72.0%	72.4%
(5-c) Faltemier et al. (ICP) [b] [9] [56]	**98.0%**	-	-	94.4%	93.5%	72.4%	72.9%
(6-a) Huang et al. (SI) [15]	91.8%	-	-	92.1%	93.0%	83.2%	83.2%
(6-b) Huang et al. (eLBP) [51]	**97.2%**	-	-	91.1%	93.5%	77.1%	78.5%
(6-c) Huang et al. (Range PFI) [57]	95.5%	-	-	91.6%	93.9%	68.7%	71.0%
(7) Gokbert et al. (Normal + Geometry) [35]	87.8%	-	-	62.6%	63.6%	54.2%	59.4%
(8-a) Kakadiaris et al. (Normal + Geometry) [12] [58]	**97.0%**	-	-	**98.1%**	**98.1%**	91.6%	93.5%
(8-b) Kakadiaris et al. (Normal) [c] [59]	**97.9%**	**98.2%** (2797/105)	**99.7%**	-	-	-	-
(9) Alyuz et al. [7]	**97.5%**	**98.2%** (2814/105)	-	-	-	-	-
(10) C. Maes et al. [60]	89.6%	**97.7%** (3186/105)	-	-	-	-	-
(11) H. Li et al. [28]		94.1% (4561/105)	-	-	-	-	-
(12) Queirolo et al. [29]	**98.4%**	-	-	-	-	-	-
(13) Spreeuwers et al. [22]	**99.0%**	-	-	-	-	-	-
MSMC-LNP + SRC	-	-	-	94.9%	96.3%	89.3%	88.3%
MSMC-LNP + BU-W-SRC	-	**95.4%** (2797/105)	-	93.9%	96.3%	90.7%	91.6%
MSMC-LNP + BO-W-SRC	**96.3%**	-	**92.21%**	94.4%	96.7%	90.7%	92.5%
MSMC-LNP + BOS-W-SRC	-	-	-	**95.8%**	**96.7%**	**95.3%**	**95.3%**

[a] match scores normalization using E_{pkn}, $E_{pkn}(p, g_k) = Emin(p, g_k) / \sum_{j=1, j\neq k}^{N} (Emin(g_i, g_k)/(N-1))$, where p is a probe image, g_k are the gallery images, and N is the number of gallery images. $Emin(p_1, p_2) = min(E(p_1, p_2), E(p_2, p_1))$, and $E(p_1, p_2)$ be the match score of point clouds p_1 and p_2.

[b] match scores normalization using E_{minmax}, which is the min-max normalization over the resulting match score from the E_{pkn} normalization

[c] Bosphrous, BU-3DFE and Bosphrous as the training sets corresponding FRGC v2.0, Bosphrous and BU-3DFE as the testing sets respectively.

MSMC-LNP + W-SRC) consistently achieved competitive rank-one recognition rates on the FRGC v2.0, Bosphrous, BU-3DFE, and 3D-TEC databases and thereby displays a good generalization skill with respect to various types of facial expressions.

In the future, we will focus on the following two aspects to further improve the proposed method. 1) In this work, a basic 3D face alinement method, *i.e.* few iterations of ICP with respect to a reference model, was used. Recent works [22, 30] suggest that a better 3D face alignment can further lead to significant performance improvement. We want to improve the proposed approach with a better alignment method; 2) In this paper, we only presented the identification results of the proposed method. Notice that there are very few works on the study of sparse representation-based classifier for 2D and 3D face verification. Reference [61] is perhaps the only work stressing this issue for 2D face verification. In our future work, we will study sparse representation-based classifier for 3D face verification and report the verification results on the FRGC v2.0, Bosphrous, BU-3DFE and 3D-TEC databases. 3) In this work, only the challenge of facial expression variations was considered. However, in real-life applications, 3D face scans from uncooperative subjects under uncontrolled conditions can be in an arbitrary pose, thus with missing face data, along with possible external occlusions (*e.g.*, glasses, scarf, *etc.*). We want to extend this work to face those new challenges.

Acknowledgments

This work is in part jointly supported by the French research agency, Agence Nationale de Recherche (ANR) and Natural Science Foundation of China (NSFC), within the 3D Face Analyzer (grant ANR 2010 INTB 0301 01; grant NSFC 61061130560), and the 3D Face Interpreter project supported by the LIA 2MCSI lab between the group of Ecoles Centrales and Beihang University.

References

[1] W. Zhao, R. Chellappa, P. J. Phillips, A. Rosenfeld, Face recognition: A literature survey, ACM Computing Surveys 35 (2003) 399–458.
[2] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: Proc. Int. Conf. Computer Vision and Pattern Recognition, 2005.
[3] K. W. Bowyer, K. Chang, P. Flynn, A survey of approaches and challenges in 3d and multi-modal 3d+2d face recognition, Computer Vision and Image Understanding 101 (2006) 1–15.
[4] B. Gokberk, H. Dutagaci, A. Ulas, L. Akarun, B. Sankur, Representation plurality and fusion for 3-d face recognition, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 38 (1) (2008) 155–173.
[5] X. Zhao, E. Dellandréa, L. Chen, I. A. Kakadiaris, Accurate landmarking of three-dimensional facial data in the presence of facial expressions and occlusions using a three-dimensional statistical facial feature model, IEEE Transactions on Systems, Man, and Cybernetics, Part B 41 (5) (2011) 1417–1428.
[6] P. Szeptycki, M. Ardabilian, L. Chen, A coarse-to-fine curvature analysis-based rotation invariant 3d face landmarking, in: Proc. IEEE Int. Conf. Biometrics: Theory Applications and Systems, 2009.
[7] N. Alyüz, B. Gökberk, L. Akarun, Regional registration for expression resistant 3-d face recognition, IEEE Transactions on Information Forensics and Security 5 (3) (2010) 425–440.
[8] Y. Wang, J. Liu, X. Tang, Robust 3d face recognition by local shape difference boosting, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (10) (2010) 1858–1870.
[9] T. C. Faltemier, K. W. Bowyer, P. J. Flynn, A region ensemble for 3d face recognition, IEEE Transactions on Information Forensics and Security 3 (1) (2008) 62–73.
[10] A. S. Mian, M. Bennamoun, R. A. Owens, An efficient multimodal 2d-3d hybrid approach to automatic face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (11) (2007) 1927–1943.
[11] C. Samir, A. Srivastava, M. Daoudi, Three-dimensional face recognition using shapes of facial curves, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (11) (2006) 1858–1863.
[12] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis, 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 29 (4) (2007) 640–649.
[13] A. S. Mian, M. Bennamoun, R. A. Owens, Keypoint detection and local feature matching for textured 3d face recognition, International Journal of Computer Vision 79 (1) (2008) 1–12.
[14] S. Berretti, A. D. Bimbo, P. Pala, 3d face recognition using isogeodesic stripes, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (12) (2010) 2162–2177.
[15] D. Huang, G. Zhang, M. Ardabilian, Y. Wang, L. Chen, 3d face recognition using distinctiveness enhanced facial representations and local feature hybrid matching, in: Proc. Int. Conf. Biometrics: Theory Applications and Systems, 2010.
[16] D. Huang, M. Ardabilian, Y. Wang, L. Chen, 3-d face recognition using elbp-based facial description and local feature hybrid matching, IEEE Transactions on Information Forensics and Security 7 (5) (2012) 1551–1565.
[17] A. M. Bronstein, M. M. Bronstein, R. Kimmel, Expression-invariant representations of faces, IEEE Transactions on Image Processing 16 (1) (2007) 188–197.

[18] H. Drira, B. B. Amor, A. Srivastava, M. Daoudi, R. Slama, 3d face recognition under expressions, occlusions and pose variations, IEEE Transactions on Pattern Analysis and Machine Intelligence 99 (PrePrints) (2013) 1. doi:http://doi.ieeecomputersociety.org/10.1109/TPAMI.2013.48.
[19] B. Ben Amor, M. Ardabilian, L. Chen, Toward a region-based 3d face recognition approach, in: IEEE International Conference on Multimedia and Expo, 2008, pp. 101–104.
[20] B. B. Amor, M. Ardabilian, L. Chen, Enhancing 3d face recognition by mimics segmentation, in: 6th International Conference on Intelligent Systems Design and Applications (ISDA), 2006, pp. 150–155.
[21] K. I. Chang, K. W. Bowyer, P. J. Flynn, Multiple nose region matching for 3d face recognition under varying facial expression, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (10) (2006) 1695–1700.
[22] L. Spreeuwers, Fast and accurate 3d face recognition using registration to an intrinsic coordinate system and fusion of multiple region classifiers., International Journal of Computer Vision 93 (3) (2011) 389–414.
[23] V. Vijayan, K. W. Bowyer, P. J. Flynn, D. Huang, L. Chen, M. Hansen, O. Ocegueda, S. K. Shah, I. A. Kakadiaris, Twins 3d face recognition challenge, in: Proc. Int. Joint Conf. on Biometrics, 2011.
[24] V. Vijayan, K. W. Bowyer, P. J. Flynn, 3d twins and expression challenge, in: Proc. Int. Conf. on Computer Vision Workshops, 2011.
[25] G.G.Gordon, Face recognition based on depth and curvature features, in: Proc. Int. Conf. Computer Vision and Pattern Recognition, 1992.
[26] D. Sun, W. Sung, R. Chen, 3d face recognition based on local curvature feature matching, Applied Mechanics and Materials 121 (126) (2011) 609–616.
[27] G. Zhang, Y. Wang, Robust 3d face recognition based on resolution invariant features, Pattern Recognition Letters 32 (7) (2011) 1009–1019.
[28] H. Li, D. Huang, P. Lemaire, J. Morvan, L. Chen, Expression robust 3d face recognition via mesh-based histograms of multiple order surface differential quantities, in: Proc. IEEE Int. Conf. on Image Processing, 2011.
[29] C. Queirolo, L. Silva, O. Bellon, M. Segundo, 3d face recognition using simulated annealing and the surface interpenetration measure, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (2) (2010) 206–219.
[30] H. Mohammadzade, D. Hatzinakos, Iterative closest normal point for 3d face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (2) (2013) 381 –397.
[31] A. F. Abate, M. Nappi, S. Ricciardi, G. Sabatino, Fast 3d face recognition based on normal map, in: Proc. IEEE Int. Conf. on Image Processing, 2005.
[32] A. F. Abate, M. Nappi, S. Ricciardi, G. Sabatino, Multi-modal face recognition by means of augmented normal map and pca, in: Proc. IEEE Int. Conf. on Image Processing, 2006.
[33] A. F. Abate, M. Nappi, S. Ricciardi, G. Sabatino, Fast 3d face alignment and improved recognition through pyramidal normal map metric, in: Proc. IEEE Int. Conf. on Image Processing, 2007.
[34] A. F. Abate, M. D. Marsico, S. Ricciardi, D. Riccio, Normal maps vs. visible images: Comparing classifiers and combining modalities, Journal of Visual Languages and Computing 20 (3) (2009) 156–168.
[35] B. Gokberk, M. O. Irfanoglu, L. Akarun, 3d shape-based face representation and feature extraction for face recognition, Image and Vision Computing 24 (8) (2006) 857–869.
[36] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (7) (2002) 971–987.
[37] T. Ahonen, A. Hadid, M. Pietikäinen, Face recognition with local binary patterns, in: Proc. Int. Conf. European Conference on Computer Vision, 2004.
[38] D. Huang, C. Shan, M. Ardabilian, Y. Wang, L. Chen, Local binary patterns and its application to facial image analysis: A survey, IEEE Transactions on Systems, Man, and Cybernetics, Part C 41 (6) (2011) 765–781.
[39] A. Martinez, Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (6) (2002) 748 –763.
[40] Z. Lei, S. Liao, M. Pietikäinen, S. Z. Li, Face recognition by exploring information jointly in space, scale and orientation, IEEE Transactions on Image Processing 20 (1) (2011) 247–256.
[41] D. Huang, K. Ouji, M. Ardabilian, Y. Wang, L. Chen, 3d face recognition based on local shape patterns and sparse representation classifier, in: Advances in Multimedia Modeling:17th International Multimedia Modeling Conference (MMM), 2011, pp. 206–216.
[42] H. Li, D. Huang, J.M.Morvan, L. Chen, Learning weighted sparse representation of encoded facial normal information for expression-robust 3d face recognition, in: Proc. IEEE Int. Joint Conf. on Biometrics, 2011.
[43] K. Klasing, D. Althoff, D. Wollherr, M. Buss, Comparison of surface normal estimation methods for range sensing applications, in: Proc. IEEE Int. Conf. on Robotics and Automation, 2009.
[44] R. Hoffman, A. K. Jain, Segmentation and classification of range images, IEEE Transactions on Pattern Analysis and Machine Intelligence 9 (5) (1987) 608–620.
[45] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2) (2009) 210–227.
[46] Y. C. Pati, R. Rezaiifar, P. S. Krishnaprasad, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition., in: Proc. 27th Asilomar Conf. on Signals, Systems and Computers, 1993.
[47] L. Yin, X. Wei, Y. Sun, J. Wang, M. Rosato., A 3d facial expression database for facial behavior research, in: Proc. 7th Int. Conf. on Automatic Face and Gesture Recognition, 2006.
[48] A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, L. Akarun, 3d face recognition benchmarks on the bosphorus database with focus on facial expressions, in: Proc. Workshop on Biometrics and Identity Management, 2008.
[49] Z. Zhang, Iterative point matching for registration of free-form curves and surfaces, International Journal of Computer Vision 13 (2) (1994) 119–152.
[50] O. Pele, M. Werman, The quadratic-chi histogram distance family, in: Proc. Int. Conf. European Conference on Computer Vision, 2010.
[51] D. Huang, M. Ardabilian, Y. Wang, L. Chen, A novel geometric facial representation based on multi-scale extended local binary patterns, in: Proc. Int. Conf. Automatic Face and Gesture Recognition, 2011.

[52] K. W. Bowyer, K. Chang, P. Flynn, Adaptive rigid multi-region selection for handling expression variation in 3d face recognition, in: Proc. Int. Conf. Computer Vision and Pattern Recognition, 2005.
[53] J. Cook, V. Chandran, C. Fookes, 3d face recognition using log-gabor templates, in: Proc. Int. Conf. British Machine Vision Conference, 2006.
[54] F. R. Al-Osaimi, M. Bennamoun, A. S. Mian, Integration of local and global geometrical cues for 3d face recognition, Pattern Recognition 41 (3) (2008) 1030–1040.
[55] F. Al-Osaimi, M. Bennamoun, A. Mian, An expression deformation approach to non-rigid 3d face recognition, Int. J. Comput. Vision 81 (3) (2009) 302–316.
[56] R. McKeon, Three-dimensional face imaging and recognition: A sensor design and comparative study, in: Ph.D. dissertation, University of Notre Dame, 2010.
[57] D. Huang, W. B. Soltana, M. Ardabilian, Y. Wang, L. Chen, Textured 3d face recognition using biological vision-based facial representation and optimized weighted sum fusion, in: Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops, 2011.
[58] O. Ocegueda, S. Shah, I. Kakadiaris, Which parts of the face give out your identity?, in: IEEE Conference on Computer Vision and Pattern Recognition, 2011, pp. 641–648.
[59] O. Ocegueda, G. Passalis, T. Theoharis, S. Shah, I. Kakadiaris, Ur3d-c: Linear dimensionality reduction for efficient 3d face recognition, in: Proc. IEEE Int. Joint Conf. on Biometrics, 2011.
[60] D. Smeets, J. Keustermans, D. Vandermeulen, P. Suetens, meshsift: Local surface features for 3d face recognition under expression variations and partial data, Computer Vision and Image Understanding 117 (2) (2013) 158–169.
[61] H. Guo, R. Wang, J. Choi, L. Davis, Face verification using sparse representations, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops(CVPRW), 2012, pp. 37–44.

The main contributions of this paper can be summarized as follows:

- Multi-Scale Multi-Component Local Normal Patterns based 3D facial descriptor is proposed. Local Normal Patterns (LNP) is more discriminative than Local Shape Patterns (LSP) feature and the original facial normal feature.
- Quantification weights of local patches of 3D facial surface are learned, which are quite different from those of their 2D counterparts, especially in the nose region.
- A weighted sparse representation classifier (W-SRC) is formulated, making the proposed method robust enough to tolerate various facial expressions and discriminating enough for distinguishing identical twins.
- Extensive experiments were carried out using various 3D face datasets, including FRGC v2.0, BU-3DFE, Bosphorus and 3D-TEC, to benchmark the effectiveness and the generalization skills of the proposed approach with respect to 3D face scans captured under different scenarios and conditions with different 3D sensors, depicting in particular different challenges in terms of facial expressions.