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 well-aligned and preprocessed facial range image, three components of normal vectors are first estimated, leading to three normal images. Then, each nor-

*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)

mal 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 resist the facial expression variations, 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 reformulate a Weighted Sparse Representation-based Classifier (W-SRC). Extensive experiments were carried out on the FRGC v2.0, Bosphorus, BU-3DFE and 3D-TEC databases, enclosing 3D face data captured under various sensors with different resolutions and depicting in particular different challenges with respect to facial expressions. The proposed approach achieves competitive rank-one identification rates over these datasets despite their heterogeneous nature, and demonstrates thereby its effectiveness and its robustness.
Keywords: facial surface normal, local normal patterns (LNP), weighted sparse representation, 3D face recognition, identical twins

1. Introduction

Biometry 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, are still its big block [1]. The advent of 3D sensors, in providing geometrical information of facial surfaces, has opened a new avenue to handle these unsolved issues in 2D. As such, 3D face recognition (FR) has attracted increasing attention in recent years [2, 3].

1.1. Related work

A typical 3D FR algorithm comprises the following major components although they are strongly interwoven each other [**?**]: 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 [4], 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 [5, 6, 7]. 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 [8, 9] although they are generally computationally expensive. The extraction of facial features is to generate a feature vector

which should comprehensively describe each 3D face scan for the latter stage of matching. As all human faces are similar 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, 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 [10], geometry and normal maps [11], tensor based representations [12], iso-geodesic stripes [13], Multi-Scale Local Binary Pattern (MS-LBP) Depth maps and Shape Index (SI) maps [14], Multi-Scale extended Local Binary Pattern (MS-eLBP) maps [15] *etc.* Other essays try to explicitly account for facial expression variations. An original tentative was made by Bronstein *et al.* [16] 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 [**?**]. 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 [**?**] [**?**]. Chang *et al.* [17] selected three regions around the nose for 3D face matching whereas Faltemier *et al.* [**?**] 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 [**?**] [**?**].

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 [18], was reported on that dataset. However, although FRGC v2.0, with its 4007 3D face scans from 466 subjects, is the most comprehensive 3D face dataset so far known in the literature, all the scans were captured in frontal position under controlled lighting conditions, depicting only neutral and smiling facial expressions. 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 other types of facial expressions, being subtle or exaggerated. As a result, 3D FR algorithms with high performance on FRGV v2.0 can vastly degrade under other settings. The recent experiments carried out on the 3D Twins Expression Challenge (3D-TEC) database [19, 20] is quite illustrative from this point of view. 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. [19] 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 degrade on 3D-TEC, especially in the cases combining factors related to facial similarity and the variations of facial expressions. Their results show that benchmarking 3D FR algorithms on FRGC v2.0 is certainly necessary but not sufficient to evaluate their performance and robustness with respect to the challenges of real-life applications, including in particular facial expressions which de-

pict not only smiling expressions but uncountable other facial expressions, leading to subtle, moderate and exaggerated facial surface deformations.

1.2. Motivations and Our Solutions

This paper focuses on the issues iii and iv, i.e. exploring a discriminative facial surface representation and an expression-robust method to handle expression variations. To address the former issue, we propose a novel facial shape descriptor, named 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 original facial surface can be then represented as a certain number of local normal patterns based maps or histograms of Local Normal Patterns (LNP).

As we know, surface curvatures [6, 23, 24] and shape index values [14, 25, 26] 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 for 3D face representation [1]. To the best of our knowledge, Abate et al. [27, 28, 29, 30] 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. [31] used surface normal variance at each pixel location as a distance measure between face images and report a rank-one score of 87.8%

[1] Note that, recently, normal constraint based surface registration for 3D face recognition methods such as [21, 22] have achieved very high performance. In this paper we stress the normal based facial representation.

on the whole FRGC v2.0 database, while this performance vastly degrades on the 3D-TEC database [19]. Kakadiaris et al. [11] 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 face recognition [32, 33], we propose to encode surface normal information, namely x, y, and z component normal images, in a local manner to generate histograms of LNP, similar to the 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, some works proposed to choose rigid facial regions such as the nose and forehead regions [8, 17] since they are expected to remain stable in occurrence of facial expressions. However, the useful information conveyed in non-rigid facial areas is ignored. Some other 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 [16]. 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 useful information which is important for face recognition.

In our view, a better alternative to solve this problem is to find the average quantification weights for all facial regions or facial physical components such as eyes, nose, mouth, etc. according to their importance in 3D face recognition. This kind of quantification weights of local patches for 2D face recognition has been investigated in several works [33, 34, 35] (see Fig. 1 (a) and (b) for an example). However, to the best of our knowledge, the corresponding weights and the effects for expression-robust 3D face recognition has not been well stressed. As shown in Fig. 1 (c) and (d), in this paper, we will show that the weights for 3D face are largely different from the ones of 2D face, 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 the weighted sparse representation model and compute the weighted reconstruction errors, namely Weighted Sparse Representation-based Classifier (W-SRC).

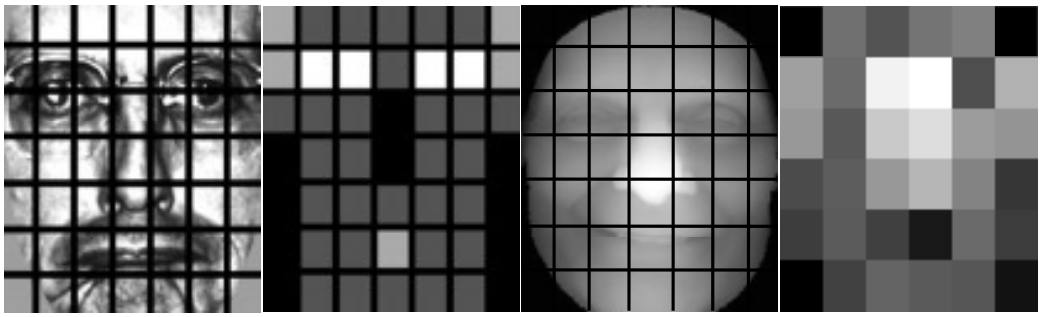

Figure 1: Illustration of patch weights for 2D and 3D face recognitions: (a-b) a 2D face image and its corresponding patch weights [33]; (c-d) a 3D face depth image and its corresponding patch weights learned by our method. All images are split to 6×6 local patches. Darker patches indicate lower weights, while brighter ones indicate higher weights.

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

1. We propose 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. An important conclusion obtained in this paper is: 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) [36] (see Tab. 3). We also illustrate that the fusion of both multiple scales and multiple components is a help way to improve final performance and demonstrate its competency on 3D face identification as well as the challenging issue of recognizing identical twins.

2. By learning strategy, we find that the importance of local patches of 3D facial surfaces is quite different from that of 2D based ones, especially in the nose region. Given a training database, the patch weights associated with different facial regions, encoding scales, and normal components can be achieved by normalizing the patch scores, which computed by running the sparse representation-based classifier (SRC) over MSMC-LNP features of local patches. Combining with the learned patch weights and the MSMC-LNP feature, we build the weighted sparse representation-based classifier. Our experimental results demonstrate that W-SRC is quite efficient to resist the variations of facial expressions even for that of identical twins.

This paper is an extension of our preliminary work in [37]. The main extensions can be highlighted as follows. First, we reformulated the weighted sparse representation-based classifier (W-SRC) in this paper. Then, we evaluated the robustness of the proposed system to different expression intensity

levels using the BU-3DFE database, as well as to the combination of six prototypical expressions and action units using the Bosphrous database. Finally, we validate the discriminative power of the proposed approach in the challenging issue of distinguishing identical twins using the 3D-TEC database.

The rest 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 show experimental and algorithmic settings as well as the 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. the training phase and the testing phase. Before the training and testing phases, each raw face scan is preprocessed to be a range image with a predefined size, including spike and noise removing, holes filling, nose tip localization, face cropping and alignment. The training process is carried out to learn the quantitative weights of facial physical components using a predefined training set. It includes three procedures: feature extraction, identification and score normalization within different patches. Their descriptions are as follows:

(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 facial scan, we first launch the preprocessing pipeline (see Sec. 5.2) to normalize the range image to an $m \times n \times 3$

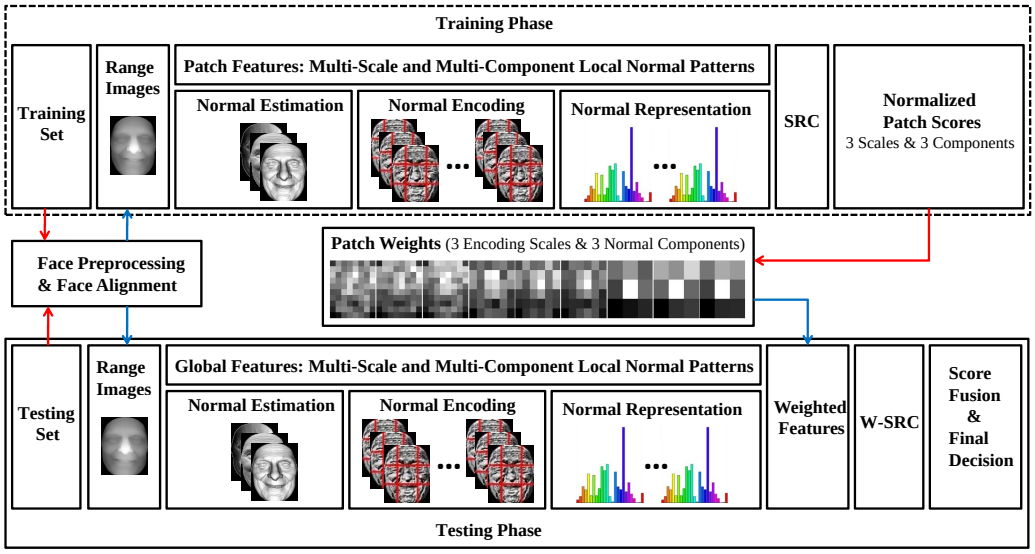

Figure 2: Overview of the proposed approach.

matrix (i.e., x, y and z coordinates). Based on the range image, we estimate its three normal components (x, y, and z) by local plane fitting method (see Sec. 3.1). For each normal component map, we roughly split it to 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 describe 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 can thus be measured by those quantitative patch weights.

During the testing phase, given a split range image in the testing set, we first compute MSMC-LNP features over all the patches as 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 is formulated as seeking the sparse solution of the sum of the weighted patch based sparse representation (as (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 computed by score level fusion of three encoding scales and three normal components is used for decision making.

3. Local Normal Patterns (LNP) based Facial Descriptor

3.1. Facial Normal Estimation

Recall that in order to highlight local variations of facial surfaces, we do analysis based on 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 [38].

The basic idea of optimization based ones is: 1) the normal vector of one point can be calculated by the normal vector of a plane or quadratic surface which it belongs to. 2) the formulate of plane or surface can be estimated by fitting its local neighboring points. 3) the fitting problem then can be solved by minimizing a cost functional penalizing a certain criteria, e.g. the

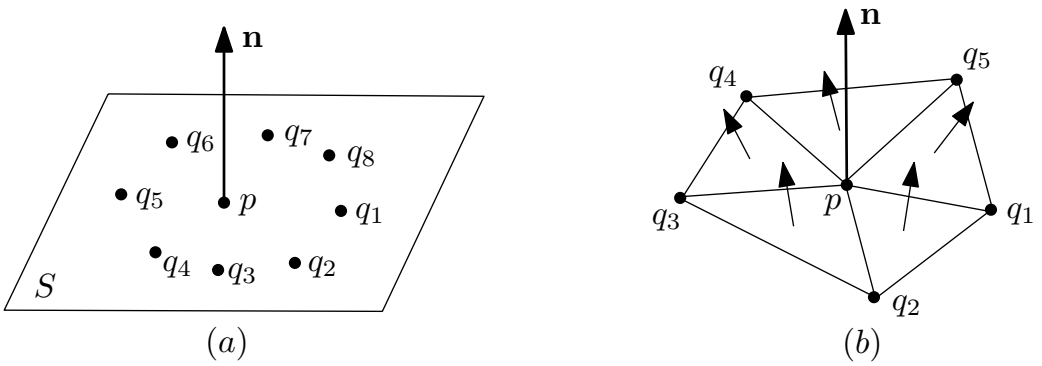

Figure 3: Illustration of two approaches for normal estimation: (a) a plane is fitted to a vertex p and its neighbors; (b) the normal vectors of triangles in one-ring of p are averaged.

distance of points to a local plane (see Fig. 3 (a)). While the averaging methods estimate the normal vector of one point by computing the weighted average of the normal vectors of the triangles in its one-ring neighbors, and the weight is equal to the inverse ratio of the areas or the surrounding angles of the triangles in its one-ring neighbors (see Fig. 3 (b)).

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 our system; however, considering the diversity in data formats of the existing databases, we make use of the former one.

Given a range image based face model represented by an $m \times n \times 3$ matrix as follows,

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

where $p_{ij}(x, y, z) = (p_{ijx}, p_{ijy}, p_{ijz})^T, (1 \leq i \leq m, 1 \leq j \leq n, i, j \in \mathbb{Z})$ 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 $n(p_{ij}(x, y, z)) = (n_{ijx}, n_{ijy}, n_{ijz})^T, (1 \leq i \leq m, 1 \leq j \leq n, i, j \in \mathbb{Z})$ denotes the unit normal vector of p_{ij}. As described in [39], the normal vector $\mathbf{N(P)}$ of range image $\mathbf{P}$ can be estimated by using 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 ($5\times$ 5 window is used in our paper) of point p_{ij} and $d = n_{ijx}p_{ijx}+n_{ijy}p_{ijy}+n_{ijz}p_{ijz}$. To simplify, each normal component in equation (2) can be represented by an $m \times n$ matrix:

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

where $\|(n^x_{ij}, n^y_{ij}, n^z_{ij})^T\|_2 = 1$.

Fig. 4 shows one example of the three normal component matrices (images) estimated from a given range image. The range image is sampled from the 3D-TEC database. From the figure, intuitively, we can see that normal component images contain more informative geometric information than their corresponding range image. For example, the geometric information around the eyes, mouth and forehead regions are highlighted.

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 encode each normal component, x, y, and z respectively as local normal patterns (LNP). Thanks to the matrix form of these normal components in equation (4), we can encode and characterize each of them using the similar way of feature extraction as to 2D texture images. Based on this kind of matrix form of normal representation, it 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 located in the same way as pixels in 2D images. Specifically, the value of every point in each 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 the neighborhood of LNP, where the notation $Q_{n,m}$ denotes a

neighborhood of m sampling points on a circle of radius of n.

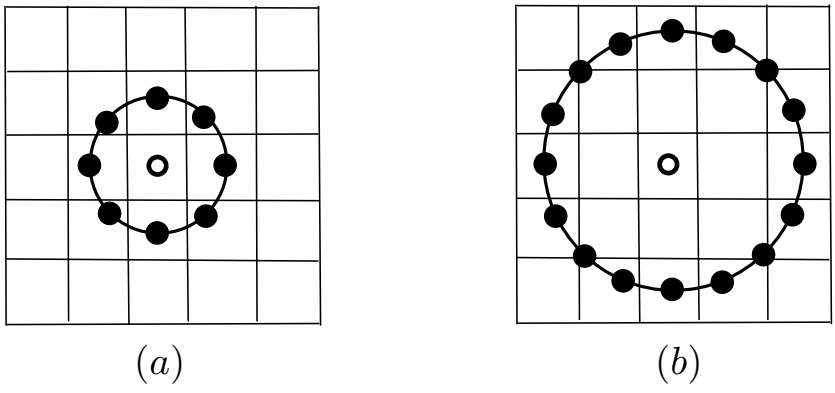

Figure 5: Examples of the neighborhood of LNP: (a) $Q_{1,8}$ and (b) $Q_{2,16}$.

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 derived binary numbers 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 $t(x) = 1, \text{if } x \geq 0$ and $t(x) = 0, \text{if } x < 0$.

$LNP(Q_{n,m})$ encodes local normal variations of each normal component as decimal values, denoted as $e([n_{ij}^k]_{m\times n}), k \in \{x, y, z\}$. See Fig. 6 for one example of $LNP(Q_{1,8})$ on three facial normal components of the same subject.

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 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 $e([n_{ij}^k]_{m\times n})$ 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; for $Q_{1,8}, R = 2^8 = 256$. $I\{A\} = 1$, if A is true, else $I\{A\} = 0$. This histogram contains the local micro-patterns of normal component over the whole face model.

3.3. Facial Normal 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 LNP histograms H are extracted and then concatenated by facial configuration to form a global histogram G to represent 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.

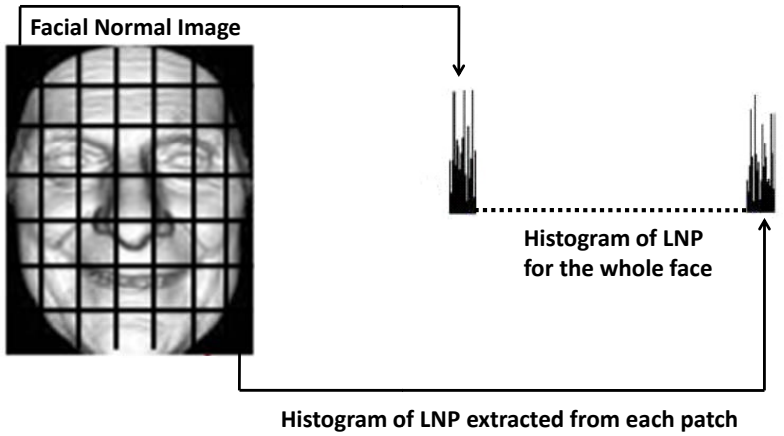

Figure 7: Illustration of facial normal representation: histogram of LNP.

4. Weighted Sparse Representation-based Classifier

In this section, we first introduce 3D face subspace based sparse representation model and its corresponding SRC. Then, we formulate the 3D face subspace and patch weight based weighted sparse representation model and its corresponding W-SRC. The effect of W-SRC will be proved in the following section.

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

Based on 2D face subspace model, 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. Wright et al. [40] proposed a sparse representation model and its corresponding SRC for robust 2D face recognition. In 3D case, we assume that a well-aligned frontal 3D face scan under different facial expressions approximately lies 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 as 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, $[v_{i,1}, v_{i,2}, \ldots, v_{i,n_i}] \in \mathbb{R}^{m \times n_i}$, according to (7), any test sample $y \in \mathbb{R}^m$ of i-th subject can be represented 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, there is only one training sample of each subject (i.e. the gallery) according to the experimental setting of the state-of-the-art 3D face recognition. Meanwhile, without occlusion, the only difference between two frontal well-aligned 3D face scans from the same subject is the local shape distortion caused by expression variations. This problem of insufficient training samples plus 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.

Based on model (9), sparse representation model and its corresponding SRC for 3D face recognition can be modeled as follows. Considering the whole gallery set with n 3D faces, 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:

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

In practice, we employ Orthogonal Matching Pursuit (OMP) [41] algorithm to solve (11) and compute the reconstruction residuals:

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

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

4.2. Weighted Sparse Representation-based Classifier

Assume each face scan is divided into K different patches, denote w_k as the learned weight for patch k, according to 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 $v_{ik} \in \mathbb{R}^{(m/K)\times 1}$, 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 probe y can be denoted as

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

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

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

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

The corresponding weighted reconstruction residuals

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

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

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

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 equation (15) equals to

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

Eq. (16) can be solved by the OMP [41] algorithm. Once we achieve 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. Experimental Settings and Results

5.1. Databases and Preprocessing

In our experiments, three databases, namely FRGC v1.0 [2], BU-3DFE [42] and Bosphorus [43], 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 [20] are used as testing sets for cross database validation and evaluation. Raw samples from each of these databases are displayed in Fig. 8 and the database introductions are briefly given as follows.

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 FRGCv1.0 is used for data acquisition, and the resolution of each range image is also 640×480.

- ***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.

- ***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 $1,600 \times 1,200$.

- ***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 [19].

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. [5]. The preprocessing pipeline contains: spike and noise removing, holes filling, nose tip localization and face cropping. As introduced in [5], 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. And then, we perform curvature analysis-based coarse search and generic face model-based fine search steps to locate the nose tips (for 3D-TEC, the manually labeled nose tips provided by the database are used). 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 project as range images by interpolation algorithm. Then, we also perform nose tip localization and face cropping steps for all the scans of BU-3DFE and Bosphrous databases by using the same tool. Then, for each of the five databases, we select a face scan with neutral expression and frontal pose to be a reference model, and all the other face scans are aligned to the reference model using the Iterative Closest Point (ICP) [44] algorithm, See Fig. 9 for some examples of the preprocessed face models.

5.2. Experimental Settings

To comprehensively evaluate the proposed approach, six experiments are designed. 1) The discriminative power of the proposed LNP descriptor; 2) The effectiveness of SRC; 3) The patch weights learning and the effectiveness

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

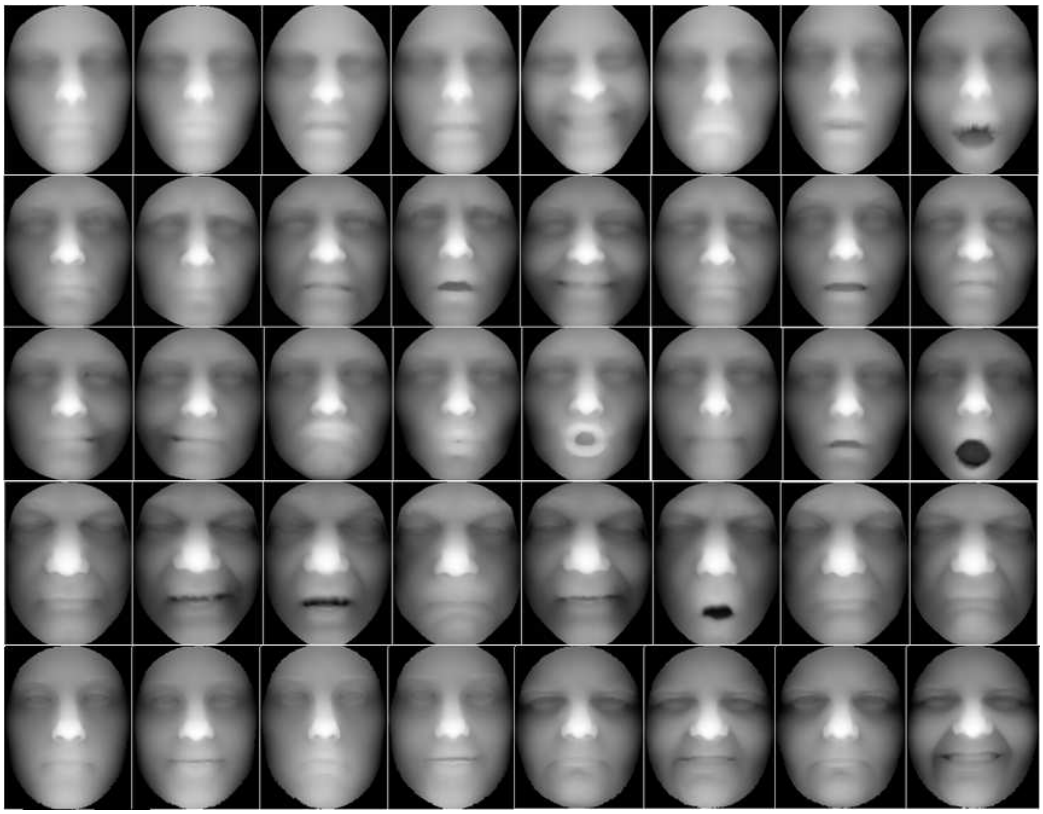

Figure 9: Illustration of the preprocessed face models: first row: models of one subject with different facial expressions (BU-3DFE); second and third rows: models of one subject with different facial expressions and action units (Bosphorus); forth row: models of one subject with different facial expressions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

of W-SRC; 4) The robustness analysis of facial expression variations; 5) The performance of distinguishing identical twins; 6) The comparison with the state-of-the-art.

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 probes; 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, since expression-robust 3D face recognition is stressed in this paper, we select the first neutral scans to make a gallery set, and the remaining scans with frontal pose and without occlusions to be a probe set. Table 1 shows the summary of these protocols, and presents the sizes of gallery sets and probe sets of these 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 [19]. 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 probes compose the probe set. Case IV reverses these roles of Case III. As pointed out in [19], 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 [19].

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 as 120×96 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, these three kinds of local patches corresponding to three normal encoding scales: i.e., 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. Thus, for each normal component, we encode it with three

different scales, achieving three histograms of local normal patterns (LNP). Similar to LBP, in order to reduce the dimensionality of final facial features, the uniform pattern strategy [32] 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 to compute the final accuracy of MSMC-LNP. To solve (11) and (16), the Orthogonal Matching Pursuit (OMP) [41] algorithm with the sparse number $L = 30$ of the sparse representation coefficient $\hat{x}$ is used for all experiments.

5.3. Experimental Results

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

To highlight the discriminative power of the proposed LNP based facial feature, we compare it with other two kinds of facial features: i) The original normal information based facial features Nx, Ny and Nz, 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 binary Patterns (LSP), i.e. LBP histograms extracted directly on 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, noted as LNPx, LNPy and LNPz, and their fusion, i.e. Multi-Component Local Normal Patterns (MC-LNP). All features were finally fed into SRC classifier. Noted that in this work, 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) $\text{LSP}_{2,16}$ + 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%**

Tab. 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 classifier

For histogram based feature vector (e.g. LNP), Chi-Square distance is the preferred similarity measurement [45]. Tab. 4 compares 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 by LNP using the same encoding scale (i.e. $Q_{2,16}$).

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%**

As it can be seen from Tab. 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 were used for learning the patch weights: FRGC v1.0, BU-3DFE, and Bosphorus. The experimental protocol introduced in Tab. 1 is used, and according to the proposed framework, the patch weights are achieved by the following four steps: 1) divide each normal component into local patches (10×8, 6×6, and 3×3 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 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 respect to the three encoding scales are 10×8, 6×6 and 3×3. Bosphorus database is used for the training database. 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 similar to each other among different normal components and different encoding scales, which have 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, take 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

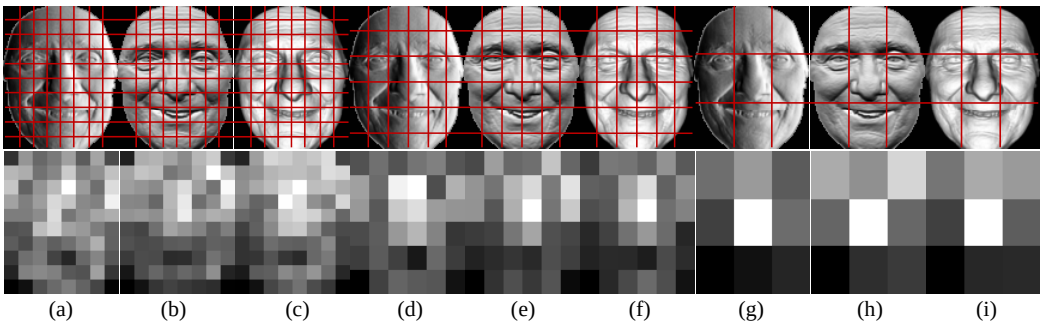

Figure 10: Illustration of the patch weights learned from the Bosphorus database. Columns (a-c), normal images x, y and z and their patch weights (10×8 patches); columns (d-f), normal images x, y and z and their patch weights (6×6 patches); columns (g-i), normal images x, y and z and their patch weights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

in 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 as compared with Fig. 1 (b). The differences are probably caused by the different data form between 2D and 3D faces; for example, the nose region of the 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 Tab. 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.

Tab. 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 LNPx($Q_{2,16}$), LNPy($Q_{2,16}$), and LNPz($Q_{2,16}$) and the one

Table 5: Comparison of rank-one score improvements on the FRGC v 2.0 database: patch weights are learned from 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%**

of their fusion MC-LNP($Q_{2,16}$) are reported. We can see from Tab. 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.

Tab. 6 presents the rank-one scores on the BU-3DFE database using S-RC 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 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%**

Tab. 7 presents the rank-one scores on the Bosphorus database using SRC and BU-W-SRC. We can see that BU-W-SRC largely 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 degradation influenced by 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 [12], [14] and [46], 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 , is utilized to analyze the robustness to facial expression variations. From Tab. 8, we can see 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 worse than [12, 14, 46], while the degradations are competitive to them.

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. [12]	99.0%	86.7%	12.3%
(2) Huang et al. [14]	99.1%	92.5%	6.6%
(3) Huang et al. [46]	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), in this experiment, we also evaluated the degradation influenced by the intensities of facial expression variations on the BU-3DFE database. We divided all the

probe faces 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 prototypic expressions.

Table 9: Comparing the degradations of rank-one scores influenced by changing the facial expression intensities 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 Tab. 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, 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. All the three degradations using BO-W-SRC are smaller than the ones using SRC, indicating the improved robustness to expression changes of the weighted sparse representation strategy.

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 that there are only neutral and smile scans in the 3D-TEC dataset, we designed another special training set 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 Tab. 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 Tab. 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 con-

tains identical twin samples 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 reason, 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 evaluation the performance of the proposed method. We display a comprehensive comparisons of the rank-one recognition rates on the FRGC v2.0, Bosphrous, BU-3DFE, and 3D-TEC databases. All results are shown in Tab. 11, where we highlight our best results and the results which are better than ours. From this table, we can find that:

(i) There are many results reported on the FRGC v2.0 database (here we just list some of them), while a very limited results reported on the other three databases. Note that the BU-3DFE and Bosphrous databases are initially designed for 3D facial expression recognition. The samples display informative expression and expression intensity variations. The performance of existing 3D face recognition methods are not well evaluated on these two databases. To the best of our knowledge, except our method, only (8-a) and (8-b) report their results on all the four databases.

(ii) Similar to our method, facial normal information is also used in (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), the authors use surface normal

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. [47]	91.9%	-	-	-	-	-	-
(2) Cook et al. [48]	92.9%	-	-	-	-	-	-
(3-a) Mian et al. [12]	93.5%	-	-	-	-	-	-
(3-b) Mian et al. [9]	96.2%	-	-	-	-	-	-
(3-c) Mian et al. [49]	93.8%	-	-	-	-	-	-
(3-d) Mian et al. [50]	**96.5%**	-	-	-	-	-	-
(4) Abate et al. (Normal) [30]	92.2%	-	-	-	-	-	-
(5-a) Faltemier et al. (ICP) [8]	**97.2%**	-	-	-	-	-	-
(5-b) Faltemier et al. (ICP) [a] [8] [51]	**98.0%**	-	-	93.5%	93.0%	72.0%	72.4%
(5-c) Faltemier et al. (ICP) [b] [8] [51]	**98.0%**	-	-	94.4%	93.5%	72.4%	72.9%
(6-a) Huang et al. (SI) [14]	91.8%	-	-	92.1%	93.0%	83.2%	83.2%
(6-b) Huang et al. (eLBP) [46]	**97.2%**	-	-	91.1%	93.5%	77.1%	78.5%
(6-c) Huang et al. (Range PFI) [52]	95.5%	-	-	91.6%	93.9%	68.7%	71.0%
(7) Gokbert et al. (Normal + Geometry) [31]	87.8%	-	-	62.6%	63.6%	54.2%	59.4%
(8-a) Kakadiaris et al. (Normal + Geometry) [11] [53]	**97.0%**	-	-	**98.1%**	**98.1%**	91.6%	93.5%
(8-b) Kakadiaris et al. (Normal) [c] [54]	**97.9%**	**98.2%** (2797/105)	**99.7%**	-	-	-	-
(8) Alyuz et al. [6]	**97.5%**	**98.2%** (2814/105)	-	-	-	-	-
(9) C. Maes et al. [55]	89.6%	**97.7%** (3186/105)	-	-	-	-	-
(10) H. Li et al. [26]		94.1% (4561/105)	-	-	-	-	-
(11) Queirolo et al. [21]	**98.4%**	-	-	-	-	-	-
(12) Spreeuwers et al. [18]	**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}, $E_{pkn}(p, g_k) = E_min(p, g_k) / \sum_{j=1, j \neq k}^{N} (E_min(g_i, g_k)/(N-1))$, where p is a probe image, g_k are the gallery images, and N is the number of gallery images. $E_min(p_1, p_2) = min(E(p_1, p_2), E(p_2, p_1))$, 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.

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, only around 60% rank-one scores are achieved. 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 face recognition method in the literature, in which the authors 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 [53, 54]. Compared with the proposed method, they obtained a better results on all the four databases except the Case III and Case IV over the 3D-TEC database.

(iii) Compared with the proposed method, methods (5-a), (5-b), (5-c) and (6-b) perform better on the FRGC v2.0 database while worse on the 3D-TEC database. the performance of method (10) is better on the Bosphrous database while worse on the FRGC v2.0 database.

(iv) The performance of method (9) is better than our method on both FRGC v2.0 and Bosphrous databases. And methods (3-d), (12) and (13) perform better than our method on the FRGC v2.0 database. Other methods not mentioned above perform worse than our method on the FRGC v2.0 database.

In conclusion, our method achieve a competitive rank-one scores on the FRGC v2.0, Bosphrous, BU-3DFE and 3D-TEC databases.

6. Conclusions and Future Works

This paper presented an expression-robust 3D face identification approach based on the proposed novel 3D facial surface descriptor, i.e. Multi-Scale and Multi-Component Local Normal Patterns (MSMC-LNP), and the displayed the effectiveness of patch-weight learning strategy for W-SRC. Our experimental results 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) Our system (i.e. MSMC-LNP + W-SRC) achieved competitive rank-one recognition rates on the FRGC v2.0, Bosphrous, BU-3DFE, and 3D-TEC databases.

In the further, we will focus on the following two aspects to improve the proposed method. 1) In this paper, we only stressed the robustness of the proposed face recognition method under expression variations. In more real scenarios, 3D face scans captured in unconstrained environments can depict not only facial expression variations but also arbitrary pose changes. Moreover, they can also be severely altered by external occlusions, e.g., hand, scarf, etc. In our further, we will evaluate the robustness of our method under pose and occlusion variations. 2) In this paper, we only presented

the identification results of the proposed method. Notice that there are few works on the study of sparse representation-based classifier for 2D and 3D face verifications. Reference [56] is perhaps the only work stressing this issue for 2D face verification. In our further, 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.

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] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] A. M. Bronstein, M. M. Bronstein, R. Kimmel, Expression-invariant representations of faces, IEEE Transactions on Image Processing 16 (1) (2007) 188–197.

[17] 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.

[18] 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.

[19] 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.

[20] V. Vijayan, K. W. Bowyer, P. J. Flynn, 3d twins and expression challenge, in: Proc. Int. Conf. on Computer Vision Workshops, 2011.

[21] G.G.Gordon, Face recognition based on depth and curvature features, in: Proc. Int. Conf. Computer Vision and Pattern Recognition, 1992.

[22] D. Sun, W. Sung, R. Chen, 3d face recognition based on local curvature feature matching, Applied Mechanics and Materials 121 (126) (2011) 609–616.

[23] G. Zhang, Y. Wang, Robust 3d face recognition based on resolution invariant features, Pattern Recognition Letters 32 (7) (2011) 1009–1019.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] T. Ahonen, A. Hadid, M. Pietikäinen, Face recognition with local binary patterns, in: Proc. Int. Conf. European Conference on Computer Vision, 2004.

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] R. Hoffman, A. K. Jain, Segmentation and classification of range images, IEEE Transactions on Pattern Analysis and Machine Intelligence 9 (5) (1987) 608–620.

[40] 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.

[41] 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.

[42] 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.

[43] 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.

[44] Z. Zhang, Iterative point matching for registration of free-form curves and surfaces, International Journal of Computer Vision 13 (2) (1994) 119–152.

[45] O. Pele, M. Werman, The quadratic-chi histogram distance family, in: Proc. Int. Conf. European Conference on Computer Vision, 2010.

[46] 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.

[47] 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.

[48] J. Cook, V. Chandran, C. Fookes, 3d face recognition using log-gabor templates, in: Proc. Int. Conf. British Machine Vision Conference, 2006.

[49] 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.

[50] 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.

[51] R. McKeon, Three-dimensional face imaging and recognition: A sensor design and comparative study, in: Ph.D. dissertation, University of Notre Dame, 2010.

[52] 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.

[53] 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.

[54] 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.

[55] 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.

[56] 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.