Signal Processing: Image Communication 57 (2017) 84–90

Contents lists available at ScienceDirect

Signal Processing: Image Communication

journal homepage: www.elsevier.com/locate/image

Pose-invariant 3D face recognition using half face

Yan Liang [a,b,*], Yun Zhang [a], Xian-Xian Zeng [a]

[a] Guangdong University of Technology, School of Automation, No.100, Waihuan Xi Road, Guangzhou Higher Education Mega Centre, Guangzhou 510006, China
[b] South China Normal University, School of Software, Nanhai Information Technology Park, Foshan 528225, China

ARTICLE INFO

Keywords:
3D face recognition
Pose variation
Facial landmark localization
Half face
Iso-geodesic stripes

ABSTRACT

Pose variations are still challenging problems in 3D face recognition because large pose variations will cause self-occlusion and result in missing data. In this paper, a new method for pose-invariant 3D face recognition is proposed to handle significant pose variations. For pose estimation and registration, a coarse-to-fine strategy is proposed to detect landmarks under large yaw variations. At the coarse search step, candidate landmarks are detected using HK curvature analysis and subdivided according to a facial geometrical structure-based classification strategy. At the fine search step, candidate landmarks are identified and labeled by comparing with a Facial Landmark Model. By using the half face matching, we perform the matching step with respect to frontal scans and side scans. Experiments carried out on the Bosphorus and UND/FRGC v2.0 databases show that our method has high accuracy and robustness to pose variations.

1. Introduction

3D face recognition has been an active research area during the last decade [1,2]. It has attracted many researchers due to its promising applications, such as surveillance, access control, and machine–human interaction [3]. Numerous face recognition algorithms based on 3D facial models have been proposed [4]. High accuracy can be obtained on frontal, expressionless views of faces. However, in cases of uncontrolled conditions such as occlusions, large pose and expression variations the system performances drop [5,6]. Particularly, variation of pose is one of the most challenging problems in 3D face recognition. It will cause self-occlusion, which results in missing data (see Fig. 1). In this work, we focus on the 3D face recognition problems caused by pose variations.

Most methods addressing pose variations were presented [8]. Zhang et al. [9] introduced a pose-independent face recognition approach which took advantage of the random forest-based pose estimation and the 3D morphable modelbased pose normalization. Both gallery and probe scans were normalized to the frontal pose. Therefore conventional face recognition method could be performed. Similar approach was presented in [10] and [11]. They first normalized scans in the probe to frontal view, and then matched the normalized probe to the gallery. For example, Panchal et al. [10] handled the pose correction problem with rotation matrix based on Euler angle. Then PCA was used to generate the feature vector of the frontal face and the recognition was done by the Euclidean distance algorithm. Cheraghian et al. [12] presented a face recognition method by combining 2D and 3D models. Features were extracted by Gabor wavelet from 2D and 3D scans, respectively. PCA and LDA were used for dimension reduction. 2D system and 3D system were combined in the decision level. Experiments were conducted on the FRAV3D database [13] and the average recognition rate was 95%. However, none of these methods was capable of handling extreme yaw variations.

Just a few methods were applicable to missing facial data resulting from pose self-occlusion For instance, Mahmood et al. [14]. presented a matching approach against large yaw variations (up to 60 degrees) An average nose model was built for alignment purpose. Facial surfaces were represented by local shape descriptors extracted from keypoints. However, instead of the whole face, they focused only on the nose region, which led to the loss of discriminative information. In [15], face models were represented by radial curves. To overcome the challenge of missing data they used a statistical model in the shape space of radial curves to complete the partially observed curves. The method worked well on scans with pose variations up to 35 degrees. However, the recognition rate decreased to 70.49% for scans from the right side and to 86.89% for scans from the left side. Moreover, a limitation of the approach is that the nose tips in side scans need to be annotated manually. Passalis et al. [16] used an automatic landmark detector to

* Corresponding author at: Guangdong University of Technology, School of Automation, No.100, Waihuan Xi Road, Guangzhou Higher Education Mega Centre, Guangzhou 510006, China.
E-mail addresses: 2005jinrongxue2@163.com (Y. Liang), yun@gdut.edu.cn (Y. Zhang), zengxianxian727@outlook.com (X.-X. Zeng).

http://dx.doi.org/10.1016/j.image.2017.05.004
Received 19 June 2016; Received in revised form 18 February 2017; Accepted 7 May 2017
Available online 12 May 2017
0923-5965/

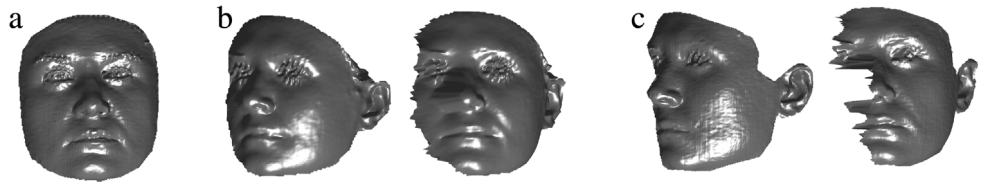

Fig. 1. Facial scans from the Bosphorus database from a single subject [7]. (a) Frontal face scan. (b) 30 degree right side scan, and (c) 45 degrees right side scan. In (b) and (c) the original side scan and the pose normalized frontal view are shown. Note the missing data in (b) and (c).

estimate pose and detect occluded areas. Facial symmetry was used to solve the missing data problem The method could be applied to scans with up to 80 degrees yaw variations. However, because half of the geometry and normal image is the mirror of the other half there is redundant information for side scans. Additionally, to deal with pose variations, they used three FLMs (i.e. FLM8, FLM5L, and FLM5R) for landmark localization, which increased the complexity of the calculation. Berretti et al. [17] dealt with missing data by sparse comparison of facial curves defined across inlier pairs of matching keypoints between probe and gallery scans. The performance of the method was evaluated on the combined UND/FRGC v2.0 dataset [18,19] and on the Gavab database [20]. However, the representation of facial curves depended on the depth information. It affected the accuracy of curve matching.

Several local feature-based SIFT-like matching methods have been recently proposed to handle the problem of missing data due to occlusion or pose variations [21–23]. Berretti et al. [21] used the meshDOG as detector of keypoints of the facial scan and built the HOG, SHOT, and GH local descriptors for 3D meshes. Li et al. [22] provided two principal curvature-based keypoint detectors to detect feature points on 3D face scans. They designed three keypoint descriptors and fused them to describe the local shape of the keypoints. The matching was performed through a multi-task sparse representation-based algorithm. Promising results were reported on the Bosphorus database [24] and the FRGC v2 database. However, a common limitation of these methods is the large number of keypoints, which results in a demanding computation cost.

In our previous work [25], we presented a 3D face recognition method using 3D Weighted Walkthrough (3DWW) [26,27] and centroid distance. A 3D face scan was partitioned into a set of isogeodesic stripes and the facial features were described by the 3DWWs and centroid distances of each pair of stripes. Good results were obtained for scans with expression variations However, the method was not able to deal with problems caused by pose variations, since under these cases facial scans might have missing areas and the values of facial features would definitely be affected.

In this paper, we extend our previous work to present a face recognition approach which can handle significant pose variations. The method relies on three fundamental components: (a) A landmark detection step using a coarse-to-fine strategy. (b) A registration step to perform face alignment. (c) A half face matching step to handle missing data Different stages and components of our method are laid out in Fig. 2. The experimental results on the Bosphorus and UND/FRGC v2.0 databases demonstrate the effectiveness of our method

The main contributions of this paper are the following.

(1) We propose a coarse-to-fine strategy for 3D landmark detection in the presence of large yaw variations. This strategy consists of facial geometrical structure-based coarse search and modified Facial Landmark Modelbased fine search. It can automatically locate landmarks on the facial scans with missing data.
(2) To deal with missing data caused by large yaw variations, a half face matching algorithm is proposed, which can handle side scans as well as frontal scans.

The rest of this paper is organized as follows. Section 2 describes the proposed approach for 3D face recognition. Section 3 analyzes the experimental results. Comments and conclusions are provided in Section 4.

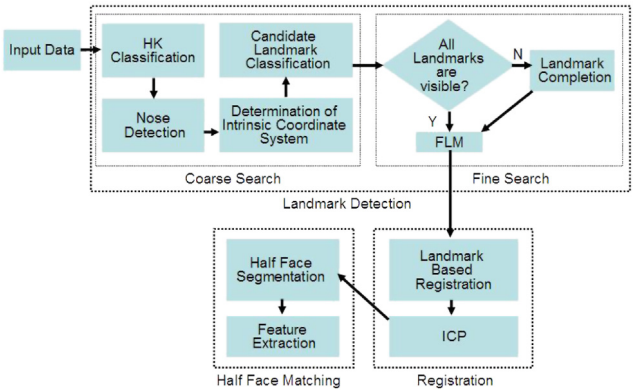

Fig. 2. Overview of the proposed method.

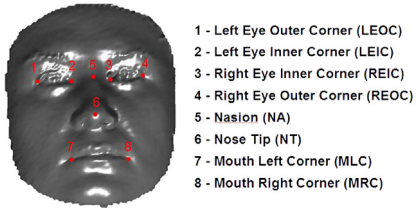

Fig. 3. The 8 landmark points on a facial scan.

2. Method

2.1. 3D Facial Landmark Detection

We use 8 landmarks for registration i.e. left eye outer corner (LEOC), left eye inner corner (LEIC), right eye inner corner (REIC), right eye outer corner (REOC), nasion (NA), nose tip (NT), mouth left corner (MLC), and mouth right corner (MRC), on a facial scan (Fig. 3). Note that large yaw variations will make some landmarks invisible. We propose a coarse-to-fine strategy for 3D landmark detection to solve this problem. At the coarse search step, candidate landmarks are detected and further subdivided according to facial geometrical structure. At the fine search step, a Facial Landmark Model (FLM) is used to localize landmark points. Specifically, candidate landmarks are detected by HK classification firstly. Secondly, nose tip and nasion are localized as auxiliary points. Thirdly, an intrinsic coordinate system is defined by nose tip, nasion and the normal vector at nose tip According to the position in the coordinate system the candidate landmarks can be subdivided. Meanwhile, we can determine which landmark point is visible. Finally, if eight landmarks are visible, candidate landmark sets are matched with the FLM directly. Otherwise, we consider the existence of large yaw variations Candidate landmark sets are complemented and then matched with the FLM.

2.1.1. Facial geometrical structure-based coarse search

Feature points can be detected from the geometric perspective. For example, nose tip appears as an elliptical convex point while eye

inner corner as elliptical concave point. There are two main kinds of geometry-based methods which make use of HK classification (H stands for Mean curvature and K stands for Gaussian curvature) and shape index respectively [28,29]. Compared with shape index HK curvature has less calculation and it reflects the curved degree. Therefore, our approach uses HK classification to detect candidate landmarks.

The elliptical convex points, elliptical concave points, hyperbolic concave points, and hyperbolic convex points are detected on a facial surface according to HK classification which are candidate landmarks for nose tip, eye inner corner/mouth corner, eye outer corner, and nasion respectively. To reduce the number of candidates a thresholding process is used. Details of HK classification and the thresholding process can be found in [30] and [28].

In this way, candidate landmark points can be detected and divided into four classes: nose tip, eye inner corner/mouth corner eye outer corner, and nasion However, since candidate landmarks of mouth corner and eye inner corner are elliptical concave, they belong to the same class In addition, it cannot catch the relative position, e.g. candidate landmarks of left and right eye outer corner are in the same class. Therefore, the HK classification alone is insufficient. Candidate landmarks need to be further classified.

Notice that nose is one of the most stable regions under pose and expression variations and it is also easier to be detected [31]. Once the nose is coarsely localized, we can further classify the candidate points according to the priori knowledge on facial shape. For example, for a frontal facial scan, eyes are always above nose and mouth is always below nose. Here, we propose a facial geometrical structure-based classification strategy for candidate landmarks subdivision. Firstly, the areas of nose tip and nasion are detected. The centroids of the two areas are regarded as the coarse positions of nose tip and nasion. Secondly, an intrinsic coordinate system is defined by nose tip, nasion and the normal vector at nose tip. Then, according to the position in the coordinate system, the candidate landmarks can be further subdivided.

In order to identify the nose, we localize nose tip and nasion as auxiliary points. The candidates of nose tip and nasion have been detected by using the HK classification. However, the candidate points are discrete. Furthermore, there are a large number of outliers. Therefore, we first perform mesh segmentation to detect the areas of nose tip and nasion.

let $M = (V, T)$ be a 3D mesh, where $V = \{p_i | p_i = (x_i, y_i, z_i), 1 \le i \le N\}$ is the set of mesh vertices and $T = \{t_{ijk} | 1 \le i \le N, 1 \le j \le N, 1 \le k \le N, i \ne j \ne k\}$ is the set of triangles formed by adjacent vertices. N is the number of mesh vertices. $AV(p_i)$ is the set of adjacent vertices of p_i. $C = \{c_i\}$ is the set of HK classification labels. If the values of Mean and Gaussian curvature of p_i conform to a certain combination, $c_i = 1$, otherwise $c_i = 0$. The method of mesh segmentation is summarized in *Algorithm 1*.

Algorithm 1. Mesh segmentation

Input: a 3D mesh M, the set of adjacent vertices AV and the set of HK classification labels C.

Procedure:

(1) Set a label l_i for each vertex. Initialize $l_i = 0$ and the number of connected regions $n = 1$.
(2) Choose a vertex p_i which $l_i = 0$ and $c_i = 1$. Set $p = p_i$ and $l_i = 1$. p_i is added to the set of connected regions RE_n.
(3) For each vertex $p_j \in AV(p)$, if $c_j = 1$ and $l_j = 0$, p_j is added to the set of connected regions RE_n.
(4) For each unprocessed vertex $p_k \in RE_n$. Set $p = p_k$ and $l_k = 1$. (Repeat step (3)).
(5) $n = n + 1$. Repeat steps (2), (3), and (4) until every vertex has been processed.

Output: the set of connected regions RE_n.

We usually get more than one connected region because of outliers. But the nose tip and nasion connected regions are the largest (having the largest number of points). Fig. 4 shows examples of nose tip and nasion connected regions in red and outlier connected regions in green. Therefore, we choose the largest peak connected region and saddle ridge connected region as nose tip area and nasion area, respectively.

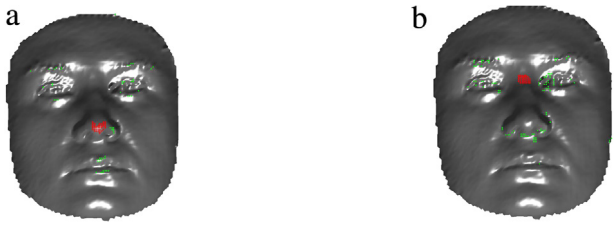

Fig. 4. Connected regions of high curvature: (a) peak connected regions (nose tip connected regions in red and outlier connected regions in green), (b) saddle ridge connected regions (nasion connected regions in red and outlier connected regions in green).

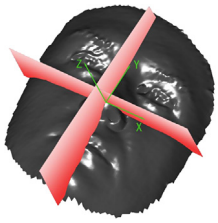

Fig. 5. An intrinsic coordinate system defined by nose tip, nasion, and the normal vector at nose tip.

Once the areas of nose tip and nasion are detected, the centroid of nose tip area G_{tip} and the centroid of nasion area G_n are computed as the coarse positions of nose tip and nasion, respectively.

Nose tip, nasion, and the normal vector at nose tip are needed to define an intrinsic coordinate system. Note that the centroid of nose tip area G_{tip} may not be located on the facial surface and there is no normal vector at G_{tip}. Thus we choose the closest vertex on the surface, which is denoted by G'_{tip}, to replace G_{tip}. The normal vector at G'_{tip} is given by

$$NV = PD_1 \times PD_2, \tag{1}$$

where PD_1 and PD_2 are the two principal directions at G'_{tip}. Then, the intrinsic coordinate system can be defined by G'_{tip}, G_n, and NV. Its origin is G'_{tip}.

As can be seen from Fig. 5, a facial surface is divided into four areas by the XZ and YZ planes. Left eye outer/inner corner is located above the XZ plane and on the left of the YZ plane. Right eye outer/inner corner is located above the XZ plane and on the right of the YZ plane. Mouth left corner is located below the XZ plane and on the left of the YZ plane. Mouth right corner is located below the XZ plane and on the right of the YZ plane.

Thus, according to the position in the coordinate system, the candidate landmarks obtained by the HK classification are subdivided into eight classes: nose tip, nasion, left eye outer corner, right eye outer corner, left eye inner corner, right eye inner corner, mouth left corner and mouth right corner.

The facial geometrical structure-based classification strategy is beneficial to improve the efficiency of landmark localization. In addition, the candidate points which do not conform to the facial geometrical structure will be filtered out. An example of candidate landmark classification is illustrated in Fig. 6. Hyperbolic concave regions are first detected by the HK classification. Then, according to the position in the intrinsic coordinate system, we can easily distinguish the candidates of left eye outer corner and right eye outer corner. The candidate points, which are located below the XZ plane, cannot be the eye outer corners. Therefore they will be regarded as outliers and filtered out.

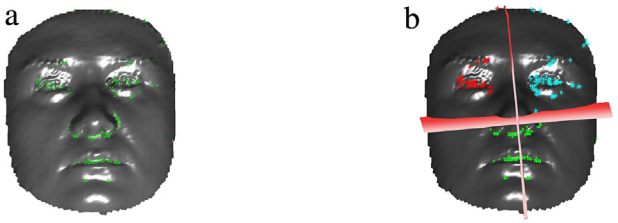

Fig. 6. An example of candidate landmark classification: (a) hyperbolic concave regions (in green), (b) candidates of left eye outer corner (in red), candidates of right eye outer corner (in blue), and outliers (in green).

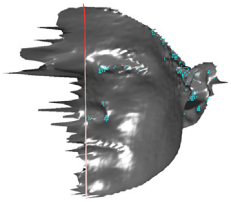

Fig. 7. Detected elliptical concave regions at a left side scan (the pose normalized frontal view). Elliptical concave regions are labeled in blue.

2.1.2. Modified Facial Landmark model-based fine search

Once the candidate landmarks are detected and subdivided into eight classes, we proceed to a fine search step by matching them with a Facial Landmark Model (FLM) [16]. In the traditional FLM-based method, the plausible landmark combinations are selected by matching with the FLM. They must have the same dimensions. However, eight landmarks are visible on frontal facial scans, while only five landmarks are visible on side facial scans. The traditional FLM-based method cannot handle frontal and side facial scans simultaneously. Perakis et al. [16,29] used three FLMs to deal with this problem, i.e., FLM8 is used for frontal scans, while FLM5L and FLM5R are used for side scans. This increased the computational burden.

In this work, candidate landmarks have been subdivided into eight classes according to the facial geometrical structure-based classification strategy. Candidate landmark set is an empty set, which means the corresponding landmark is invisible. Fig. 7 shows a left side scan. Elliptical concave regions which correspond to eye inner corners and mouth corners are labeled in blue. It is clear that there is no elliptical concave region on the left of the *YZ* plane. The candidate landmark sets of left eye inner corner and mouth left corner are empty sets. Therefore, these two landmarks are likely invisible. If eight landmarks are visible, candidate landmark sets are matched with the FLM directly. Otherwise, candidate landmark sets should be complemented first by mirroring before matched with the FLM. Our method only uses one FLM to deal with frontal and side facial scans simultaneously, which greatly decreases the complexity of the calculation.

The modified FLM-based fine search contains the training phase and the detection phase. In the training phase, a FLM is created to represent the landmark positions [16,29]. We manually labeled the target landmarks for each facial scan of the training set. Each example is represented by a shape vector of concatenated 3D coordinates:

$$\Omega = [\omega_1, \omega_2, \ldots, \omega_m]^T, \tag{2}$$

where $\omega_i = (x_i, y_i, z_i)$ is the coordinate of each landmark and m is the number of landmarks. In this paper, $m = 8$.

Shape has to be invariant to 3D euclidean similarity transformations. Therefore we use Procrustes analysis to align the training shapes to their mean shape. It is performed by minimizing the Procrustes distance D:

$$D^2 = |\Omega^k - \overline{\Omega}|^2 = \sum_{i=1}^{m} (\omega_i^k - \overline{\omega_i})^2, \tag{3}$$

where Ω^k is the kth training shape. $\overline{\Omega} = \frac{1}{L}\sum_{k=1}^{L}\Omega^k$ is the mean shape of all L examples in the training set.

PCA is then applied to the aligned shape vectors to dimensionality reduction [32,33]. If ϕ contains the f eigenvectors ϕ_i corresponding to the f largest eigenvalues λ_i of the covariance matrix CM of the aligned shape vectors, then the FLM can be modeled by

$$\Omega' = \overline{\Omega} + \phi b, \tag{4}$$

where $b = [b_1, b_2, \ldots, b_f]$ is a set of parameters controlling the shape variation of FLM. The value of f is typically determined by the number of components which is required to account for 99% of the training variation, and it is computed as

$$\frac{\sum_{i=1}^{f}\lambda_i}{\sum_{j=1}^{3m}\lambda_j} \geq 0.99. \tag{5}$$

In the detection phase, the candidate landmarks are identified and labeled by being matched with the FLM. Let α be the number of nonempty candidate landmark sets of a test facial scan. The nonempty landmark class is referred to as L_N and the empty landmark class is referred to as L_E. We create combinations Ψ by selecting one candidate landmark point from each L_N class. Notice that if Ψ and $\overline{\Omega}$ have different dimensions, Ψ cannot be aligned to $\overline{\Omega}$ and be projected onto the PCA feature subspace directly. Thus, in the case of $\alpha < 8$, candidate landmark sets are complemented first by mirroring. Specifically, the right eye corners are obtained by mirroring the left eye corners and the mouth right corner are obtained by mirroring the mouth left corner, and vice versa. We use Ψ' instead of Ψ, where Ψ' contains the α points of Ψ and the other 8-α points obtained by mirroring.

Ψ' is aligned to $\overline{\Omega}$ by minimizing the Procrustes distance. Project Ψ' onto the PCA feature subspace and its deformation parameters can be computed as

$$b = \phi^T \cdot (\Psi' - \overline{\Omega}). \tag{6}$$

The probability of Ψ which belongs to the shape class is computed as

$$\Pr(\Psi) = \frac{\sum \lambda_i}{E_{sum}}, \tag{7}$$

where λ_i are the eigenvalues that satisfy the deformation constraints $|b_i| \leq 3\sqrt{\lambda_i}$ and E_{sum} is the sum of the eigenvalues of the FLM. If $\Pr(\Psi)$ exceeds 0.99, the landmark set Ψ is considered plausible, otherwise it is rejected.

Finally, the plausible landmark set which has the minimum Procrustes distance from FLM is selected as the landmark localization result.

2.2. Registration

In order to normalize the test scans to a common pose and define them in the same coordinate system, we register the test scans with a reference model. The reference model is obtained by averaging a selected set which is composed of neutral frontal scans. Here, we use the landmarks detected in Section 2.1 for initializing the registration process. Then the ICP algorithm is used to refine the registration.

Let Ω_{ts} and Ω_{md} be the set of landmarks on a test scan and the corresponding landmarks on the reference model, respectively. The nose tip of Ω_{md} is located at the origin. If Ω_{ts} is translated by T so that its nose tip is at the origin and rotated by R to minimize the Procrustes distance between Ω_{ts} and Ω_{md} then the transformation to register a test face with vertices p_i to the reference model is computed by

$$p_i^* = R \cdot T \cdot p_i. \tag{8}$$

After initial registration, a fine registration is obtained by the ICP algorithm, whose goal is to minimize the mean square error between the two faces. For more details about ICP algorithm please refer to [34].

Once the test scans are registered with the reference model, they are normalized to a common pose. Meanwhile, according to the yaw angle,

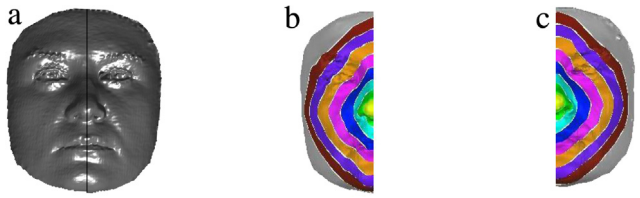

Fig. 8. The representation of the frontal scan. (a) A frontal scan is divided into two parts with respect to the symmetry plane. (b) Calculated isogeodesic stripes on the left part of the face. (c) Calculated isogeodesic stripes on the right part of the face.

we can divide the scans into three categories, i.e., frontal, left side, and right side. In our experiments, the scan whose yaw angle between −20 and +20 degrees is marked as frontal scan. Otherwise, the scan is marked as side scan.

2.3. Half face matching

3DWW is a kind of descriptor which can capture the relative displacement between two 3D spatial entities [26,27]. It was first proposed by Berretti Motivated by the work of Berretti, in Ref. [25], we presented a 3D face recognition method by using 3DWW and centroid distance. A 3D face was first partitioned into a set of equal width isogeodesic stripes starting from the nose tip. Then the 3DWWs and centroid distances of each pair of stripes were calculated, which described the spatial distributions information and the spatial Euclidean distance information, respectively. Finally, the similarity between two facial scans was measured by comparing the 3DWWs and centroid distances of homologous pairs of the two faces.

This method works well for scans which have frontal poses and various facial expressions. However, it cannot handle missing data caused by self-occlusion in non-frontal poses. To account for this problem, we improve the method in [25] by incorporating the notion of *half face matching*. To the scan with large yaw rotations, we only use the complete half face for matching. The other half face with missing data is discarded. Human face is approximately bilaterally symmetrical [35]. Given the left or right side of a facial scan, the other side can be obtained by mirroring. Therefore, side scan has enough discriminative information to distinguish different identities. The experimental results presented in Section 3 justified the theory.

The matching step can now be directly applied to the left and right side scans to handle missing data. Specifically we partition each scan into two parts, i.e. left and right, with respect to the vertical symmetry plane defined by nose tip, nasion and the normal vector at nose tip [27]. If the probe scan is classified as left side, we only compute the 3DWWs and centroid distances between stripes on the left part of the face. Differences between two scans are measured by comparing the 3DWWs and centroid distances of homologous pairs of the left parts, and vice versa for right side scans For frontal scan, we regard it as a pair of independent left and right side scans and the two side scans are computed respectively. The final dissimilarity measure is obtained by averaging distances in the two parts (Fig. 8).

3. Experimental results

3.1. Databases

3.1.1. Bosphorus database

The Bosphorus database [7] is utilized to evaluate the system performance. It consists of 4666 facial range scans from 105 subjects, including expression, pose, and occlusion variations. The scans are given as point clouds of approximately 35000 points.

The Bosphorus database contains side scans with yaw rotations of 10, 20, 30, 45, and 90 degrees. Note that the 90 degrees side scans have

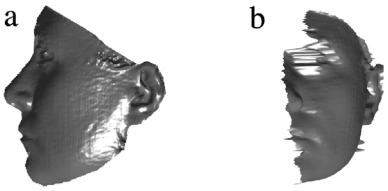

Fig. 9. 90 degrees side scan from the Bosphorus database. (a) Original side scan. (b) Pose normalized frontal view.

Table 1
Effect of pose variations on landmark localization accuracy.

Testing set	Mean (mm)	Std. (mm)	Success rate
Neutral	4.23	1.38	99.87%
YR10	4.26	1.42	99.52%
YR20	4.37	1.52	99.29%
YR30	4.55	1.71	98.48%
YR45	4.68	1.88	97.71%
PR	4.29	1.36	99.76%
CR	5.14	2.13	97.16%

massive missing data and contain no information of nose tip and nasion (Fig. 9). The coarse-to-fine strategy is invalid for the 90 degrees side scans. Therefore, we do not use them in the experiments. To test the effects of pose variations on the performance, we define six testing sets

(1) *Neutral*, contains neutral frontal scans (194 scans)
(2) *YR10* , contains scans with yaw rotations of 10 degrees (105 scans)
(3) *YR20* contains scans with yaw rotations of 20 degrees (105 scans)
(4) *YR30*, contains scans with yaw rotations of 30 degrees (105 scans)
(5) *YR45*, contains scans with yaw rotations of 45 degrees (210 scans)
(6) *PR*, contains scans with pitch rotations of 5 degrees and 10 degrees (419 scans)
(7) *CR*, contains scans with cross rotations (yaw rotations of 45 degrees and pitch rotations of 20 degrees) (211 scans).

3.1.2. UND/FRGC v2.0 database

In addition to the Bosphorus database, we used the FRGC v2.0 database [19] combined with the Ear database from the University of Notre Dame (UND), collections F and G [18]. The FRGC v2.0 database consists of 4007 scans collected from 466 subjects. These scans have frontal poses and various facial expressions. The UND database contains side scans with yaw rotations of 45, 60, and 90 degrees. The institution collecting the UND dataset is the same that defined the FRGC v2.0 dataset. There is a partial overlap between subjects in the two datasets.

3.2. Landmark detection results

The performance of the landmark detection algorithm was evaluated on the Bosphorus database firstly. The first neutral frontal scan of each subject is used to construct the training set (105 scans), which can be used to train a facial landmark model We manually annotate landmarks on the scans of testing sets as ground truth. The error is computed as the Euclidean distance between the automatically detected landmark and the corresponding manually labeled landmark. If the error is under a certain threshold ($T = 10$ mm), it is deemed to be a successful detection.

Table 1 shows the mean error, standard deviation, and success rates of landmark localization on probe sets with pose variations. The mean error is less than 5.14 mm and the standard deviation is less than 2.13 mm in all testing sets. The success rate of landmark localization is at least 97.16%. The results show that the proposed algorithm has high accuracy to pose variations. As can be seen from the table, least

Table 2
Comparison of localization results on the UND/FRGC v2.0 database.

Mean (std.) mm Method	Perakis et al. [29]	Our method
Testing set (scans)	FRGC v2 (975) + UND (117)	FRGC v2 (1500) + UND (410)
LEOC	6.06(4.13)	4.67(2.71)
REOC	5.32(3.71)	4.03(2.40)
LEIC	4.90(2.96)	4.52(2.79)
REIC	4.65(2.45)	4.53(2.41)
NT	4.41(2.68)	4.06(2.32)
MLC	4.91(2.88)	4.54(2.43)
MRC	5.01(2.97)	4.69(2.67)
CT	4.80(3.52)	4.66(2.83)

Table 3
Rank-1 recognition rates on the Bosphorus database.

Probe set	Recognition rate
Neutral	100%
YR10	100%
YR20	98.10%
YR30	97.14%
YR45	93.33%
PR	99.52%
CR	94.79%

error is obtained with frontal scans, while the highest error is obtained with scans with cross rotations. This is because the latter has significant pose variations along the yaw axis and pitch axis. Many useful data are missing and the proportion of non-facial regions increases. More noise and outliers are brought in, which leads to the failure of landmark localization.

We also compare our method with Perakis' method [29] on the UND/FRGC v2.0 database. The FRGC v2.0 database is used for frontal facial scans and the UND database is used for side facial scans. We test our method on 1500 facial scans which are randomly selected from the FRGC v2.0 database. In addition, we test our method on the 45 degree side scans (118 left and 118 right from 118 subjects) and the 60 degree side scans (87 left and 87 right from 87 subjects) from the UND database. Table 2 shows the comparative results. Compared with [29], our method obtains smaller mean localization distance errors and standard deviations for all landmarks. Furthermore, we use more scans for testing. These results indicate that our method is more accurate and robust to pose variations.

3.3. Face recognition results

We evaluate the performance of the half face matching algorithm on the Bosphorus database. The first neutral frontal scan of each subject is manually divided into a pair of left and right side scans with respect to the vertical symmetry plane, which compose the gallery set. The probe sets are the same as the testing sets described in Section 3.1.1

The Rank-1 recognition rates are depicted in Table 3. In the case of *Neutral* and *YR10*, the proposed approach achieves a 100% rank-1 recognition rate. When the probe sets consist of scans with pose variations along the pitch axis, the rank-1 recognition rate is 99.52%. The results show that in the cases of pitch rotations, since the whole face is almost visible and the missing of facial data is limited, our method achieves high accuracy. When the probe sets consist of scans with yaw rotations of 30 and 45 degrees, the recognition rate is 97.14% and 93.33%, respectively. The results above show that our method can handle large yaw variations which lead to missing facial data.

Note that the accuracy decreases when the yaw angle increases. Specifically, the increase of the yaw angle occludes more facial data, which will lead to the failure of registration and matching. In addition, human faces are not perfectly symmetric. The use of only half of face has an impact on recognition rates.

3.4. Overall performance comparison

We compare the performance of the proposed 3D face recognition system with that of the state-of-the-art methods on the Bosphorus database. Results of the comparative evaluation are summarized in Table 4. The results of other methods are quoted from their papers.

From Table 4, we notice that our method outperforms [21] and [23]. Compared with [22], our scores are equal to it at the *Neutral*, *YR10*, and *PR* sets, whereas the *YR20*, *YR30*, *YR45*, and *CR* sets our scores are a little less than it. However, the large number of keypoints, high dimensionality of local descriptors, and the large number of keypoints matching lead to a demanding computation cost (about 75 s to process a facial scan), which reduces the practicability of Li's method [22]. We implement experiments on a PC with the CPU by Intel 530, 2.93 GHz, 8GB RAM. The average runtime for each step of our method is: landmark detection 10 s, registration 6 s, and half face matching 8 s. The overall processing time is about 24 s. Therefore, our method has achieved a compromise between computation cost and the accuracy of recognition.

4. Conclusions

In this paper, we present a new method for pose-invariant 3D face recognition. It addresses the problem of missing data caused by large yaw variations which is very common in real-world applications with unconstrained acquisition. A coarse-to-fine strategy is proposed to detect landmarks under large yaw variations. Then the landmarks are used for pose estimation and registration. By using half face matching, we can perform the matching step in the presence of missing data.

The landmark detection algorithm is tested on the Bosphorus and UND/FRGC v2.0 databases which both contain scans with large yaw variations. The results show that our algorithm is very robust and accurate for 3D facial landmark detection under large yaw variations. We also perform experiments on the Bosphorus database to evaluate the performance of the half face matching algorithm. When the probe sets contain frontal and almost-frontal scans (up to ±20 degrees yaw variation), the recognition rate is no less than 98.10% Even when the probe set contains scans with cross rotations (yaw rotations of 45 degrees and pitch rotations of 20 degrees), we achieve a 94.79% recognition rate, which represents a high performance

The limitation of our method is that half of the face with respect to the yaw axis is visible in the scan. In addition, the areas of nose tip and nasion cannot be occluded SIFT-like matching methods support partial face match (e.g., Berretti et al. [21], Li et al. [22], and Werghi et al. [23]). Instead of landmarks, they depend on keypoints. Local descriptors constructed at keypoints are compared to find the best matches. SIFT-like matching methods apply to the case that more than half of the face is missing. However the computation cost of these methods is quite high, as we have demonstrated it in Section 3.4. In contrast, our method can perform the identification task efficiently.

In future work, we plan to further improve the accuracy and robustness of the automatic landmark detection, so as to improve the overall performance of the system.

Table 4
Comparison of rank-1 recognition rates on the Bosphorus.

	Berretti et al. [21]	Li et al. [22]		Werghi et al. [23]	Our Method
		CGM	FGM		
Neutral	97.9%	100%	100%	100%	100%
YR10	–	100%	100%	98.10%	100%
YR20	–	99.05%	100%	89.52%	98.10%
YR30	–	98.10%	99.05%	73.33%	97.14%
YR45	–	90.95%	97.62%	46.67%	93.33%
PR	98.3%	98.81%	99.52%	92.84%	99.52%
CR	93.4%	94.31%	99.05%	85.37%	94.79%

References

[1] A.F. Abate, M. Nappi, D. Riccio, G. Sabatino, 2D and 3D face recognition: A survey,, Pattern Recognit. Lett. 28 (14) (2007) 1885–1906.
[2] O. Ocegueda, T. Fang, S.K. Shah, I.A. Kakadiaris, 3D face discriminant analysis using gauss-markov posterior marginals, IEEE Trans. Pattern Anal. Mach. Intell. 35 (3) (2013) 728–739.
[3] F. Al-osaimi, M. Bennamoun, A. Mian, An expression deformation approach to non-rigid 3D face recognition, Int. J. Comput. Vis. 81 (2009) 302–316.
[4] D. Huang, M. Ardabilian, Y. Wang, L. Chen, 3-D face recognition using eLBP-based facial description and local feature hybrid matching, IEEE Trans. Inf. Forensics Secur. 7 (5) (2012) 1551–1565.
[5] N. Alyuz, B. Gokberk, L. Akarun, 3-D face recognition under occlusion using masked projection, IEEE Trans. Inf. Forensics Secur. 8 (5) (2013) 789–802.
[6] L. Ballihi, B.B. Amor, M. Daoudi, A. Srivastava, D. Aboutajdine, Boosting 3-D-Geometric features for efficient face recognition and gender classification, IEEE Trans. Inf. Forensics Secur. 7 (6) (2012) 1766–1779.
[7] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gokberk, B. Sankur, L. Akarun, Bosphorus database for 3D face analysis, in: Biometrics and Identity Management, 2008, pp. 47–56.
[8] A.F. Abate, M. Nappi, D. Riccio, G. Sabatino, 2D and 3D face recognition: A survey, Pattern Recognit. Lett. 28 (14) (2007) 1885–1906.
[9] W. Zhang, D. Huang, D. Samaras, J. Morvan, Y. Wang, L. Chen, 3D assisted face recognition via progressive pose estimation, in: IEEE International Conference on Image Processing, 2014, pp.728–732.
[10] K. Panchal, H. Shah, 3D face recognition based on pose correction using Euler angle method, in: International Conference on Machine Intelligence and Research Advancement, 2013, pp.467–471.
[11] A. Asthana, T.K. Marks, M.J. Jones, K.H. Tieu, M.V. Rohith, 3d pose normalization, in: IEEE International Conference on Computer Vision, 2011, pp.937–944.
[12] A. Cheraghian, K. Faez, H. Dastmalchi, F.B. Oskuie, An efficient multimodal face recognition method robust to pose variation, Comput. Inform. (2011) 431–435.
[13] Frav3d, http://www.frav.es/databases/FRAV3D/.
[14] S.A. Mahmood, R.F. Ghani, A.A. Kerim, 3D face recognition using pose invariant nose region detector, in: Computer Science and Electronic Engineering Conference, 2014, pp.103–108.
[15] H. Drira, B.B. Amor, A. Srivastava, M. Daoudi, R. Slama, 3D face recognition under expressions, occlusions, and pose variations, IEEE Trans. Pattern Anal. Mach. Intell. 35 (9) (2013) 2270–2283.
[16] G. Passalis, P. Perakis, T. Theoharis, I.A. Kakadiaris, Using facial symmetry to handle pose variations in real-world 3D face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 33 (10) (2011) 1938–1951.
[17] S. Berretti, A.D. Bimbo, P. Pala, Sparse matching of salient facial curves for recognition of 3-D faces with missing parts, IEEE Trans. Inf. Forensics Secur. 8 (2) (2013) 374–389.
[18] U. of Notre Dame, University of Notre Dame Biometrics Database, http://www.nd.edu/~cvrl/UNDBiometricsDatabase.html, 2008.
[19] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: IEEE International Conference on Comput. Vis. Pattern Recognition, 2005, pp.947–954.
[20] A.B.A. Moreno, A. Sanchez, GavabDB: A 3D face database, in: Workshop Biometrics on the Internet, 2004, pp.77–85.
[21] S. Berretti, N. Werghi, A.D. Bimbo, P. Pala, Matching 3D face scans using interest points and local histogram descriptors, Comput. Graph. 37 (5) (2013) 509–525.
[22] H. Li, D. Huang, J.M. Morvan, Y. Wang, L. Chen, Towards 3D face recognition in the real: A registration-free approach using fine-grained matching of 3D keypoint descriptors, Int. J. Comput. Vis. 113 (2) (2015) 128–142.
[23] N. Werghi, C. Tortorici, S. Berretti, A.D. Bimbo, Boosting 3D LBP-based face recognition by fusing shape and texture descriptors on the mesh, IEEE Trans. Inf. Forensics Secur. 11 (5) (2016) 1–16.
[24] A. Savran, N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gokberk, B. Sankur, L. Akarun, Bosphorus database for 3D face analysis, in: The First COST 2101 Workshop on Biometrics and Identity Management, 2008, pp.47–56.
[25] Y. Liang, Y. Zhang, Expression-invariant face recognition using three-dimensional weighted walkthrough and centroid distance, J. Electron. Imaging 24 (5) (2015).
[26] S. Berretti, A.D. Bimbo, E. Vicario, Weighted walkthroughs between extended entities for retrieval by spatial arrangement, IEEE Trans. Multimedia 5 (1) (2003) 52–70.
[27] S. Berretti, A.D. Bimbo, P. Pala, 3D face recognition using iso-geodesic stripes, IEEE Trans. Pattern Anal. Mach. Intell. 32 (12) (2010) 2162–2177.
[28] T. Khadhraoui, F. Benzarti, H. Amiri, A novel approach to nose-tip and eye-corners detection using HK-classification in case of 3D face analysis, Comput. Appl. Inform. Syst. (2014) 1–4.
[29] P. Perakis, G. Passalis, T. Theoharis, I.A. Kakadiaris, 3D facial landmark detection under large yaw and expression variations, IEEE Trans. Pattern Anal. Mach. Intell. 35 (7) (2013) 1552–1564.
[30] P.J. Besl, R.C. Jain, Invariant surface characteristics for 3D object recognition in range images, Comput. Vis. Graph. Image Process. 33 (1) (1986) 33–80.
[31] M. Emambakhsh, A.N. Evans, Self-dependent 3D face rotational alignment using the nose region, in: International Conference on Imaging for Crime Detection and Prevention, 2011, pp.1–6.
[32] P. Nair, A. Cavallaro, 3-D Face detection, landmark localization, and registration using a point distribution model, IEEE Trans. Multimedia 11 (4) (2009) 611–623.
[33] W. Lin, H. Chu, J. Wu, B. Sheng, Z. Chen, A heat-map-based algorithm for recognizing group activities in videos, IEEE Trans. Circuits Syst. Video Technol. 23 (11) (2015) 1980–1992.
[34] P.J. Besl, N.D. McKay, A method for registration of 3-D shapes, IEEE Trans. Pattern Anal. Mach. Intell. 14 (2) (1992) 239–256.
[35] Y. Liu, J. Palmer, A quantified study of facial asymmetry in 3D faces, in: IEEE International Workshop on Analysis and Modeling of Faces and Gestures, 2003, pp.222–229.