Iterative Closest Normal Point for 3D Face Recognition

Hoda Mohammadzade, *Student Member, IEEE*, and Dimitrios Hatzinakos, *Senior Member, IEEE*

Abstract—The common approach for 3D face recognition is to register a probe face to each of the gallery faces, and then calculate the sum of the distances between their points. This approach is computationally expensive and sensitive to facial expression variation. In this paper, we introduce the iterative closest normal point method for finding the corresponding points between a generic reference face and every input face. The proposed correspondence finding method samples a set of points for each face, denoted as the closest normal points. These points are effectively aligned across all faces enabling effective application of discriminant analysis methods for 3D face recognition. As a result, the expression variation problem is addressed by minimizing the within-class variability of the face samples while maximizing the between-class variability. As an important conclusion, we also show that the surface normal vectors of the face at the sampled points contain more discriminatory information than the coordinates of the points. We have performed comprehensive experiments on the Face Recognition Grand Challenge database which is presently the largest available 3D face database. We have achieved verification rates of 99.6% and 99.2% at a false acceptance rate of 0.1% for the all vs. all and ROC III experiments, respectively, which, to the best of our knowledge, have seven and four times less error rates, respectively, compared to the best existing methods on this database.

Index Terms—Three-dimensional, face recognition, expression variation, point correspondence, 3D registration, normal vector, LDA.

1 Introduction

Because of its non-intrusive and friendly nature and also its applicability for surveillance, face recognition has drawn enormous attention from the biometrics community. Despite the significant advances over the past two decades, 2D face recognition still bears limitations due to pose, illumination, expression and age variation between the probe and gallery images. Due to the advances in 3D imaging sensors, researchers are now paying more attention to 3D face recognition to overcome the pose and illumination issues existing in 2D face recognition. Moreover, the geometric information available in 3D data provides more discriminatory information for face recognition.

While some of the 3D face recognition methods take advantage of using 2D face images in combination with 3D data, the focus of this paper is on using only the 3D data for recognition.

Expression variation is also one of the main challenges in 3D face recognition because the geometry of a face changes drastically under expression variation. Existing approaches to the expression variation problem can be categorized into rigid and non-rigid methods. The rigid methods [1], [2], [3], [4], [5], [6], [7] use regions of the face that are almost invariant to expression, such as the nose region. These regions are matched with their corresponding ones on the galley faces using the iterative closest point (ICP) method [8], and the resulting mean square error (MSE) of the alignment is used for recognition. In the non-rigid methods [9], [10], [11], [12], a deformable model is used to deform an expression face to its neutral face. The deformable model is learned using training subjects displaying both neutral and expression faces. The deformed faces are matched with neutral faces using the ICP method.

Registration of a probe face to every gallery face makes the existing approaches computationally very expensive. Moreover, the existing rigid methods do not take into account the fact that the expression-invariant regions of the face are subject-specific, that is, depending on the subject, a specific region of the face may or may not change under a particular expression. Similarly, the existing non-rigid methods do not take into account the fact that the deformation of the face under expression is also subject-specific, that is, the deformation of the face is different across population.

Our solution to 3D face recognition is to use discriminant analysis (DA) methods [13], [14], [15], [16], [17], [18], [19], [20]. An important advantage of these methods over the existing methods is that, by learning from training faces displaying various expressions for different subjects, these methods are able to automatically find *subject-specific* expression-invariant regions of the face [21]. Such methods, in the high dimensional face space, find projection directions that are invariant to expression changes *for each subject*. As an evidence for this claim, we have shown in [21] for 2D face recognition that, when the training images belonging to a specific subject are removed from the training set of a DA method,

• *The authors are with The Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, ON M5S 3G4, Canada.*
Email: {hoda,dimitris}@comm.utoronto.ca

This work has been supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

the recognition rate deteriorates for that subject. This means that, expression-invariant projection directions obtained for some subjects might not be suitable for some other subjects, and therefore, are subject-specific, to some extent.

Another important advantage of the proposed method is that one-to-one alignment between a probe face and all of the gallery faces is not required for recognition, which enables fast database searches.

However, the use of a DA method requires a proper alignment of its inputs. This is also a requirement for the use of principal component analysis (PCA). Early attempts to apply PCA for 3D face recognitions involved applying PCA to the range images of the 3D face data [1], [22], [23], [24]. In these methods, the range images are either aligned in 2D space by detecting the facial features, e.g., the eyes, or the 3D faces are first aligned to a generic reference face using ICP and then their range images are used for applying PCA. The problem with the former alignment approach is that the size information of the face, which is useful for recognition, is dismissed. The problem with the later alignment approach is inaccurate alignment of 3D faces using ICP, especially under expression variation.

More recent approaches to apply PCA for 3D face recognition involve finding point-to-point correspondence between 3D faces and a generic reference face [25], [26], [27]. PCA is then applied to the coordinates of the surface points of the face. These points should be corresponded across all faces in such a way that if, for instance, the ith point of one face corresponds to the left corner of the left eye, then the ith point of all faces corresponds to the same feature. In [25], the point-to-point correspondence is achieved by finding the closest points of a 3D face to the points of a *scaled* generic reference face. The closest points are found using the ICP approach. However, because ICP does not result in accurate correspondence between the points of two faces, scaling the reference face is an important required step of this approach. For scaling the reference face for each input face, five feature points of the input face are used. However, because of the lack of an accurate facial feature detector, in [25], the feature points for each face were manually detected. In applying PCA using the morphable models [27], the point-to-point correspondence is established using an optical flow technique. The cost function for the optical flow algorithm includes the texture value and the 3D coordinates of the face points. For the initialization of this algorithm, seven feature points of the face are required to be determined, which was performed manually [27].

The recognition performance of the approaches that use PCA for 3D face recognition is not satisfactory especially under expression variation. This is mainly because PCA is an unsupervised approach and does not take into account the within-class variability of the subjects. On the other hand, by finding the projection directions that maximizes the between-class variability of the faces while minimizing their within-class variability, DA methods are able to address the expression variation problem. In order to use DA methods for 3D face recognition, we establish correspondence across different faces by means of a generic reference face. For each point on the reference face, one corresponding point on every face is selected using our proposed correspondence finding method denoted as the iterative closest-normal point (ICNP) method. This method uses two criteria to find the corresponding points: the Euclidian distance between the two faces and the angle between their *surface normal vectors*. (For the rest of the paper, we call the surface normal vectors simply as the normal vectors.) In particular, we will show that the surface orientation factor works well under expression variation. Compared to the other point-to-point correspondence finding approaches, our method does not require determining any facial feature point and does not use the texture values of the face points.

Our ICNP method is distinguished from the registration method of Chen and Medioni [28] in which the distance between the points on one surface and the tangent planes on other surface is minimized. This minimization, which replaces minimizing the point-to-point distances done in the ICP method [8], results in a faster global optimization. Moreover, in the method of Chen and Medioni, no corresponding point is obtained. On the other hand, in our proposed ICNP method, the angle between the normal vectors on the two surfaces is minimized followed by low-pass filtering the distance vectors between the points of the two surfaces, which results in a smooth *point-to-point* correspondence between the two surfaces.

In our proposed 3D face recognition approach, first the points corresponding to the points of a generic reference face are found for each face. These corresponding points are denoted as the closest normal points (CNPs). Then, a DA method is applied to the *normal vectors* at the CNPs of each face for recognition. We will show that the normal vectors contain more discriminatory information than the coordinates of the points of a face.

The main contributions of this work can be summarized as follows.

- The ICNP method is proposed to establish effective correspondence across the points of the faces.
- Only a sampled set of points, i.e., the CNPs, are used for an effective application of DA methods.
- The normal vectors of the face at the sampled points are input to a DA method for recognition.

We have used the Face Recognition Grand Challenge (FRGC) v2.0 [29] database to evaluate the performance of our proposed face recognition method. The FRGC is an international benchmarking database containing 4007 3D faces from 466 subjects displaying various facial expressions. These images have been taken during Fall 2003 and Spring 2004. We performed various experiments on this database to show the effectiveness of the proposed method.

We present a complete 3D face recognition system including the preprocessing and recognition stages. Comprehensive experiments are also performed to support the proposed approach. This paper is organized as follows. Section 2 reviews the related work. Section 3 describes the preprocessing steps including the detection of the nose tip and the correction the face pose. Section 4 presents the proposed method for finding the CNPs and Section 5 describes how the iterations of the proposed ICNP method work. Section 6 presents the face recognition results and Section 7 concludes the paper.

2 Related Work

For a comprehensive review of 3D face recognition methods one may refer to [30], [31]. We review here the state-of-the-art 3D face recognition approaches with high recognition rates on the FRGC database. It is common in the related literature to perform either of the following two experiments on the FRGC database to evaluate the performance of a 3D face recognition method. One is the ROC III experiment which was suggested by the FRGC program [29]. In this experiment, the images taken in Fall 2003 form the gallery set and the images taken in Spring 2004 form the probe set. The other experiment is the all-versus-all experiment in which every two images in the dataset (4007 images) are matched to each other.

Mian et al. [2] proposed a multimodal 2D+3D face recognition system. For 3D face images, they proposed the spherical face representation (SFR) to eliminate a large number of candidates for efficient recognition at a later stage. After this elimination, the remaining faces are automatically segmented into two regions: the eyes and forehead region and the nose region. To match two faces, these regions are matched with their corresponding ones using a modified ICP method. The final matching score is obtained by combining minimum MSE from each region. With only using the 3D scans, they achieved a verification rate (VR) of 86.6% at 0.1% FAR on the FRGC database for the all vs. all experiment.

Maurer et al. [3] proposed a method to combine 2D and 3D face images for recognition. To match a 3D probe face to a 3D gallery face, they first used the ICP method to align the two faces. Then a distance map is computed using the distance of the two faces on every pixel. The statistics of the distance map is then used to generate a matching score. Using only the 3D scans, they achieved a VR of 87.0% at 0.1% FAR on the FRGC database for the all vs. all experiment.

Husken et al. [4] proposed to fuse 2D and 3D hierarchical graph matching (HGM) for face recognition. An elastic graph is automatically adapted to the depth image using the landmarks found on the texture image. With only using the 3D scans, they achieved a VR of 86.9% at 0.1% FAR on the FRGC database for the ROC III experiment.

Lin et al. [5] proposed fusion of summation invariant features extracted from different regions of a face. They proposed an optimal fusion of the matching scores between corresponding regions for recognition. They achieved a VR of 90.0% at 0.1% FAR on the FRGC database for the ROC III experiment.

Cook et al. [6] proposed Log-Gabor templates for robust face recognition. They achieved relative robustness to occlusions, distortions and facial expressions by breaking a face into 75 semi-independent observations in both the spatial and frequency domains. The matching scores from each observation are combined using linear support vector machines (SVM). They achieved a VR of 92.3% at 0.1% FAR on the FRGC database for the all vs. all experiment.

Kakadiaris et al. [9] proposed an annotated deformable model for the expression variation problem in 3D face recognition. The 3D model is first fitted to an aligned 3D face and then their difference is measured using a combination of three matching procedures including spin images, ICP and simulated annealing on z-buffers. The 3D geometric differences are then mapped onto a 2D regular grid and two wavelet transformations including, Pyramid and Haar, are used for recognition. They achieved a VR of 97.0% at 0.1% FAR on the FRGC database for the ROC III experiment.

Ocegueda et al. [32] proposed an extension to the method of Kakadiaris et al. [9] by reducing the number of wavelet coefficients used for recognition. They reduced this number by adding a feature selection step and projecting the signatures to LDA projection bases. They achieved a VR of 96.8% at 0.1% FAR on the FRGC database for the ROC III experiment.

Al-Osamini et al. [10] proposed an expression deformation model by learning from pair of neutral-nonneutral faces for each subject. A PCA subspace is built using the shape residues of the pairs of the scans which are aligned using the ICP algorithm. The learnt patterns of the expression deformation are then used to morph out the expression from a non-neutral face for recognition. They achieved a VR of 94.1% at 0.1% FAR for the ROC III experiment on the FRGC database.

Faltemier et al. [33] proposed fusion of results from 28 spherical regions of the face. They performed score-based fusion on the matching scores of the individual regions for recognition. They showed that the Borda Count and Consensus Voting result in the best performance compared to other combination approaches. They achieved a VR of 94.8% at 0.1% FAR on the FRGC database for the ROC III experiment and a VR of 93.2% for the all vs. all experiment.

McKeon [34] proposed an extension to the method of Faltemier et al. [33] by using fusion and score normalization techniques. He achieved a rank-one recognition rate 98.6% for an experiment on the FRGC database, where the first scan of each subject forms the gallery set and the subsequent scans form the probe set. No performance rate is reported for the standard ROC III or all vs. all experiment. The experiment performed in [34] was also performed in [33]. Comparing the results

of this experiment with the results of the ROC III and all vs. all experiment in [33] shows that, the VRs at 0.1% FAR for the ROC III and all vs. all experiment should be lower than 98.6% using the method of McKeon [34].

Spreeuwers [35] proposed an approach for registration to an intrinsic coordinate system of the face, which is defined by the vertical symmetry plane through the nose, the tip of the nose and the slope of the bridge of the nose. He also proposed a classifier based on the fusion of 120 dependent region classifiers for overlapping face regions. He achieved a VR of 94.6% at 0.1% FAR on the FRGC database for the all vs. all experiment.

Huang et al. [36] proposed a method for face representation based on multi-scale extended local binary patterns for describing the local shape variations on range images. They achieved VRs of 97.2% and 98.4% at 0.1% FAR on the FRGC database for the neutral vs. non-neutral and neutral vs. all experiments, respectively. No performance rate is reported for the standard ROC III or all vs. all experiment. The experiments performed in [36] were also performed in [2]. Comparing the results of these experiments with the results of the ROC III and all vs. all experiments in [2] shows that, the VRs at 0.1% FAR for the ROC III and all vs. all experiment should be lower than 97.2% using the method of Huang et al. [36].

Queirolo et al. [7] proposed a simulated annealing-based approach for registration of 3D faces and used surface interpenetration measure for recognition. They segmented the face into four regions: the entire face, the forehead, the circular and the elliptical regions around the nose. They modified the simulated annealing approach for the entire face to address the expression variation problem. They achieved a VR of 96.6% at 0.1% FAR on the FRGC database for the ROC III experiment and a VR of 96.5% for the all vs. all experiment.

In summary, the method of Queirolo et al. [7] produced the best results on the FRGC database for the all vs. all experiment with a VR of 96.5% at 0.1% FAR and the method of Kakdiaris et al. [9] produced the best results on the FRGC database for the ROC III experiment with a VR of 97.0% at 0.1% FAR.

3 Preprocessing

The FRGC v2.0 [29] database contains 4007 3D faces of 466 subjects. The data are acquired from the frontal view and from the shoulder level up. The 3D data are stored in the form of four matrices of size 480×640. The first matrix is a binary mask indicating the valid points, and the remaining three matrices contain the x, y, and z coordinates of the points. Only the points corresponding to the subject are valid points, i.e., the points corresponding to the background are invalid points. The origin of the coordinate system is the 3D camera, and the z-axis corresponds to the depth of the image.

We found two persons with inconsistent subject ID's in the FRGC v2.0 database by visually inspecting their color images. The first one is the person with the ID numbers 4643 and 4783, and the second one is the person with the ID numbers 4637 and 4638. The first case has also been mentioned in [7]. The first person has a total of six and the second person has a total of five 3D face scans in the entire database. Because of the large number of subjects (i.e., 466) and also the very large number of 3D faces (i.e., 4007) in this database, the effect of correcting these labels on the VRs for the ROC III and all vs. all experiments is negligible. Therefore, we can validly compare the results of our method with those of the other methods, regardless of whether they have done this correction. However, this correction is useful for investigating the individual matching scores.

3.1 Nose Detection

The first step in preprocessing is to localize the face in the 3D image. Because the nose tip usually has the smallest depth, it can be easily detected and used to localize the face. However, in some images, because of the out-of-plane rotation of the head or the presence of the hair in the face, the point with the smallest depth does not correspond to the nose tip. To verify whether a point is actually a nose tip, we use the PCA technique described in [37]. In this method, first, a PCA space is constructed using a set of training nose regions. Then, if the distance between a region and its projection onto this space is smaller than a threshold, it is verified as a nose region. This technique was originally proposed for detecting facial features in 2D face images. We adapt this method here for 3D faces as follows.

Unlike other parts of the proposed 3D face recognition method, in which 3D points are directly used in the procedures, the proposed nose detection algorithm uses the range image representation of 3D faces. For training, we visually verify the nose tip in a number of range images (40 subjects here). We then crop the 3D points inside a sphere of radius 60 mm centered at the visually verified nose tips. After that, the z value of the cropped points are resampled on a uniform rectangular grid in the xy-plane at 1 mm resolution with the nose tip shifted to the (0,0) coordinate. Also, the z value of the nose tip point is subtracted from the z value of all of the points (i.e., the nose tip is at (0,0,0) coordinate.). We use linear interpolation for resampling. The range of the rectangular grid is [-30,30] in the x direction and [-20,50] in the y direction. An example of the resulting range image of a nose region is shown in Fig. 1 (a). The resulting range images form a set of training nose region images.

We then construct a PCA space by applying PCA [38] to the training nose region images. Because this space represents the space of the nose regions, we call it *eigen-nose space*. In order to apply PCA, we form a vector representation for each training nose region image by column-wise concatenation of its depth values. The eigen-nose space is represented by the mean nose, which is the mean of the training nose vectors, and the set of

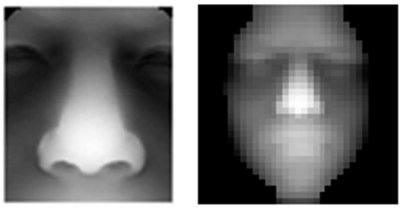


Fig. 1. (a) Example of the range image of a training nose region and (b) example of the range image of a low-resolution training face region.

the leading eigenvectors of the covariance matrix of the training nose vectors.

To verify whether a point is a nose tip, we first crop and resample the region surrounding the point as described above. We then project the vector representation of the candidate region into the eigen-nose space. The projection of the candidate vector $\boldsymbol{x}$ into the eigen-nose space is obtained as

$$\mathcal{P}(\boldsymbol{x}) = [\boldsymbol{u}_1 ... \boldsymbol{u}_k][\boldsymbol{u}_1 ... \boldsymbol{u}_k]^t(\boldsymbol{x} - \bar{\boldsymbol{n}}) + \bar{\boldsymbol{n}} \qquad (1)$$

where $\boldsymbol{u}_{1...k}$ are the normalized eigenvectors corresponding to the k largest eigenvalues of the covariance matrix and $\bar{\boldsymbol{n}}$ is the mean nose. If the MSE between a candidate vector and its projection (i.e., the reconstruction error) is smaller than a threshold (we use 100 as the threshold here), we verify it as a nose region. Otherwise, the point with the next smallest depth is selected and tested. This procedure can be repeated until the nose tip is detected.

There are two issues associated with this nose verification system. One issue is that sometimes a very large number of points (as large as 10,000) should be tested for a face until the nose tip is found, which is time-consuming. This happens when a part of the face is covered with the hair or when the head has excessive out-of-plane rotation. The other issue is that sometimes, because of the excessive rotation of the head, the reconstruction error for the nose region falls above the specified threshold. We address these issues by using a low-resolution and wider-nose-region PCA space as follows.

We first create a low-resolution *eigen-face* space using the training 3D faces as follows. We crop a sphere of radius 100 mm centered at the visually verified nose tips and then resample the z values on a square grid in the xy-plane at 5 mm resolution. The range of the grid is [-80,80] centered at the nose tip in both directions. An example of the training low-resolution face regions is shown in Fig. 1 (b). We then construct the eigen-face space in a similar way to the eigen-nose space.

If the verification using the eigen-nose space fails for a certain number of points (20 points here), we reduce the resolution of the 3D points of the input face, i.e., the x, y and z matrices, by five times, using the value of the nearest neighbor entries. Then, we start with the point with the smallest depth. We crop and resample the region around the candidate point at 5 mm resolution and project the range image vector to the low-resolution eigen-face space. If the reconstruction error falls below a lower threshold (150 here), the point is verified as the nose tip. Otherwise, if the reconstruction error falls below an upper threshold (320 here), then the pose of the candidate face region is corrected using the pose correction procedure described in Section 3.3 and the reconstruction error is calculated again. If now the reconstruction error falls below the lower threshold, it is verified as the nose tip, otherwise the point with the next smallest depth is selected and tested. This procedure is repeated until either the nose tip is detected or the number of points tested reaches a threshold (600 here). If finally, no point is verified as the nose tip, the point with the smallest reconstruction error is selected as the nose tip from among the 600 tested points. The nose detection procedure is summarized by the block diagram shown in Fig. 2.

The thresholds for the reconstruction error are determined by performing a 10-fold cross-validation test on the training faces.

The nose tip in 2885, out of 4007, faces in the FRGC database was verified using the first point. Also, the nose tip in 3251 faces was verified without using the low-resolution part. Clearly, this number can be much higher if we increase the thresholds for the high-resolution part. However, choosing conservative thresholds guarantees an error-free and efficient detection. Distribution of the number of the points tested for each face until the nose tip is found is summarized in Table 1.

TABLE 1
Frequency of the number of the points (m) tested for each face until the nose tip is found.

	$m = 1$	$2 \leq m \leq 20$	$21 \leq m \leq 100$	$101 \leq m \leq 600$
No. of faces	2885	366	655	81

Examples of the reconstruction error map for faces that required testing more than 100 points before the nose tip can be detected are shown in Fig. 3. In both faces, points of the hair and/or forehead have a smaller depth than the nose tip.

The proposed nose detection algorithm is computationally efficient and simple, and successfully detect the nose tip in *all* of the 4007 3D scans in the FRGC database (i.e., 100% success rate). Segmenting the face in the 3D scans has been a major issue in the 3D face recognition. While different approaches have been proposed so far [2], [33], [7], [39], none of them achieved a 100% success rate. Errors of these methods are usually due to the presence of the hair in the face, excessive rotation of the head, or smearing of the face scan by the subject's motion. The face in all of these cases have been successfully segmented by our method. Fig. 4 shows some examples

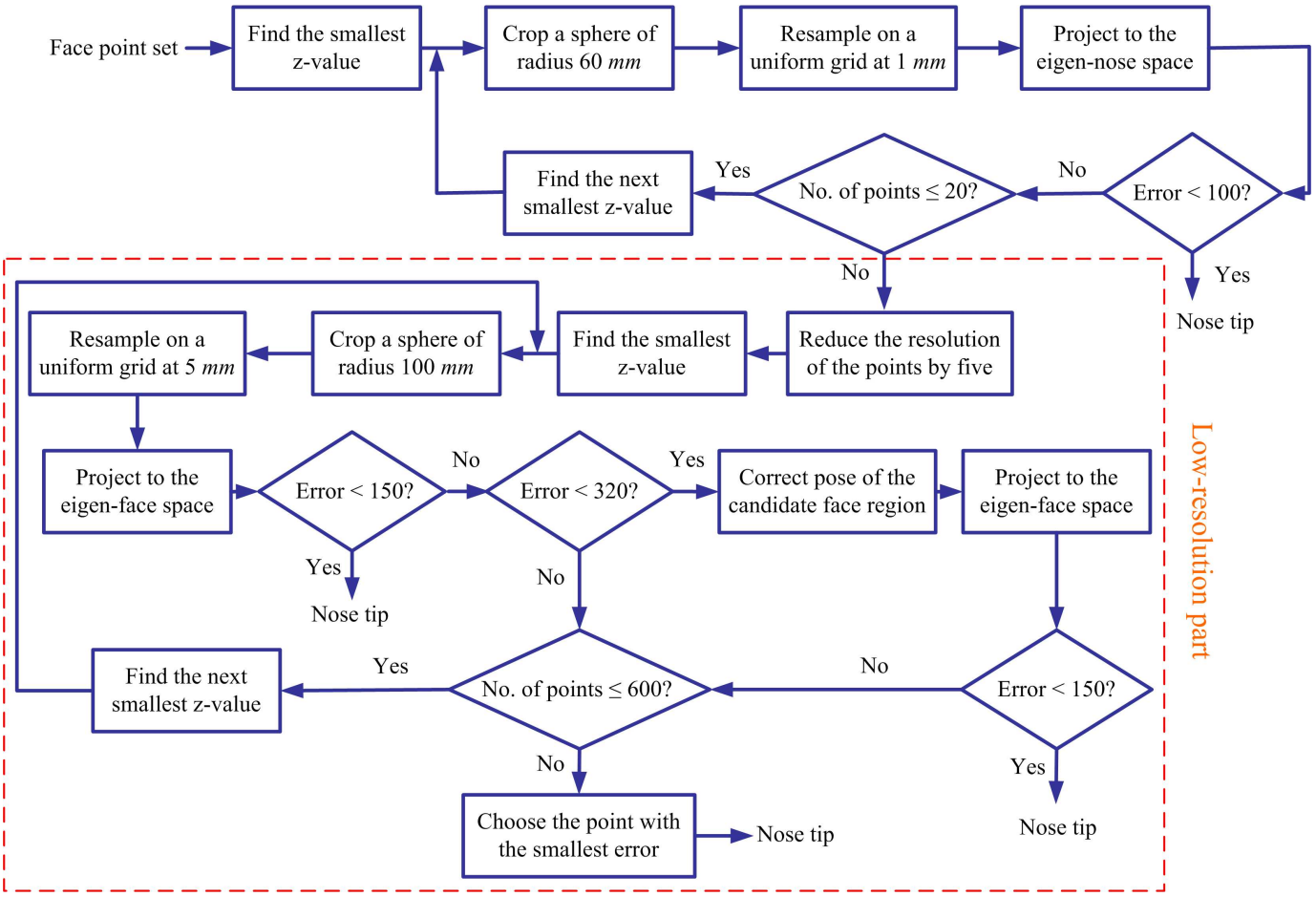


Fig. 2. Block diagram of the proposed procedure for nose detection.

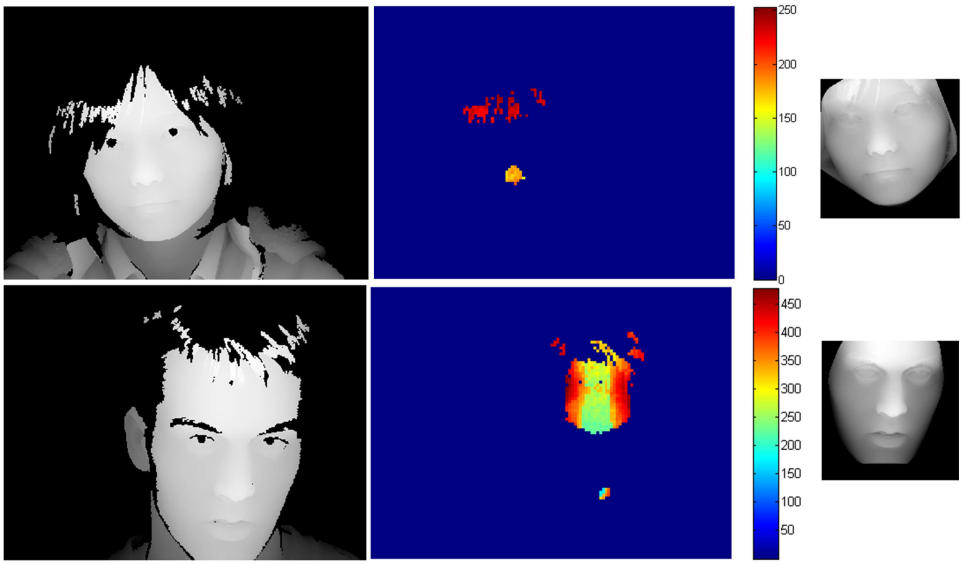


Fig. 3. (a) The range image of the raw 3D scan, (b) the reconstruction error map for the points tested, and (c) the range image of the cropped and resampled face regions.

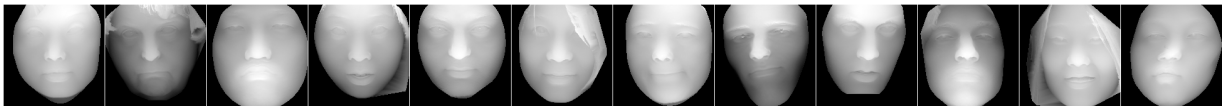


Fig. 4. Examples of the faces with hair artifact, excessive head rotation and smearing due to subject's motion. All such faces were successfully segmented by the proposed nose verification algorithm.

of these faces.

3.2 Denoising

After detecting the nose tip, a sphere of radius 100 mm centered at the nose tip is cropped as the face region. The cropped region is then resampled on a uniform square grid in the xy-plane at 1 mm resolution, where the nose tip is shifted to the (0,0) coordinate in this grid. The range of the grid is [-80,80] centered at the nose tip in both directions. Also, the z value of the nose tip is subtracted from the z value of all of the points.

Possible spikes in the 3D data are removed during the nose detection phase by removing the tested points for which there are only a few points within the cropped sphere. Also, the holes are filled by linear interpolation during the resampling phase.

After cropping the face region, the z component is denoised using 2D Wiener filtering as follows. 2D Wiener filtering is used to low-pass-filter 2D images that are degraded by additive noise [40]. In 2D Wiener filtering, for each pixel of the image, adaptive Wiener method is used based on statistics estimated from a local neighborhood of the pixel. In applying a 2D Wiener filter for denoising the z component of the 3D face, we use a window of size 4×4 mm in the xy-plane. Let μ and σ^2 denote the mean and variance of the z component of the points located within the window centered at the current point, i.e.,

$$\mu = \frac{1}{n} \sum_{i \in \eta} z_i \tag{2}$$

and

$$\sigma^2 = \frac{1}{n} \sum_{i \in \eta} z_i^2 - \mu^2 \tag{3}$$

where η denotes the set of the indices of the points located within the window centered at the current point and n is the number of points within the window. The filtered z component of the current point, denoted as $f(z)$, is then obtained as

$$f(z) = \mu + (1 - \frac{\nu^2}{\sigma^2})(z - \mu) \tag{4}$$

where ν^2 is the variance of the noise. We estimate the variance of the noise by averaging the local variance of the z component over the entire face.

We also experimented with other low-pass-filtering approaches, such as median and averaging filtering, and eventually concluded that the 2D Wiener filtering results in better definition around the facial features (i.e., eyes, nose and mouth).

3.3 Pose Correction

To correct the pose of the face, we adapt the PCA technique described in [2], which is also known as the Hotelling transform, as follows. Let $\mathbf{P}$ be a $3 \times N$ matrix of the x, y, and z coordinates of the point set of a face,

$$\mathbf{P} = \begin{bmatrix} x_1 & x_2 & \dots & x_N \\ y_1 & y_2 & \dots & y_N \\ z_1 & z_2 & \dots & z_N \end{bmatrix} \tag{5}$$

The covariance matrix $\mathbf{C}$ of the points is obtained as

$$\mathbf{C} = \frac{1}{N} \mathbf{P}\mathbf{P}^t - \bar{\boldsymbol{p}}\bar{\boldsymbol{p}}^t \tag{6}$$

where $\bar{\boldsymbol{p}} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{p}_i$ is the mean of the points and $\boldsymbol{p}_i = [x_i \; y_i \; z_i]^t$. Calculating the eigenvectors of the covariance matrix $\mathbf{C}$ gives us the rotation matrix $[\boldsymbol{v}_1 \boldsymbol{v}_2 \boldsymbol{v}_3]$, where $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, and $\boldsymbol{v}_3$ are the normalized eigenvectors of $\mathbf{C}$, which are sorted in the descending order of their eigenvalues. It is important to note that the i-th component of the eigenvector $\boldsymbol{v}_i$ should be positive, i.e., if this component is negative, the eigenvector should be multiplied by -1 in order to prevent excessive rotation of the face. Finally, the point set $\mathbf{P}$ is rotated and aligned to its principal axes as

$$\mathbf{P}' = [\boldsymbol{v}_1 \; \boldsymbol{v}_2 \; \boldsymbol{v}_3](\mathbf{P} - \bar{\boldsymbol{p}}) \tag{7}$$

After rotating the face to its principal axes, it is resampled again around the nose tip on a uniform square grid at 1 mm resolution. The eigenvectors are calculated using the resampled points. Rotating and resampling are repeated until the rotation matrix converges to an identity matrix, i.e., no more significant rotation is applied to the point set. To test the convergence, we calculate the square root of the sum of the squared differences between the current rotation matrix and the identity matrix, and we use the threshold of 0.002. To avoid the effect of the hair, for calculating the eigenvectors, we use the points inside a sphere of radius 70 mm. The pose correction procedure is summarized by the block diagram shown in Fig. 5.

Each 3D face image in the FRGC dataset has a corresponding 2D color image (texture map) of size 480×640

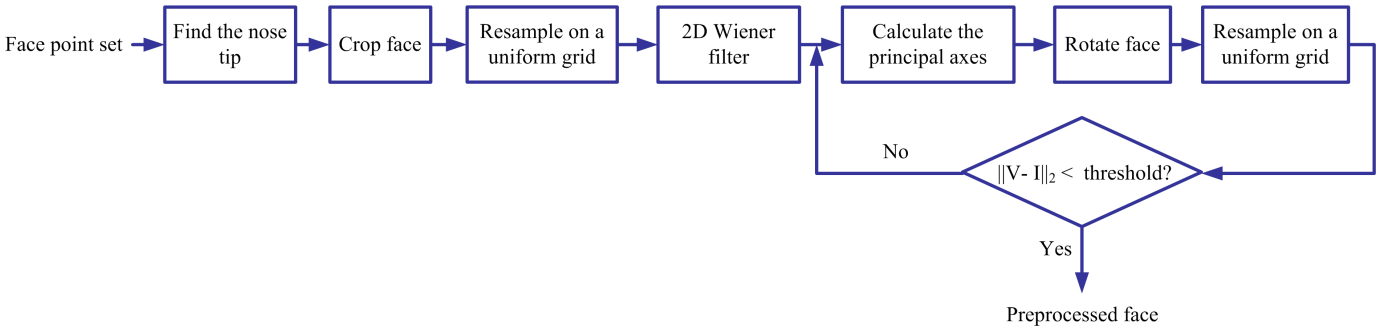


Fig. 5. Block diagram of the pose correction procedure.

pixels. Texture images usually have a pixel-to-pixel correspondence to their respective range images. Exceptions occur when the subject moves or rapidly changes expression. Although we are not using texture images for face recognition, they are very useful for the demonstration of our method. In order to preprocess texture images, after localizing the face in the 3D image, the corresponding region in the texture image is cropped and then the pose of the 2D face image is corrected using the same rigid transformation that was applied to its corresponding 3D face, as described in [2]. The 2D image is then resampled on the same grid as the 3D face. The resulting 2D face images usually have a pixel-to-pixel correspondence to their respective preprocessed range images. Fig. 6 shows some examples of preprocessed 2D and 3D images. The top two rows contain initial 3D and 2D images in which the face region was cropped around the detected nose tip. The bottom two rows contain the corresponding pose-corrected images. As it is seen, the pose correction also acts as a rough alignment of 3D faces. The pose correction is a computationally simple step and greatly speeds up the the correspondence finding step. It should be noted that, sometimes because of the presence of the hair in the face, pose correction might fail. An example of this failure is shown in the fourth column from the right in Fig. 6. However, this problem will be resolved by the rigid transformation in the correspondence finding step, as described in Section 5.

4 Correspondence Finding

4.1 Reference Face

We create the reference face by smoothing a random pose-corrected denoised face with a neutral expression. For smoothing, we use a 2D Wiener filter with the filtering window of size 6×6 mm. The reference face that was used for the experiments in this paper is shown in Fig. 7(a). This reference face was created from the 3D scan with ID 04460d260. The choice of the face was random and we expect that other faces work equally well.

4.2 Closest Normal Search

Our goal at this stage is to find for each point on the reference face one corresponding point on the input face such that it resembles the features of its reference point. For example, for the points of the eyes of the reference face, the corresponding points are the points of the eyes of the input face. Note that the corresponding points therefore represent the same facial features *across all faces*. Also note that the number of the corresponding points for every input face is equal to the number of the points of the reference face. Let $\{\boldsymbol{r}_i\}_{i=1}^{N}$ represent the point set of the reference face, and $\{\boldsymbol{p}_i\}_{i=1}^{M}$ represent the point set of the input face, where $\boldsymbol{r}_i$ and $\boldsymbol{p}_i$ indicate 3D vectors containing the coordinates of the points. Also, let $\{\boldsymbol{\rho}_i\}_{i=1}^{M}$ and $\{\boldsymbol{\pi}_i\}_{i=1}^{N}$ represent the sets of unit normal vectors of the reference face and the input face, respectively. The corresponding point for each point $\boldsymbol{r}_i$ of the reference face is obtained through the following steps.

First, the point with the smallest distance to the reference point is selected. Let $\{\boldsymbol{p}_i^c\}_{i=1}^{N}$ denote the set of the *closest points*. $\{\boldsymbol{p}_i^c\}_{i=1}^{N}$ is a subset of $\{\boldsymbol{p}_i\}_{i=1}^{M}$ such that,

$$\boldsymbol{p}_i^c = \boldsymbol{p}_k, k = \arg \min_{1 \le j \le M} (\|\boldsymbol{r}_i - \boldsymbol{p}_j\|_2). \quad (8)$$

For the fast search of closest points, a KD-tree algorithm can be used.

Fig. 7(b) shows one cross section of the reference face (blue points) and the corresponding cross section on the input face (red points). The position of the cross sections on the faces is shown by dotted lines in Fig. 7(a) and (e). The horizontal axis in Fig. 7(b) represents the z-axis (depth) of the 3D face points and the vertical axis represents the y-axis. For the purpose of demonstration, in Fig. 7(b),(c), and (d) the reference points have been shifted by 10 mm to the left. In Fig. 7(b), each vector connects a reference point to its closets point on the input face.

In Fig. 7, in order to demonstrate the closest points, we have assumed that they are located on the same cross section of the input face. However, it should be noted that, the actual closest points can be located anywhere within a neighborhood of this cross section.

As mentioned in Section 3.3, preprocessed color images have a pixel-to-pixel correspondence to their respec-

Fig. 6. Examples of the preprocessed 2D and 3D images. The top two rows are face regions cropped around the nose tip. The bottom two rows are corresponding pose-corrected images. 3D images are shown in shaded view.

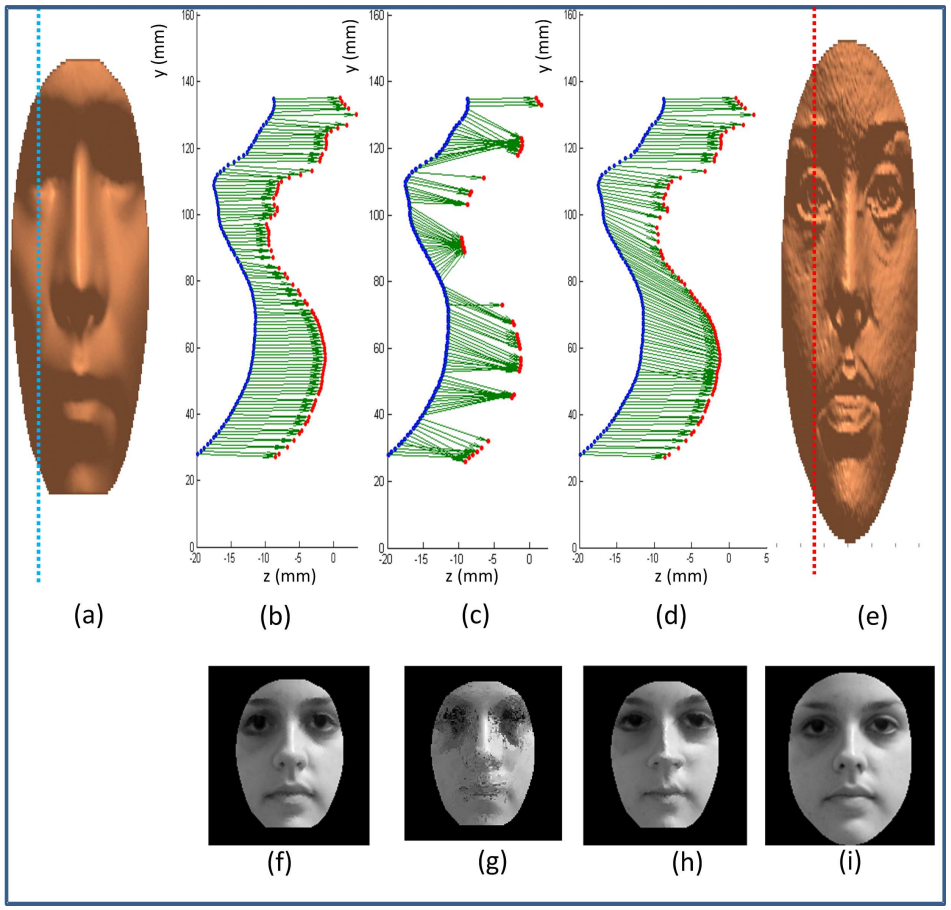


Fig. 7. Different steps of the proposed correspondence finding method on a cross section of a sample face and the corresponding texture-to-reference maps at each step. (a) Reference face, (b) closest points, (c) closest normal points, (d) filtered distance vectors, (e) input face, (f),(g),(h) texture-to-reference map of the points in (b),(c), and (d), respectively, and (i) the 2D image of the input face.

tive preprocessed range images. That is, each 3D point has a texture value. Now, in order to check whether successful correspondences have been established between the reference face and an input face, we have created a specific texture map for the corresponding points on the input face by mapping the texture of these points to their reference points in the xy-plane. We obtained such mapping by replacing the pixel values in the range image of the reference face by the texture value of their corresponding points from the input face. When the corresponding points establish correct correspondences between the reference face and the input face, this texture map should represent the texture of the *input face* mapped to the geometry of the *reference face*. For example, if the mouth is open in the input face but closed in the reference face, this texture map of the input face should display a closed-mouth face. Similarly, the other features of the input face should have been shifted to their corresponding locations within the reference face. We call this texture map the *texture-to-reference map*.

The texture-to-reference map of the closest points is shown in Fig. 7(f). This texture map is similar to the 2D face image of the input face (shown in Fig. 7(i)) and does not represent the geometry of the reference face. This observation demonstrates that the closest points cannot be considered as the corresponding points.

After finding the closest point, we search a window of size $10 \times 10mm$ centered at the closest point to find the point that its normal vector has the smallest angle with the normal vector at the reference point. Let $\{\boldsymbol{p}_i^{cn}\}_{i=1}^N$ represent the set of the points with the *closest normals* to the reference points. $\{\boldsymbol{p}_i^{cn}\}_{i=1}^N$ is again a subset of $\{\boldsymbol{p}_i\}_{i=1}^M$, but here such that,

$$\boldsymbol{p}_i^{cn} = \boldsymbol{p}_k, k = \arg\min_{j \in \eta}(\arccos(\boldsymbol{\rho}_i.\boldsymbol{\pi}_j)), \tag{9}$$

and η denotes the set of the indices of the input face points which are located within a window of size 10×10 mm centered at the point $\boldsymbol{p}_i^c$. Fig. 7(c) shows the points with the closest normals to the reference points. These points still cannot be considered as corresponding points because they are not smoothly corresponded to the reference points. The texture-to-reference map of these points is shown in Fig. 7(g), where the effect of unsmooth correspondences is evident.

4.3 Smoothing the Correspondences

In order to smooth the correspondences, a 2D Wiener filter is applied to the distance vectors between the closest normal points and the reference points, as follows. Let $\{\boldsymbol{d}_i\}_{i=1}^N$ represents the set of these distance vectors which are obtained by subtracting the coordinates of the reference points from the coordinates of their closest normal points,

$$\boldsymbol{d}_i = \boldsymbol{p}_i^{cn} - \boldsymbol{r}_i. \tag{10}$$

In fact, the vectors in Fig. 7(c) represent these distance vectors. In order to apply the 2D filter to the distance vectors, a vector field is constructed by assigning each distance vector to the x and y coordinates of its reference point. Fig. 8(a) shows a sample of this vector field which is projected to xy-plane. The filtering is then performed on each of the three components of the distance vector field. A window size of 20×20 is used for the filtering. Fig. 8(b) shows the filtered version of the vector field.

A consequence of the filtering with the large window size is that the vectors at the boundary of the face cannot be filtered. To solve this problem, the boundary distance vectors are interpolated using their nearest neighbors.

Let $\{f(\boldsymbol{d}_i)\}_{i=1}^N$ represent the set of filtered distance vectors. Then, the smoothed closest normal points can be approximated by adding the smoothed distance vectors to the reference points,

$$\boldsymbol{p}_i^{cn'} = \boldsymbol{r}_i + f(\boldsymbol{d}_i), \tag{11}$$

where $\{\boldsymbol{p}_i^{cn'}\}_{i=1}^N$ denotes the set of *approximate* smoothed closest normal points. These points are approximate because they may not lie exactly on the input face surface. To resolve this issue, the point on the input face with the smallest Euclidian distance to the approximate smoothed closest normal point is selected as the smoothed closest normal point. That is,

$$\boldsymbol{p}_i^{cn''} = \boldsymbol{p}_k, k = \arg\min_{1 \le j \le M}(\|\boldsymbol{p}_i^{cn'} - \boldsymbol{p}_j\|_2). \tag{12}$$

where $\{\boldsymbol{p}_i^{cn''}\}_{i=1}^N$ denotes the set of smoothed closest normal points. Again, for the fast search of closest points, the KD-tree algorithm can be used. Fig. 7(d) shows the smoothed distance vectors and closest normal points. As it can be seen, a fine correspondence between the eyes of the reference face and the eyes of the input face has been achieved. The texture-to-reference map of the corresponding points is shown in Fig. 7(h). Evidently, this texture map represents the texture of the input face mapped to the geometry of the reference face (for example look at the position of the eyes), showing that fine correspondences between the points of the reference face and the points of the input face has been achieved.

5 The ICNP Method

Throughout the rest of this paper, we refer to the smoothed closest normal points simply as the closest normal points (CNPs). After finding the CNPs, the input face is rotated and translated to reduce the distance between the CNPs and the reference points. The resulting better alignment between the input face and the reference face results in a more accurate correspondence between their points. We repeat the search for the CNPs, rotation and translation of the input face until no more significant rotation is applied to the input face.

We calculate the rotation matrix and the translation vector using the singular value decomposition (SVD) [41] method as follows. Let $\bar{\boldsymbol{r}} = \frac{1}{N}\sum_{i=1}^N \boldsymbol{r}_i$ and $\bar{\boldsymbol{p}} = \frac{1}{N}\sum_{i=1}^N \boldsymbol{p}_i^{cn''}$ be the mean of the reference points and the CNPs, respectively. The cross correlation matrix $\mathbf{C}$

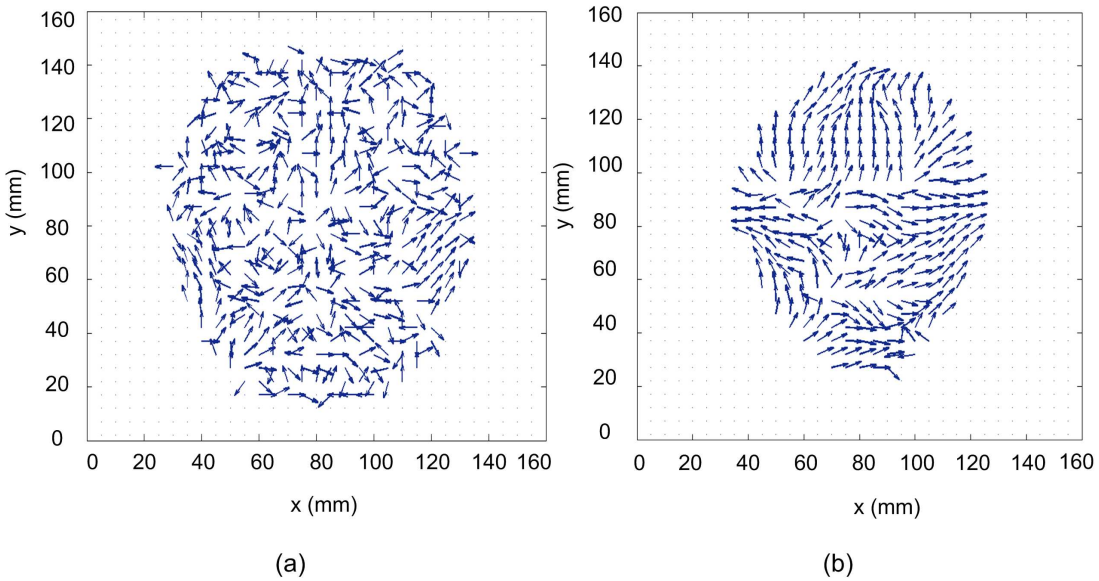


Fig. 8. The distance vector field before and after smoothing. (a) The initial distance vector field projected to xy-plane, (b) the smoothed distance vector field projected to xy-plane.

between the reference points and the CNPs is obtained as

$$\mathbf{C} = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{p}_i^{cn''} - \bar{\boldsymbol{p}})(\boldsymbol{r}_i - \bar{\boldsymbol{r}})^t. \tag{13}$$

Let $\mathbf{U}$ and $\mathbf{V}$ be the orthogonal and $\mathbf{A}$ be the diagonal matrices obtained from the SVD of $\mathbf{C}$ as

$$\mathbf{U}\mathbf{A}\mathbf{V}^t = \mathbf{C} \tag{14}$$

The rotation matrix $\mathbf{R}$ is then obtained as

$$\mathbf{R} = \mathbf{V}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \det(\mathbf{U}\mathbf{V}^t) \end{bmatrix}\mathbf{U}^t \tag{15}$$

where the middle matrix is to prevent reflection of the face when $\det(\mathbf{U}\mathbf{V}^t) = -1$. The translation vector $\boldsymbol{t}$ is obtained as

$$\boldsymbol{t} = \bar{\boldsymbol{p}} - \mathbf{R}\bar{\boldsymbol{r}} \tag{16}$$

The transformation of the CNPs is then obtained as

$$\mathrm{T}(\boldsymbol{p}_i^{cn''}) = \mathbf{R}\boldsymbol{p}_i^{cn''} + \boldsymbol{t} \tag{17}$$

This rigid transformation minimizes the mean square distance between the reference points and the CNPs.

The CNP search, rotation and translation of the reference face are repeated until the sum of squared differences between the current rotation matrix and the identity matrix is less than a threshold. After the last iteration of the ICNP algorithm, the final CNPs are obtained. The ICNP algorithm is summarized by the block diagram shown in Fig. 9.

In each iteration of the ICNP algorithm, we have also applied the same rigid transformation to the corresponding color image of the input face (as described in Section 3.3), in order to maintain the pixel-to-pixel correspondence between the 3D face and its color image.

The CNPs can be considered as a sampling of the points of a face surface according to the points of the reference face. An example of the CNPs is shown in Fig. 10 (b), where the surface has been created by interpolating the CNPs. The CNPs by themselves do not demonstrate the quality of the correspondence between the reference and input faces. On the other hand, the texture-to-reference maps seem to be helpful for this purpose as explained above.

Fig. 10. (a) The shaded view of a sample face and (b) the shaded view of the face created using the CNPs.

When the geometry of a face changes under expression, the proposed method is able to successfully find the same corresponding points that it usually finds from its neutral face. When a DA method is applied to the CNPs, this stability is a key factor in the success of the DA method to recognize expression variant faces. That is, by establishing stable correspondences, a DA method can successfully learn the geometry changes resulting from expression variation of a subject versus the geometry changes resulting from subject variations. DA methods learn these differences by minimizing the within-class

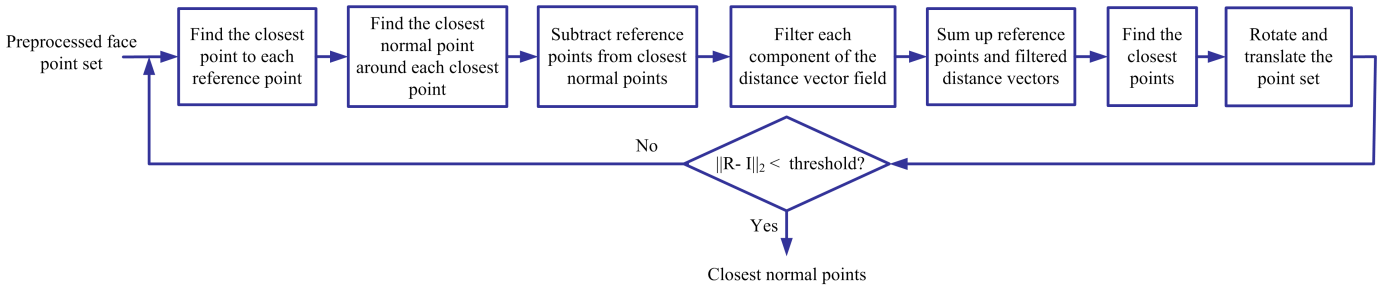


Fig. 9. Block diagram of the proposed ICNP method.

scatter matrix while maximizing the between-class scatter matrix [13].

Fig. 11 shows examples of the texture-to-reference maps of the CNPs for a number of non-neutral faces. As it can be seen, the texture maps represent the same facial geometry across all the faces. Also, it is seen that, the CNPs are almost invariant under expression, i.e., almost the same point on the face is found as the CNP, as if the face was neutral. In particular, note that, in the texture-to-reference maps, the mouth is always closed. That is, the CNPs to the points of the lips of the reference face are always the corresponding points of the lips of the input face regardless of whether the mouth is open or closed in the input face. Dealing with open mouth has been a serious issue in 3D face recognition and a number of works have been proposed to address this issue [42]. We expect that the success of the proposed method in correctly establishing correspondence between an open mouth and a reference closed mouth can greatly improve 3D face recognition.

6 Face Recognition Results

To evaluate the performance of the proposed method, we reproduce the ROC III and all vs. all experiments here (see Section 2 for the definition of these experiments). We also use LDA as an example of DA methods for the proposed face recognition system. Many state-of-the-art DA methods have been recently proposed. However the focus of this paper is to propose a proper base for the employment of DA methods in 3D face recognition and not to extensively examine different DA methods. Even though we only use a simple DA method here, the performance improvement is significant.

In order to employ LDA, database images should to be divided into gallery and probe sets. LDA is then trained using the gallery images. Gallery and probe sets are already defined in the ROC III experiment. However, there is no consensus for these sets for the all vs. all experiment. We observed that the performance rate of LDA varies by changing the number of 3D face scans that are used in the gallery set for each subject. However, the number of scans for each subject in the FRGC database varies between two to 22. Table 2 shows the number of subjects with *at least* n scans in the database. Therefore, to evaluate the performance of the proposed face recognition method, we used a maximum for the number of scans that are used for each subject in the gallery. Also, to investigate the performance change under different numbers of sample available for each subject in the gallery, we evaluated the performance for different such maximum numbers.

We therefore used the following setup to reproduce the all vs. all experiment. We randomly selected a maximum of m scans from each subject to form the gallery set and we used the remaining images to form the probe set. For the subjects with fewer than or equal to m scans, we randomly selected one image from each subject for the probe set and the remaining images were used for the gallery set. We then calculated the matching score for every pair of gallery-probe scans. We repeated the random division of the dataset into the gallery and probe sets for many times to make sure that every two images in the dataset are matched with each other, which is the consensus for the test setting for the all vs. all experiment. Finally, we calculated the VRs at different thresholds using the similarity scores over all the trials.

6.1 Normal Vectors versus Point Coordinates

Here we compare the face recognition performance using the normal vectors versus using the point coordinates. After finding the CNPs for each face in the gallery and probe sets, first we performed face recognition using the coordinates of these points as follows. We first constructed a $3 \times N$ matrix from the coordinates of the CNPs for each face, where one row corresponds to each of the x, y and z coordinates. We then concatenated the x, y and z rows of the point matrix to form a 1D vector for each face. We then trained LDA using these vectors of the gallery faces. Because of the high dimensionality of these vectors, it is infeasible to directly apply the LDA algorithm. As proposed in [13], to solve this problem, we first applied PCA [38] to these vectors to reduce their dimension. The number of the PCA components that we used was according to 99% of the eigenvalue energy.

After reducing the dimension, we obtained the LDA projection bases through generalized eigenvalue decomposition as described in [13]. We discarded the eigenvectors corresponding to the eigenvalues which are smaller than 0.01. We then obtained the feature vectors of the

Fig. 11. The first and the third rows show the 2D image of sample 3D faces, and the second and the fourth rows show the texture-to-reference maps of their CNPs.

TABLE 2
The number of the subjects with at least n scans in the FRGC database.

n	1	2	3	4	5	6	7	8	10	15	22
Number of subjects	464	409	382	353	318	286	257	230	186	95	3

gallery and probe faces by projecting their reduced-dimension vectors to the LDA bases. Finally, we used the cosine metric for measuring the similarity of each pair of gallery and probe faces.

A second time, we used the normal vectors of the faces at the CNPs instead of the point coordinates. That is, we constructed a $3 \times N$ matrix from the components of the normal vectors at the CNPs for each face.

Fig. 12 shows the verification results for the above two experiments using the ROC III scenario. Clearly, the use of normal vectors significantly outperforms the use of point coordinates. This observation is further verified in Section 6.6. Therefore, we conclude that, the normal vectors of the face contains more discriminatory information than the coordinates of the face points. One possible explanation for this observation is that, the normal vector at a point contains information regarding the neighboring points as well. This conclusion is one of the main contributions of this work.

Fig. 13 shows the x, y and z components of the normal vectors at the CNPs for a number of sample faces. As shown, each component represents the orientation of the face surface in the corresponding direction. These components can be considered as *three orthogonal channels* of the geometry of a face.

For the rest of the experiments in this paper, we use the normal vectors at the CNPs for recognition unless otherwise is stated.

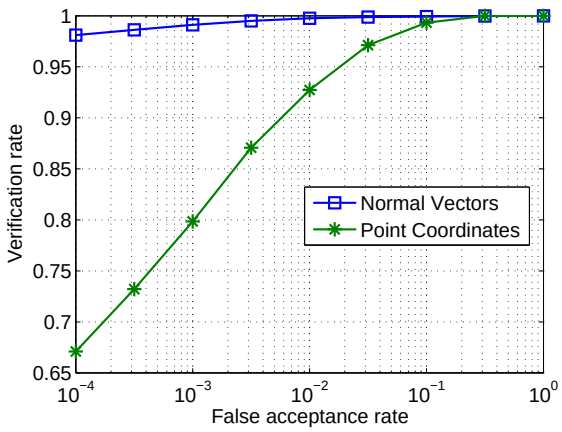


Fig. 12. VR versus FAR for the ROC III experiment using the CNPs versus using the normal vectors at the CNPs .

6.2 Face Recognition Using Different Numbers of Scans for Each Subject

Table 3 shows the VR at 0.1% FAR for the all vs. all experiment when the maximum of m scans per subject are used in the gallery set in each trial. As it can be

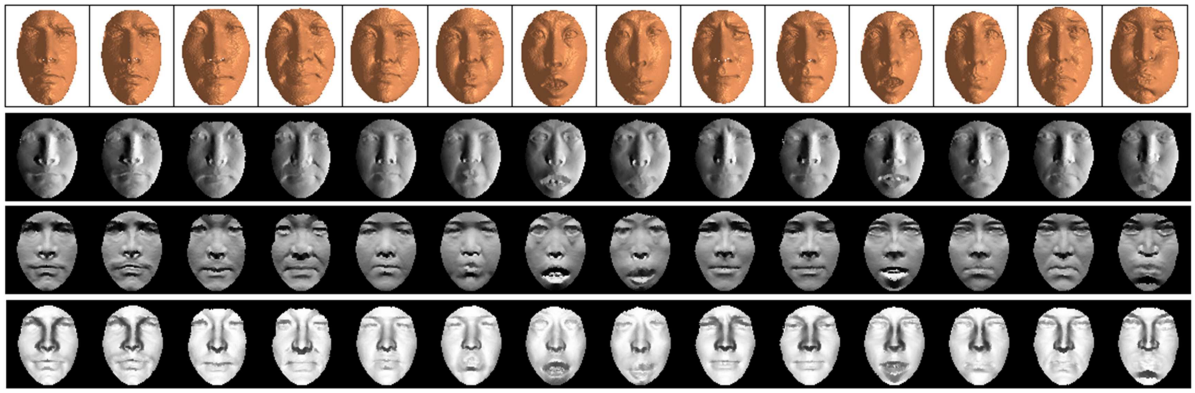


Fig. 13. Examples of each component of the normal vectors at the CNPs. From top to bottom: shaded view of the 3D face, x, y, and z component of the normal vectors at the CNPs.

seen, by increasing the number of scans per subject, LDA learns within-subject geometric variations across different expressions and better recognition is achieved.

TABLE 3
VR at 0.1% FAR for different maximum numbers of scans per subject in the gallery.

m	2	3	4	5	6	7
VR	90.6	98.4	99.2	99.5	99.6	99.6

Because capturing a 3D face scan takes only a few seconds, in most face recognition applications, the enrolees can easily provide multiple 3D scan samples of themselves. However, for situations with limited number of samples, we can use a generic training set to boost the performance. A generic training set consists of samples of subjects who are different from the testing subjects. Clearly, there is no limitation on the number of samples that can be collected for this set. By training LDA using the gallery samples and a generic set, we can improve the recognition performance. Moreover, we know that, at least two samples per subject are required for training LDA. However, as proposed in [43], by using a generic set, LDA can also be used in the so-called *single sample* scenario. Table 4 shows the VR at 0.1% FAR when a generic set with the maximum of six scans per subject is included for training. In this table, the VRs are listed for different sizes of the generic set. We randomly selected the specified number of subjects from the FRGC database for the generic set, and used the remaining subjects for testing and then formed the gallery and probe sets as described in the beginning of Section 6. Again, the random division of the database into the generic and gallery-probe sets was repeated for many times to be confident that every two images in the dataset are matched with each other. As it can be seen in Table 4, by using a generic training set, the recognition performance using only two scans per subject has been significantly improved. Moreover, a very good performance has been achieved by using only one scan per subject.

TABLE 4
VR at 0.1% FAR for cases with one and two samples per subject in gallery.

Max. no. of scans per subject in gallery	1	2
VR without using a generic set	NA	90.6
VR using a generic set of 50 subjects	82.3	95.1
VR using a generic set of 100 subjects	86.3	96.9
VR using a generic set of 300 subjects	93.1	98.3

6.3 The Effect of Expression

By visually inspecting each image in the FRGC database, we classified 2455 faces as neutral and 1552 faces as non-neutral. Non-neutral faces display various expressions including surprise, happy, puffy cheeks, anger, etc. There are also a large number of faces with closed eyes among them. Most of the non-neutral faces display an intense expression. Fig. 14 shows examples of the non-neutral faces from the first four subjects in this database. More examples of the non-neutral faces can be seen in Figs. 6, 11, and 13.

Fig. 14. Examples of non-neutral faces from the first four subjects in the FRGC database.

By including one neutral image for each subject to the set of non-neutral faces, we created a subset of the

FRGC database in which there is expression variation between every two faces of a subject. The resulting set, denoted as the expression subset, consists of 1976 faces. We also denote the set of the 2455 neutral faces as the neutral subset. Table 5 shows the VRs at 0.1% FAR for the expression and neutral subset when different maximum numbers of scans are used in the gallery. We performed an experiment similar to the all vs. all experiment on these subsets. No generic training set has been used for the results in this table. As it is shown, by increasing the number of scans per subject, the recognition performance for the expression subset approaches that for the neutral subset, which indicates an expression-invariant recognition. By increasing the number of scans per subject, LDA learns within-subject geometric variations across *different expressions*.

TABLE 5
VR at 0.1% FAR for the neutral and expression subset when different maximum numbers of scans per subject are used in the gallery.

m	2	3	4	5	6	7
Neutral	98.8	99.7	99.9	99.9	99.9	99.9
Expression	77.5	90.8	96.2	97.6	98.5	98.5

6.4 Corresponding Points versus Range Images

Here we compare face recognition performance using CNPs versus using the range images (depth map) of 3D faces. In order to use the range images, we first aligned all faces to the reference face using the ICNP method. We then used the range images of the aligned faces for applying LDA [13]. Fig. 15 shows the verification results for the ROC III experiment using the range images versus using the normal vectors at the CNPs. The verification results for the all vs. all scenario are similar but are not shown here for brevity. Clearly the use of the CNPs significantly outperforms the use of range images, showing the importance of sampling of the face points guided by a reference face.

6.5 LDA versus PCA

In order to show the role of LDA in the proposed approach, here we compare the face recognition performance using LDA versus using PCA. As mentioned before, once successful correspondences are established across faces, LDA is capable of distinguishing between geometry changes resulted from expression variation and those resulted from subject variation; whereas PCA is an unsupervised method and does not take into account the within-subject variability of the faces. Fig. 16 shows the verification results for the ROC III experiment using these two feature extraction methods. The same settings have been used for training these two methods.

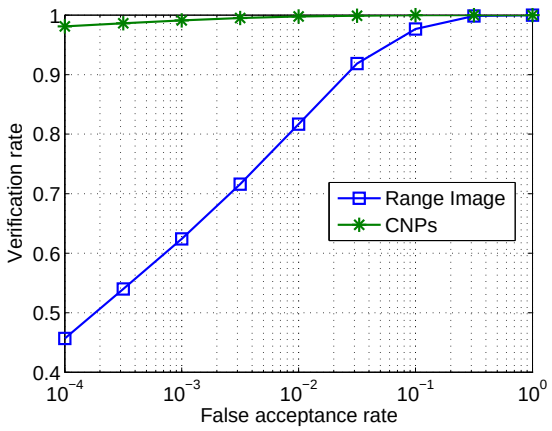


Fig. 15. VR versus FAR for the ROC III experiment using the normal vectors at the CNPs versus using the range images.

The verification results for the all vs. all scenario are again similar but are not shown here for brevity. As it is seen, LDA significantly outperforms PCA.

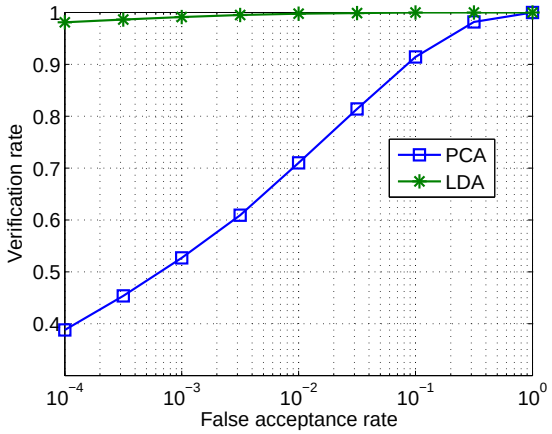


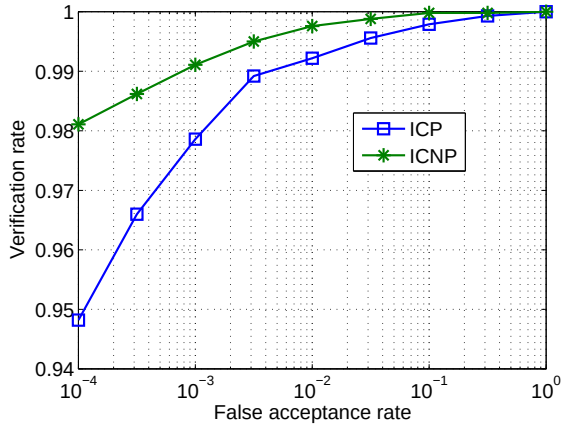
Fig. 16. VR versus FAR for the ROC III experiment using LDA versus using PCA.

6.6 ICNP versus ICP

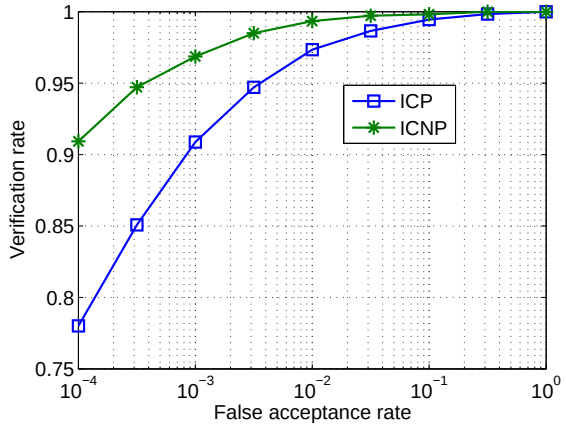
Finding the correct corresponding points across faces is a key factor for the success of LDA in the proposed face recognition system. Here we compare the face recognition performance using the ICNP method versus using the ICP method. In order to evaluate the face recognition performance using the ICP method, we experimented with coordinates of the corresponding points as well as with the normal vectors at the corresponding points. We then observed that when the ICP method is used, the use of the point coordinates results in better recognition performance; whereas as it was shown earlier, when the ICNP method is used, the use of the normal vectors results in better performance. As a result, in order to

have a fair comparison between the ICP and ICNP methods, we use the normal vectors in the case that the ICNP method is used and use the point coordinates in the case that the ICP method is used. Fig. 17 (a) shows the verification results for these two methods for the ROC III scenario. Clearly, the ICNP method significantly outperforms the ICP method. The verification results for the all vs. all scenario are again similar but are not shown here for brevity.

We also performed another experiment to compare the ICP and ICNP methods using the expression subset defined in Section 6.3. Fig. 17 (b) shows the verification results for this experiment. As it is depicted, the difference between the performance of the ICNP and ICP method is greater when expression variation is larger.



(a)



(b)

Fig. 17. VR versus FAR using the normal vectors through the ICNP method versus using the point coordinates through the ICP method (a) for the ROC III experiment and (b) on the expression subset.

6.7 Fusion of Normal Vectors and Point Coordinates

We performed another experiment to examine whether combining the normal vectors and the point coordinates can improve the recognition performance. As concluded in Section 6.6, in order to use point coordinates for recognition, we applied the ICP method and in order to use the normal vectors, we applied the ICNP method. We then used the simple sum rule for fusing these two modalities. That is, the final matching score for a pair of gallery-probe faces is the summation of the cosine from the feature vectors of the two modalities. The verification results for the ROC III and all vs. all experiments are shown in Fig. 18 (a) and (b), respectively. As it is seen, some improvement has been achieved by fusing these two modalities. We performed a similar experiment on the expression subset. The verification results for this experiment are shown in Fig. 18 (c), depicting noticeable improvement on images with more expression variations.

By using a simple fusion rule, we showed that combining normal vectors and point coordinates can improve the recognition performance even further. One may consider examining state-of-the-art combination strategies such as stack generalization [44], rule based schemes [45], and order statistics [46] to seek more improvement.

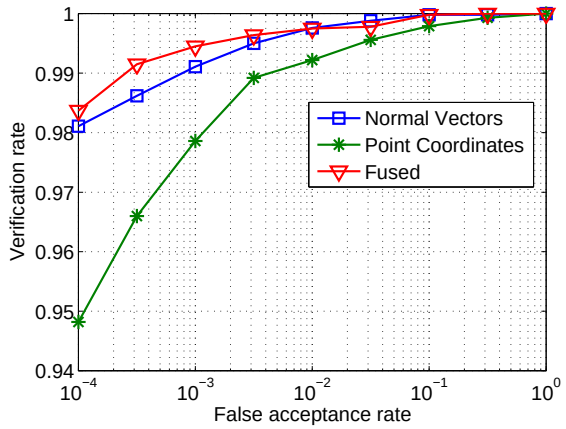
6.8 Performance Comparison with Other Methods

Here we compare the performance of the proposed 3D face recognition system with that of the state-of-the-art methods. Table 6 shows the verification results for state-of-the-art methods on the FRGC database as reported in the literature. Some methods have only reported the verification results for the all vs. all experiment while some other have only reported them for the ROC III experiment.

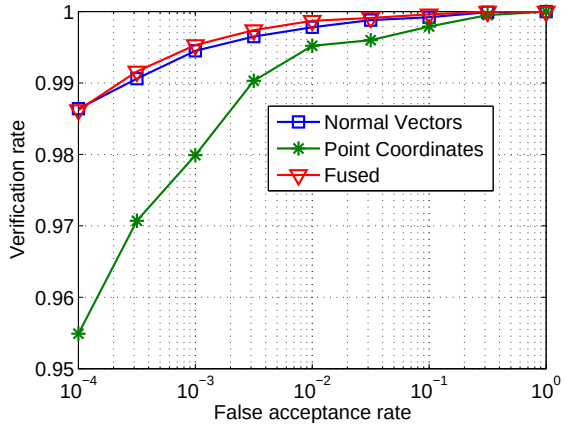
Also the VRs using the proposed method for both of these experiments are shown in Tabel 6. For these results, only the normal vectors at the CNPs have been used (i.e., without fusing with the point coordinates). For the all vs. all experiment, a maximum of six scans per subject have been used for the gallery set in each trial. A VR of 99.6% at 0.1% FAR has been achieved using the proposed method for the all vs. all experiment. Compared to the method of Queirolo et al. [7], which produces the best results for the all vs. all experiment among the existing methods, our method has reduced the error rate by 7 times. Also, a VR of 99.2% at 0.1% FAR has been achieved using the proposed method for the ROC III experiment. Compared to the method of Kakadiaris et al. [9], which produces the best results for the ROC III experiment among the existing methods, our method has reduced the error rate by 4 times.

7 Conclusions

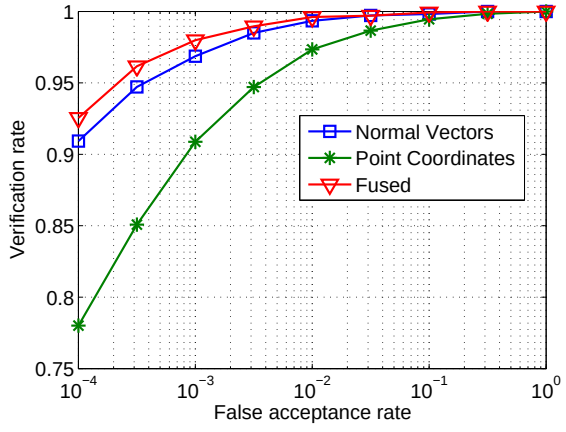
We introduced the ICNP method for effectively finding the points that correspond to the same facial features across all faces. We used a surface orientation factor to find the corresponding points denoted as CNPs. We showed that the CNPs that are detected from an expression face are the same as those detected from the



(a)



(b)



(c)

Fig. 18. VR versus FAR using the normal vectors, the point coordinates and the fusion of the two (a) for the ROC III experiment, (b) for the all vs. all experiment, and (c) on the expression subset.

TABLE 6
Verification results for the all vs. all and ROC III experiments, at 0.1% FAR.

Method	ROC III	All vs. All
Mian et al. [2]	NA	86.6
Husken et al. [4]	86.9	NA
Maurer et al. [3]	NA	87.0
Lin et al. [5]	90.0	NA
Cook et al. [6]	NA	92.3
Al.Osaimi et al. [10]	94.1	NA
Faltemier et al. [33]	94.8	93.2
Spreeuwers et al. [35]	NA	94.6
Queirolo et al. [7]	96.6	96.5
Ocegueda et al. [32]	96.8	NA
Kakadiaris et al. [9]	97.0	NA
This work	99.2	99.6

corresponding neutral face. In particular, the CNPs to the points of the lips of the reference face are the corresponding points of the lips of the input face regardless of whether the mouth is open or closed in the input face.

We also showed that a successful application of DA methods for 3D face recognition can be achieved by using the CNPs. As an straightforward DA method, we used LDA for recognition and achieved significant improvements in recognition. We expect that the use of more advanced DA methods provides more improvement.

As an important conclusion, we observed that the normal vectors at the CNPs provide a higher level of discriminatory information than the coordinates of the points, i.e., the normal vectors of the face are more useful for recognition.

Another important feature of the proposed approach is that one-to-one alignment/registration of a probe face to every gallery face is not required for recognition, which enables fast database searches.

REFERENCES

[1] K.I. Chang, K.W. Bowyer, , and P.J. Flynn, "Multiple nose region matching for 3D face recognition under varying facial expression," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1695–1700, 2006.

[2] A. Mian, M. Bennamoun, and R. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1927–1943, 2007.

[3] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, and G. Medioni, "Performance of geometrix activeid 3D face recognition engine on the FRGC data," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, p. 154, 2005.

[4] M. Husken, M. Brauckmann, S. Gehlen, and C.V. der Malsburg, "Strategies and benefits of fusion of 2D and 3D face recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, p. 174, 2005.

[5] W.-Y. Lin, K.-C. Wong, N. Boston, and Y.H. Hu, "3d face recognition under expression variations using similarity metrics fusion," *Proc. IEEE Int. Conf. Multimedia and Expo*, pp. 727–730, 2007.

[6] J. Cook, C. McCool, V. Chandran, and S. Sridharan, "Combined 2D/3D face recognition using log-Gabor templates," *Proc. IEEE Int. Conf. Video and Signal Based Surveillance*, p. 83, 2006.

[7] C.C. Queirolo, L. Silva, O.R.P. Bellon, and M.P. Segundo, "3d face recognition using simulated annealing and the surface interpenetration measure," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, pp. 206–219, 2010.

[8] P.J. Besl and N.D. McKay, "A method for registeration of 3-D shapes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[9] I. Kakadiaris, G. Passalis, G. Toderici, N. Murtuza, and T. Theoharis, "Three-dimensional face recognition in the presence of facial expression: An annotated deformable model approach," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 640–649, 2007.

[10] F. Al Osaimi, M. Bennamoun, and A. Mian, "An expression deformation approach to non-rigid 3D face recognition," *Int. J. Computer Vision*, vol. 81, pp. 302–316, 2009.

[11] J. Huang, B. Heisele, and V. Blanz, "Component-based face recognition with 3D morphable models," *Proc. Int. Conf. Audio and Video-Based Biometric Person Authentication*, pp. 27–34, 2003.

[12] X. Lu and A.K. Jain, "Deformation modeling for robust 3D face matching," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 1377–1383, 2006.

[13] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[14] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, vol. 12, pp. 2385–2404, 2000.

[15] J. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms," *IEEE Trans. Neural Networks*, vol. 14, no. 1, pp. 117–126, 2003.

[16] T. Hastie, A. Buja, and R. Tibshirani, "Penalized discriminant analysis," *Annals of Statistics*, vol. 23, pp. 73–102, 1995.

[17] M. Loog and R.P.W. Duin, "Linear dimensionality reduction via a heteroscedastic extension of LDA: The chernoff criterion," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 732–739, 2004.

[18] M. Zhu and A.M. Martinez, "Subclass discriminant analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 8, pp. 1274–1286, 2006.

[19] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data-with applications to face recognition," *Pattern Recognition*, vol. 34, pp. 2067–2070, 2001.

[20] J. Wang, K.N. Plataniotis, J. Lu, and A.N. Venetsanopoulosc, "Kernel quadratic discriminant analysis for small sample size problem," *Pattern Recognition*, vol. 41, pp. 1528–1538, 2008.

[21] H. Mohammadzade and D. Hatzinakos, "An expression transformation for improving the recognition of expression-variant faces from one sample image per person," *Proc. IEEE Int. Conf. Biometrics: Theory Applications and Systems*, pp. 1–6, 2010.

[22] T. Heseltine, N. Pears, and J. Austin, "Three-dimensional face recognition: an eigensurface approach," *Proc. Int. Conf. Image Processing*, pp. 1421–1424, 2004.

[23] C. Hesher, A. Srivastava, and G. Erlebacher, "A novel technique for face recognition using range imaging," *Proc. Int. Symp. Signal Processing and Its Applications*, vol. 2, pp. 201–204, 2003.

[24] X. Yuan, J. Lu, and T. Yahagi, "A method of 3D face recognition based on principal component analysis algorithm," *Proc. IEEE Symp. Circuits and Systems*, vol. 4, pp. 3211–3214, 2005.

[25] T. Russ, C. Boehnen, and T. Peters, "3D face recognition using 3D alignment for PCA," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2006.

[26] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 9, pp. 1063–1074, 2003.

[27] V. Blanz, K. Scherbaum, and H.P. Seidel, "Fitting a morphable model to 3D scans of faces," *Proc. Int. Conf. Computer Vision*, pp. 1–8, 2007.

[28] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," *Image and Vision Computing*, vol. 10, pp. 145–155, 1992.

[29] P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 947–954, 2005.

[30] K.W. Bowyer, K. Chang, and P.J. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Computer Vision and Image Understanding*, vol. 101, pp. 1–15, 2006.

[31] D. Smeets, P. Claes, D. Vandermeulen, and J.G. Clement, "Objective 3D face recognition: Evolution, approaches and challenges," *Forensic Science International*, vol. 201, pp. 125–132, 2010.

[32] O. Ocegueda, S.K. Shah, and I.A. Kakadiaris, "Which parts of the face give out your identity?," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 641–648, 2011.

[33] T. Faltemier, K.W. Bowyer, , and P.J. Flynn, "A region ensemble for 3D face recognition," *IEEE Trans. Information Forensics and Security*, vol. 3, no. 1, pp. 62–73, 2008.

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

[35] L. Spreeuwers, "Fast and accurate 3d face recognition," *Int. J. Computer Vision*, vol. 93, pp. 389–414, 2011.

[36] D. Huang, M. Ardabilian, Y. Wang, and L. Chen, "A novel geometric facial representation based on multi-scale extended local binary patterns," *Proc. IEEE Int. Conf. Automatic Face and Gesture Recognition*, pp. 1–7, 2011.

[37] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 19, pp. 696–710, 1994.

[38] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognitive Neuroscience*, vol. 37, no. 1, pp. 2–86, 1991.

[39] M.P. Segundo, C. Queirolo, O.R.P. Bellon, and L. Silva, "Automatic 3D facial segmentation and landmark detection," *Proc. Int. Conf. Image Analysis and Processing*, pp. 431–436, 2007.

[40] J.S. Lim, *Two-Dimensional Signal and Image Processing*, Prentice Hall, 1990.

[41] K. Arun, T. Huang, and S. Blostein, "Least-squares fitting of two 3-D point sets," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 9, no. 5, pp. 698–700, 1987.

[42] A.M. Bronstein, M.M. Bronstein, and R. Kimmel, "Expression-invariant representations of faces," *IEEE Trans. Image Processing*, vol. 16, no. 1, pp. 188–197, 2006.

[43] J. Wang, K.N. Plataniotis, J. Lu, , and A.N. Venetsanopoulos, "On solving the face recognition problem with one training sample per subject," *Pattern Recognition*, vol. 39, pp. 1746–1762, 2006.

[44] D.H. Wolpert, "Stacked generalization," *Neural Networks*, vol. 5, no. 2, pp. 241–259, 1992.

[45] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[46] K. Tumer and J. Ghosh, "Robust combining of disparate classifiers through order statistics," *Pattern Analysis and Applications*, vol. 5, no. 2, pp. 189–200, 2002.

Hoda Mohammazade received the BSc degree (with honours) from Amirkabir University of Technology (Tehran Polytechnic), in 2004 and the MSc degree from the University of Calgary, in 2007, both in Electrical Engineering. She is currently pursuing her PhD degree in Electrical Engineering at the University of Toronto. Her research interests include signal and image processing, pattern recognition and biometric systems.

Dimitrios Hatzinakos received the Diploma degree from the University of Thessaloniki, Greece, in 1983, the M.A.Sc degree from the University of Ottawa, Canada, in 1986 and the Ph.D. degree from Northeastern University, Boston, MA, in 1990, all in Electrical Engineering. In September 1990 he joined the Department of Electrical and Computer Engineering, University of Toronto, where now he holds the rank of Professor with tenure. Since November 2004, he is the holder of the Bell Canada Chair in Mutimedia, at the University of Toronto. Also, he is the co-founder and since 2009 the Director and the chair of the management committee of the Identity, Privacy and Security Institute (IPSI) at the University of Toronto. His research interests and expertise are in the areas of Multimedia Signal Processing, Multimedia Security, Multimedia Communications and Biometric Systems. Since 2008, he is an Associate Editor for the IEEE Transactions on Mobile Computing. He is a senior member of the IEEE, a Fellow of the Engineering Institute of Canada (EIC), and member of EURASIP, the Professional Engineers of Ontario (PEO), and the Technical Chamber of Greece.