Face recognition based on 3D ridge images obtained from range data

Mohammad H. Mahoor, Mohamed Abdel-Mottaleb

**ABSTRACT**

In this paper, we present an approach for 3D face recognition from frontal range data based on the ridge lines on the surface of the face. We use the principal curvature, \( k_{\text{max}} \), to represent the face image as a 3D binary image called ridge image. The ridge image shows the locations of the ridge points around the important facial regions on the face (i.e., the eyes, the nose, and the mouth). We utilized the robust Hausdorff distance and the iterative closest points (ICP) for matching the ridge image of a given probe image to the ridge images of the facial images in the gallery. To evaluate the performance of our approach for 3D face recognition, we performed experiments on GavabDB face database (a small size database) and Face Recognition Grand Challenge V2.0 (a large size database). The results of the experiments show that the ridge lines have great capability for 3D face recognition. In addition, we found that as long as the size of the database is small, the performance of the ICP-based matching and the robust Hausdorff matching are comparable. But, when the size of the database increases, ICP-based matching outperforms the robust Hausdorff matching technique.

**1. Introduction**

Face recognition has become one of the most important applications of image analysis and computer vision in recent years. This trend has caught the attention of many academic and research groups. Recently, the National Institute of Standards and Technology (NIST) initiated the face recognition grand challenge (FRGC) competition [1]. The goal of FRGC is to develop new face recognition techniques and prototype systems that have higher performance than existing systems by an order of magnitude over the face recognition vendor test (FRVT) 2002 [2].

There are three main contenders for improving face recognition algorithms: high resolution images, three-dimensional (3D) face recognition, and new preprocessing techniques. In current two-dimensional (2D) face recognition systems, changes in lighting (illumination), and pose of the face always have been challenging problems [3,4]. The 3D face recognition algorithms identify faces from the 3D shape of a person’s face. Because the 3D shape of a face is not affected by changes in lighting or pose, 3D face recognition has the potential to improve performance under these conditions [3].

As Bowyer et al. mentioned in Ref. [3], the literature appears to be split on whether using a single 3D face data outperforms using a single 2D face data or not. We believe that 3D data has more potential for face recognition and the results in Refs. [5–12] support this opinion.

In the literature, the 3D face recognition approaches using range images\(^1\) can be categorized into three categories: (1) PCA-based approaches [1,13–16], (2) feature-based approaches [11,17–20], and (3) surface matching approaches [8,10,21]. In the first category, similar to the 2D Eigenface recognition algorithm, principal component analysis (PCA) is applied to range data to reduce the dimension and then the recognition is performed by matching a probe image with gallery images in a lower dimension space. These approaches are simple, fast and straightforward, but they have low performance rate compared to the approaches in the other two categories. In the second category, the response of the range image or its representation at certain landmark points to a set of wavelet filters is calculated and considered as a set of features. Then, recognition is done based on the similarity of these features. Generally, these approaches are fast and have high performance rate compared to approaches in the other two categories, but localization of the landmarks is very important. For example in Ref. [22] the 2D texture images were used for landmark localization which means that the approach is not pure 3D. In the third category, researchers mainly utilize the ICP or Hausdorff distance (HD) to match the 3D surface points of a probe face to those of the face images in the gallery and then perform the recognition based on the minimum distance between the two faces. Also, the potential for face recognition and the results in Refs. [5–12] support this opinion.

\(^1\) The multi-modal 2D+3D approaches are excluded from this classification.
2. 3D face recognition based on ridge images

The algorithm starts by finding the locations of the ridge lines in the range images. These are the points around the eyes, the nose, and the mouth. We call a 3D binary image that contain only these lines a ridge image. For recognition, we use two different approaches, the directed robust HD and the ICP technique to find the best match for a given probe image from the facial range images in the gallery. In the following subsections, we explain in detail the process of extracting a ridge image and the process of matching ridge images using the robust HD and the ICP technique.

2.1. Extracting ridge images

For a given surface \( z = f(x, y) \), the mean curvature \( H \), the Gaussian curvature \( K \), and the principal curvatures \( k_{\text{max}} \) and \( k_{\text{min}} \) are defined as:

\[
K = \frac{f_{xx}f_{yy} - f_{xy}^2}{(1 + f_{xx}^2 + f_{yy}^2)^2} \quad H = \frac{f_{xx} + f_{yy} + f_{xx}f_{yy}^2 + f_{yy}f_{xx}^2 - 2f_{xy}f_{xx}}{2(1 + f_{xx}^2 + f_{yy}^2)^{1.5}}
\]

\[
k_{\text{max}} = H + \sqrt{(H^2 - K)} \quad k_{\text{min}} = H - \sqrt{(H^2 - K)}
\]

The mean curvature and the Gaussian Curvature of the points on 3D face surface are calculated based on the gradient defined in a local neighborhood. The principal curvatures are derived from \( H \) and \( K \).

Our goal is to extract and use the points lying on ridge lines as the feature points. These points correspond to the extreme ridge points on the considered surface. In the literature [24], the ridges are defined as the umbilic points at which the \( k_{\text{max}} \) attains a local positive maximum. An umbilic point is a point on a surface where the principal curvatures are equal and are non-zero (in the case of zero curvature, the point is called a flat point). Intuitively, ridges are the points that form the drainage patterns and are called valleys when the ridges are looked at from the opposite side.

There are different approaches to locate the ridges [25]. One of the main approaches applies thresholding which is used in this paper. We threshold the \( k_{\text{max}} \) values to find these points. The suitable threshold is obtained such that the highest recognition rate is achieved for a small training set that is different from the images in the gallery. Then, in our experiments the suitable threshold (a fixed value) is used for creating the ridge images for all the facial images in the databases under evaluation. Fig. 1 shows few examples of the ridges obtained by thresholding the \( k_{\text{max}} \) values. These are 3D binary images that show the locations of the ridge lines on the surface of the face. The lines on the boundary of the face are filtered out and are not considered as feature points for recognition.

Since captured range images have some artifacts, i.e., noise and gaps, we apply median filtering to remove sharp spikes that occur during scanning the face. Afterwards, we use interpolation to fill the gaps on the face region and finally we use a low pass filter to smooth the surface of the face that suffers from rapid changes due to facial hair or any other artifacts.

2.2. Matching ridge images using robust HD

Huttenlocher et al. originally proposed HD [26] as a measure for object matching in computer vision. Unlike other shape matching methods, HD can be calculated without knowing the exact correspondences of the points in different sets. Modifications to the HD
increase its capability to handle not only noisy points, but also missing data due to occlusion and outliers [27].

Given two sets of points \( \mathcal{A} = \{a_1, a_2, \ldots, a_{N_A}\} \) and \( \mathcal{B} = \{b_1, b_2, \ldots, b_{N_B}\} \) of size \( N_A \) and \( N_B \), respectively, the partial HD between the two sets of points \( \mathcal{A} \) and \( \mathcal{B} \) is defined as

\[
H(\mathcal{A}, \mathcal{B}) = \max\{h_K(\mathcal{A}, \mathcal{B}), h_K(\mathcal{B}, \mathcal{A})\}
\]

(2)

where \( h_K(\mathcal{A}, \mathcal{B}) \) and \( h_K(\mathcal{B}, \mathcal{A}) \) represent the directed distance between the two sets \( \mathcal{A} \) and \( \mathcal{B} \). The directed distances of the partial HD are defined as

\[
h_K(\mathcal{A}, \mathcal{B}) = K^{th} \text{ ranked value of } d_{\mathcal{A}}(a), \quad h_K(\mathcal{B}, \mathcal{A}) = K^{th} \text{ ranked value of } d_{\mathcal{B}}(b)
\]

(3)

where \( d_{\mathcal{A}}(a) \) represents the minimum distance (e.g., Euclidean distance) value at point \( a \) to the point set \( \mathcal{B} \), \( d_{\mathcal{B}}(b) \) represents the minimum distance value at point \( b \) to the point set \( \mathcal{A} \). \( K^{th} \) denotes the \( K^{th} \) ranked value of \( d_{\mathcal{A}}(a) \), and \( K^{th} \) denotes the \( K^{th} \) ranked value of \( d_{\mathcal{B}}(b) \).

After Huttonlocher et al.’s original work, researchers have proposed many different definitions and methods to realize directed HD. Dubbioso and Jain revised the original HD and investigated the performance of 24 different HD measures based on their behavior in the presence of noise [28]. They proposed the modified Hausdorff distance (MHD). Sim et al. applied the robust statistic techniques of regression analysis to the computation of the HD measures for object matching, resulting in two robust HD measures: M-HD based on M-estimation and least trimmed square-HD (LTS-HD) based on LTS [29]. Based on the experimental matching performance of these different HD measures, robust LTS-HD based on the LTS measure [29] is adopted in our work. In the proposed LTS-HD [29], the directed distance \( h_{LTS}(\mathcal{A}, \mathcal{B}) \) is defined by a linear combination of order statistics:

\[
h_{LTS} = \frac{1}{H} \sum_{i=1}^{H} d_{\mathcal{B}}(a_{(i)})
\]

(4)

where \( H \) denotes \( h \times N_{\mathcal{A}} \) (\( 0 \leq h \leq 1 \)) as in the partial HD case, and \( d_{\mathcal{B}}(x_{(i)}) \) represents the \( i \)-th distance value in the sorted sequence \( d_{\mathcal{B}}(x_{(1)}) \leq d_{\mathcal{B}}(x_{(2)}) \leq \cdots \leq d_{\mathcal{B}}(x_{(N_{\mathcal{A}})}) \). The measure \( h_{LTS}(\mathcal{A}, \mathcal{B}) \) is calculated by eliminating the large distance values and only keeping the \( h \) fraction of the smallest distances. In our experiments, the value of \( h \) that resulted in the best recognition rate was 0.8.

In our case, the calculation of LTS-HD is between the two point sets of two 3D binary images, one is the ridge image of the test face image and the second is the ridge image of a gallery face image. The process of finding the best pose between a probe ridge image and a gallery ridge image can be formulated as follows:

\[
\arg \min_{\alpha, \beta, \gamma, x, y, z} h_{LTS}(Tr(\mathcal{A}), \mathcal{B})
\]

(5)

where

\[
Tr = \begin{bmatrix}
\alpha R & \\
\beta R & \\
\gamma R & \\
0 & 1
\end{bmatrix}
\]

is a 3D similarity transformation, \( \alpha \) is a scale factor, \( T = [x, y, z] \) is the 3D translation, and \( R \) is a 3D rotation matrix with \( x, \beta, \gamma \) as rotation angels.

The process of finding the optimum pose between a probe ridge image and a gallery ridge image is achieved by an iterative approach as shown in Table 1.

![Fig. 1. Sample of ridges image extracted for different subjects.](image)

### Table 1

<table>
<thead>
<tr>
<th>Iterative algorithm to find the optimum pose in Hausdorff distance matching</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Set ( h_{3T} := +\infty ), and ( t := 0 )</td>
</tr>
<tr>
<td>(2) Initially align the 3D ridge image of the test image ( P ) (i.e., translate, rotate and scale), to the gallery image ( P' ), by using the three labeled feature points and the auxiliary point. This similarity transformation is calculated by procrustes analysis [30]</td>
</tr>
<tr>
<td>(3) Set Success := 0</td>
</tr>
<tr>
<td>(4) Place the aligned probe ridge image, ( T(P) ), over the gallery ridge image. For all the points in the aligned probe image, find the distance to the closest point in the gallery image, ( P' ), using:</td>
</tr>
<tr>
<td>( d_{\mathcal{A}}(x) = \min_{y \in P'}</td>
</tr>
<tr>
<td>where the ( | \cdot | ) denotes the L2 norm</td>
</tr>
<tr>
<td>(5) Sort the minimum calculated distances and then calculate the robust Hausdorff distance, ( h_{3T} ), using Eq. (4)</td>
</tr>
<tr>
<td>(6) If ( h_{3T} &lt; h_{3T} ), set the following items:</td>
</tr>
<tr>
<td>( h_{3T} := h_{3T} )</td>
</tr>
<tr>
<td>( t := t + 1 )</td>
</tr>
<tr>
<td>Success := 1</td>
</tr>
<tr>
<td>(7) Change the parameters of the similarity transformation, (i.e., translation, rotation, and scale)</td>
</tr>
<tr>
<td>(8) If Success = 1 AND (( t &lt; \text{Max}_\text{Iterations} )) goto 3</td>
</tr>
<tr>
<td>(9) Return ( h_{3T} )</td>
</tr>
</tbody>
</table>
We used the Matlab optimization toolbox (i.e., fminsearch Matlab function) to solve this problem. The fminsearch uses the simplex search method of Ref. [31]. This is a direct search method that does not use numerical or analytic gradients. This procedure is repeated to find the matching distance between a probe image and all the images in the gallery. The gallery face image that results in the minimum matching distance, is considered the best match.

2.3. Ridge matching using ICP

The ICP algorithm is widely used for geometric alignment of 3D models when an initial estimate of the relative pose is known. Many variants of ICP have been proposed, where the differences are in the phases of selecting, matching the feature points, and/or the minimization strategy. In this work, we use a fast ICP variant [32]. Instead of using random sampling of the feature points as in Ref. [32], we use all the feature points in the 3D ridge image in the matching process.

For the initial alignment of the ridge points, we utilize the similarity transformation between a set of labeled facial feature points on the probe and gallery images. Procrustes analysis [30] is used to estimate the parameters of this similarity transformation (scale, rotation, and translation). After the initial alignment, we use the aforementioned ICP algorithm to finely align a 3D ridge probe image with a given 3D ridge gallery image and compute the MSE between the points. The smaller the MSE the closer the probe image to the gallery image.

3. Experiments and results

We use the GavabDB database [33] and the FRGC2.0 [1] 3D face database for our experiments. In the following subsections we review these two databases and present our experiments and results.

3.1. Experiments on GavabDB

The GavabDB database contains 427 3D facial surface images corresponding to 61 individuals (45 males and 16 females). For each person, there are nine different images, two neutral frontal images, two neutral images with pose (looking down and up), two profile images, and three frontal images in which the subject presents different and accentuated facial expressions. The digitizer is a Minolta Vi-700 digitizer, a laser sensor which captures in less than a second a range image of the scene with color information. The individuals were placed near a large window without special focus. Fig. 2 shows the range images for one of the subjects in the database along with the textured images. The texture images for each person are not released and only the range images are available for public access.

In our experiments, we used the two neutral frontal images (the 1st and the 2nd captures), the two neutral looking up and down images (the 5th and the 6th captures), the frontal images with smile expression (the 7th capture), the frontal images with laughing expression (the 8th capture), and a frontal image with random gesture (the 9th capture). The images in the 2nd capture are used as gallery images and the images in the 1st, 3rd, 4th, and 7–9th captures are used as the probe images for recognition.

For recognition, we compared between the robust HD and the ICP techniques. For the initial alignment, we manually label three feature points (the two inner corners of the eyes and the tip of the nose) and use them for initial alignment. Table 2 presents the results of the experiments. For neutral frontal images, the rank-one identification rates were 93.5% and 95%, based on the HD and ICP techniques, respectively. In another experiment, we projected the frontal ridge images to 2D (ignoring the 3rd dimension) and the recognition process was tested. By ignoring the 3rd dimension,

Table 2

<table>
<thead>
<tr>
<th>Facial expression</th>
<th>1st rank recognition (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Robust HD</td>
<td>ICP</td>
</tr>
<tr>
<td>Neutral (3D)</td>
<td>93.5</td>
</tr>
<tr>
<td>Neutral (2D)</td>
<td>82.0</td>
</tr>
<tr>
<td>Smiling (3D)</td>
<td>82.0</td>
</tr>
<tr>
<td>Laughing (3D)</td>
<td>73.8</td>
</tr>
<tr>
<td>Random gesture (3D)</td>
<td>63.4</td>
</tr>
<tr>
<td>Looking up (3D)</td>
<td>75.4</td>
</tr>
<tr>
<td>Looking down (3D)</td>
<td>70.5</td>
</tr>
</tbody>
</table>

Fig. 2. 3D views of an individual: 3D images with the texture mapped, the same images without texture at 1/4 of the original resolution (both from scanner’s point of view) and a rotated version without the texture [33].
we obtained rank-one identification rate of 82.0% and 86.9% using robust HD and ICP, respectively. This result supports the opinion that 3D data has more potential for face recognition than 2D data.

For faces with expressions, we considered only the upper part of the face (i.e., the 3D ridge lines around the eyes and the nose) for recognition and excluded the lower part of the face (i.e., the mouth), which is highly affected by the expression. We achieved a recognition rate of 83.6% using the ICP technique and 82.0% using the robust HD for the smiling expression.

Furthermore, we evaluated the performance of our approach for recognition of facial images with pose (looking up/down) based on both the ICP and the robust HD techniques. As a result, the recognition rate for the facial images with looking up (down) pose are 88.6% (70.5%) and 75.4% (85.3%) using the robust HD and the ICP technique, respectively.

Our experiments show that the ICP technique outperforms the robust HD (except for the laughing expression).

We compared our algorithm with three different approaches for 3D face recognition that were presented by Moreno et al. in Refs. [15,16,34] based upon the GavabDB dataset. In Ref. [34], they segmented the range images into isolated subregions using the mean and the Gaussian curvatures. Then, they extracted 86 descriptors such as the areas, the distances, the angles, and the average curvature of the subregions. They selected 35 best features and utilized them for face recognition based on the minimum Euclidean distance classifier. They achieved a first rank recognition rate of 78.0% for neutral frontal images and 62% for images with smile expression (only 60 subjects out of 61 from the database were utilized). In Ref. [15], they selected a set of 30 features out of the 86 features and obtained recognition rates of 82.0% and 90.16% when the images are frontal views with neutral expression using principal component analysis (PCA) and support vector machines (SVM), respectively. The recognition rates decreased to 76.2% and 77.9%, using PCA and SVM matching schemes, respectively, when using probe images with expressions and slight face rotation. In Ref. [16], the authors represented the face using 3D voxels. Experiments were performed on both images with neutral expression and images with either pose variations or facial expressions. The best recognition rates that they achieved were 90.16% for the images with neutral expression and 77.9% for the images with pose and facial expressions. Table 3 summarizes their results as well as ours. As the results show, our method based on ridge images and the ICP technique for matching has a better recognition performance for images with neutral expression, with expressions, and with poses.

3.2. Experiments on FRGC2.0 face database

The FRGC2.0 database [1] consists of 50,000 recordings divided into training and validation partitions. The training partition is designed for training algorithms and the validation partition is for assessing the performance of systems in a laboratory setting. FRGC2.0 consists of six experiments, where the third experiment measures the performance of 3D face recognition. In experiment three, the gallery and probe datasets consist of both range and texture images for each subject. The 3D images were acquired by a Minolta Vivid 900/910 series sensor. There are 4007 pairs of images (range and texture) for 466 subjects in the validation set. The set contains images from 1 to 22 sessions per subject, including images with neutral expression and images with other expressions. 370 subjects have at least two neutral images and 432 subjects have at least one neutral image.

We investigated the performance of our method on the neutral 3D face images of the FRGC2.0 database. In the first experiment, we compared the performance of the robust HD and the ICP techniques for matching the ridge images. There are 370 subjects that have at least two neutral images captured in different sessions. For some of the subjects, there are more than two captured neutral images with a time lapse of one week between them. We chose the two farthest captured images for each subject and considered the oldest one as the gallery and the most recent captured as the probe. The result of rank-one identification using the robust HD on this selected dataset is 58.92% while the result of the ICP technique for matching is 91.8%. This means that the ICP-based matching approach not only gives the best performance, but also it is robust with the increase in the size of the database. To remind the reader, for a small size database such as Gavab, the performance of the Hausdorff matching and the ICP matching were comparable (ICP was slightly better). This conclusion is made by Yan and Bowyer in Ref. [35], where they compared ICP and Hausdorff for ear surface matching: The ICP outperforms the HD for shape matching.

In another experiment, we evaluated the capability of the ridge images for face verification on FRGC2.0 face database. Only the ICP technique was used for matching. For initial alignment of the ridge
In conclusion, there is a tradeoff between the performance and computational complexity in shape matching, our experiments show that ridge points are very promising in surface matching. More precisely, the use of ridge points results in negligible performance deterioration while reducing the computational complexity of matching.

Computation of the ICP technique for matching two 3D images had an average run time of 1.0 s and 0.11 s, for the entire face surface and for the ridge images, respectively, running on a 2.4 GHz dual core Pentium 4. The core algorithm of the ICP technique is written in C++ and compiled as a MEX function and is called in Matlab.2

4. Conclusions

This paper presented a method for 3D face recognition using range data based on 3D binary images, created using principal maximum curvature. A 3D binary image shows the locations of the ridge lines in the range facial image (i.e., lines around the eyes, the nose, and the mouth). Two different techniques for matching the 3D binary ridge images of a probe image and a gallery image were utilized: the robust Hausdorff distance (HD) and iterative closest points (ICP). Experiments on the GavabDB 3D face database and the FRGC2.0 3D face database show that the ridge lines are robust representations for 3D face recognition. In addition, the results also show that ICP matching technique outperforms the HD distance for large scale databases.

References


Table 4

<table>
<thead>
<tr>
<th>Database</th>
<th>Ridge points</th>
<th>Random points</th>
<th>Complete surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gavab (61 subjects)</td>
<td>95.0</td>
<td>67.2</td>
<td>95.0</td>
</tr>
<tr>
<td>FRGC V2.0 (370 subjects)</td>
<td>91.8</td>
<td>10.0</td>
<td>93.7</td>
</tr>
</tbody>
</table>

Table 5

Comparison between the ridge points, random points selection, and entire surface based on the ICP matching technique. Results are in terms of rank-one identification rate (%).

<table>
<thead>
<tr>
<th>Database</th>
<th>Ridge points</th>
<th>Random points</th>
<th>Complete surface</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>3D</td>
<td>FRGC baseline</td>
<td></td>
</tr>
<tr>
<td>ROC I</td>
<td>90.69</td>
<td>90.00</td>
<td></td>
</tr>
<tr>
<td>ROC II</td>
<td>88.5</td>
<td>86.01</td>
<td></td>
</tr>
<tr>
<td>ROC III</td>
<td>85.75</td>
<td>81.58</td>
<td></td>
</tr>
</tbody>
</table>


About the Author—MOHAMMAD HOSSEIN MAHOOR received the B.Sc. degree in electrical engineering from Petroleum University of technology, Iran, in 1996, the M.Sc. degree in biomedical engineering from Sharif University of technology, Iran, in 1998, and the Ph.D. in electrical and computer engineering from University of Miami, Florida in 2007. He then worked as Post-Doctoral fellow at the University Miami till August 2008. Currently, he is an assistant professor at the department of electrical and computer engineering at the University of Denver. His research interests are pattern recognition, image processing and computer vision.

About the Author—MOHAMED ABDEL-MOTTALEB received his Ph.D. in computer science from University of Maryland, College Park, in 1993. He is a professor in the department of Electrical & Computer Engineering, University of Miami, Coral Gables, where his research focuses on 3D face recognition, dental biometrics, visual tracking, and human activity recognition. Prior to joining the University of Miami, from 1993 to 2000, he worked at Philips Research, Briarcliff Manor, NY. At Philips Research, he was a Principal Member of Research Staff and a Project Leader, where he led several projects in image processing, and content-based multimedia retrieval. He holds 20 US patents and published over 85 papers in the areas of image processing, computer vision, and content based retrieval. He is an associate editor for the Pattern Recognition Journal.