Combining Statistics of Geometrical and Correlative Features for 3D Face Recognition

Yonggang Huang[1], Yunhong Wang[2], Tieniu Tan[1]
[1]National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences, Beijing, China
{yghuang,tnt}@nlpr.ia.ac.cn
[2]School of Computer Science and Engineering
Beihang University, Beijing, China
yhwang@buaa.edu.cn

Abstract

In this paper, we present a new method for face recognition using range data. The proposed method is based on both global statistics of geometrical features and local statistics of correlative features of facial surfaces. Firstly, we analyze the performances of common geometrical representations by using global histograms for matching. Secondly, we propose a new method to encode the relationships between points and their neighbors, which are demonstrated to own great power to represent the intrinsic structure of facial surfaces. Finally, the two kinds of features are supposed to be complementary to some extent, and the combination of them is proven to be able to improve the recognition performance. All the experiments are performed on the full 3D face dataset of FRGC 2.0 which is the largest 3D face database so far. Promising results have demonstrated the effectiveness of our proposed method.

1 Introduction

Face recognition attracted much attention from researchers recently, due to its broad application and non-intrusive nature. A large amount of work has been done on this topic based on 2D color or intensity images. Though a certain level of success has been achieved by many algorithms with restricted conditions, severe problems caused by large variations of illumination, pose, and expression remain unsolved. To deal with these problems, many researchers move attention to some other facial modalities, such as range data, infrared data, and etc.

Human face is actually 3D deformable object with texture. Most of its 3D shape information is lost in 2D intensity images. The 3D shape information should not be ignored since they may provide distinct feature to distinguish different faces. With recent advances of 3D acquisition technology, 3D facial data capture is becoming easier and faster, and 3D face recognition is attracting more and more attention[1].In earlier days, many researchers worked on curvatures from range data[2]. Chua et al. [3] divided the

face into rigid and non-rigid regions, and used point signature to represent the rigid parts. Since recognition was conducted only based on those rigid parts, their method achieved a certain level of robustness to expression variance. Medioni et al. [4] used Iterative Closest Point (ICP) to align and match face surfaces. ICP-based methods were also adopted by many other researchers, such as Lu [5]. Xu et al. [6] used Gaussian-Hermite moments to describe shape variations around important areas on face, then combined them with the depth feature for recognition. Brostein et al. [7] achieved expression-invariant 3D face recognition by using a new representation of facial surface ,which is invariant to isometric deformations. Subspace methods were also used for 3D face recognition based on range images [8].

From previous work, we can see that extracting discriminative features from range data is one of the key problems for 3D face recognition. Some of the researchers used geometrical features, such as curvature, distances of fiducial points, profile etc [2]; while some used feature learned by subspaces [8]. Some feature extraction algorithms are model-based [4,5,6]; while some are range images based[8]. Many researchers applied methods developed for 2D intensity images to range images. But the effects were not so promising as expected, because range images and intensity images are intrinsically different. It is of great significance to extract discriminative and robust features from 3D face surface according to their special property.

In this paper, we firstly analyze some common geometrical features of facial surface by using global histograms for matching. Though the expression variance can affect the local points' geometrical features, we suppose that global statistics of those geometrical features should remain some intrinsic invariance. Moreover, we suppose that the relationships between points and their neighbors on the facial surface should own more potential power to describe the structure of faces than individual points. We define these relationships between points as "correlative features" in this paper. To obtain correlative features of 3D facial points for recognition, we propose a new method: 3DLBP. This method is based on Local Binary Patterns (LBP), which is originally proposed for texture analysis [9]. Finally, we suppose the global statistics of geometrical features (GSGF) and the local statistics of correlative features (LSCF) can complement each other to some extent. By combining them for 3D face recognition, we achieved encouraging results. All of our experiments in this paper were conducted on the 3D face database of FRGC 2.0, which contains 4007 range data of 466 different persons.

The rest of this paper is organized as follows. Section 2 illustrates the global statistics of geometrical features. Section 3 presents the details of 3DLBP. Experimental results and discussion are given in Section 4. Finally we conclude this paper in Section 5.

2   Global statistics of geometrical features (GSGF)

2.1   Geometrical features

Since human face is a 3D object, it is natural to represent it with geometrical features of their surfaces. Different geometrical features can capture different characteristics of their shapes. We will test four categories of common geometrical features here: depth, surface normal, gradient and curvature.

Depth is directly given by the range image after the original 3D point cloud is preprocessed. Each pixel value of the range image represents the relative distance from a

facial point to the camera. All pixel values are normalized to [0 255]. The nose tips are set to 255, which means that the nose tip is the nearest point to the camera after pose alignment. When expression changes, some facial regions' depth values will get smaller, while some will get bigger. So there should be some neutralization as a whole. If the neutralization can make variations within person be smaller than variations between persons, global statistics can be qualified for recognition.

Surface normals and gradients are simple geometrical description for surfaces. They are easy to calculate. Both horizontal and vertical gradients were tested. Normal of a point on the range image is represented as a vector (n_x, n_y, n_z). For the convenience and efficiency of comparison, we use a pair of angles (ϕ, θ) in sphere coordinates to represent a normal [11]. ϕ and θ are calculated as follows:

$$\phi = \arctan(\frac{n_z}{n_y}), \theta = \arctan(\frac{\sqrt{(n_y^2 + n_z^2)}}{n_x}) \tag{1}$$

Curvatures of facial surface have been studied much for 3D face recognition in the past two decades. Common curvatures include mean, Gaussian, and principle curvatures. Most previous work used the signs of mean and Gaussian curvatures for surface segmentation and interpretation. Exact values of curvatures were seldom used because they are unstable on the noisy data. Curvatures can be easily calculated from the first and second fundamental forms of range images. Based on curvatures, Shape Index (S_I) is a quantitative measure of a surface at a point, and curvedness (R) describes how curved a surface is [12]. They are defined as follows.

$$S_I(p) = \frac{1}{2} - \frac{1}{\pi} \arctan(\frac{k_1(p) + k_2(p)}{k_1(p) - k_2(p)}), R(p) = \sqrt{(k_1^2(p) + k_2^2(p))/2} \tag{2}$$

where k_1 and k_2 are the principal curvatures at point p, with $k_1 \geq k_2$. The range of the shape index is [-1 1].

2.2   Globally statistical analysis using histograms

Histograms have been widely used in color images for object recognition and image retrieval. In spite of their extreme simplicity, they have been shown to be efficient. There are two main reasons for us to choose histograms as a tool for statistical analysis of these geometrical features: One is that histograms take these local features as global statistics. So even if local shape changes, global statistics of these features may preserve some stable properties as a whole; the other is that histograms are robust to noise and occlusions to some extent.

All features are normalized to [0 255] for histograms establishment. There are many similarity measures for histogram matching. Here we use χ^2−divergence:

$$\chi^2(Q, V) = \sum_i \frac{(q_i - v_i)^2}{q_i + v_i} \tag{3}$$

where Q and V denote histograms, q_i and v_i denote the values of ith bin of Q and V.

3 3D Local Binary Patterns

Local Binary Patterns (LBP) operator is first proposed by Ojala et al. [9] for texture description. It has been successfully applied to 2D face recognition by Timo et al. [10]. Also we have tested it for 3D face recognition. Motivated by the original LBP, we propose 3D Local Binary Patterns (3DLBP) operator to obtain local correlative features of facial surfaces. In 3DLBP, not only the original LBP are included, but also the information of depth differences is encoded into binary patterns.

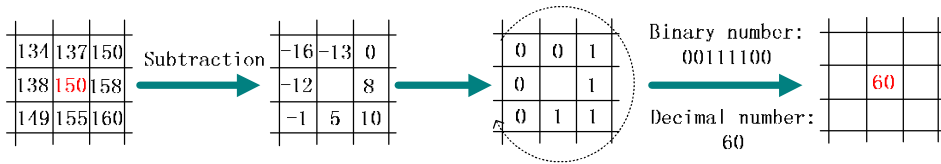The main idea of the original LBP operator is described in Figure 1.

Figure 1: An example of Basic LBP operator.

As shown in Figure 1, every pixel in the image is subtracted by their neighbors at the first step. Then the differences are converted to binary units: 0 or 1 according to their signs. At the third step, binary units are arranged clockwise. By doing this, we can obtain a set of binary units as the local binary pattern of the pixel. The binary pattern is further transformed to decimal number or "uniform pattern" (see [9] for more details). Two parameters (P, R) are used to control the selection of the number of neighbors (P) and their locations' radius (R). They can be varied as (8, 2), (16, 2), (24, 3) etc.

From the process above, we can see that LBP operator actually encodes pixels' relationships with their neighbors. We name these "relationships" as "correlative features" in this paper. So LBP can be seen as a kind of local correlative features. In our opinion, the structure information of facial surface should exist in the correlative features of points on the surface. According to this intrinsic property of LBP analyzed above, we suppose LBP operator should own the potential power to encode structure information of 3D face surface.

Furthermore, in the process of LBP operator, only encoding signs of depth differences is not adequate for describing 3D faces, because different depth differences on the same point of the facial surfaces distinguish different faces. See a specific example in Figure 2. Though A and B are two different persons, LBP of their nose tips are the same, because all the points around the nose tips are 'lower' than them. So we can see that, if two facial regions of different persons on the same place have the same trend of depth variation, LBP is inadequate to distinguish them. However, though the signs of differences between the two nose tips and their neighbors are the same, the exact values of the differences are different. We consider this clue should be crucial for 3D face recognition.

So we further encode the exact values of differences into binary patterns as what is shown in Figure 2. According to our statistical analysis, more than 93 % percent of the depth differences between points in $R = 2$ are smaller than 7, so we add three binary units to encode each depth difference between a pixel and its neighbor. Three binary units ($\{i_2i_3i_4\}$) can correspond to the absolute value of depth difference (DD): $0 \sim 7$. All the $|DD| \geq 7$ are assigned to 7. Combining with signs of the differences denoted by 0, 1 as the

head binary unit(i_1) like what the original LBP does, we finally obtain a set of 4 binary units $\{i_1i_2i_3i_4\}$ to denote *DD* between two points as follows.

$$i_1 = \begin{cases} 1, & DD \geq 0, \qquad\qquad (4) \\ 0, & DD < 0. \qquad\qquad (4') \end{cases}$$

$$|DD| = i_2 \cdot 2^2 + i_3 \cdot 2^1 + i_4 \cdot 2^0 \qquad\qquad (5)$$

A specific example is shown in the blue window of Figure 2.

Four binary units are divided into four layers. The binary units of each layer are arranged clockwise. Finally, we will get four decimal numbers: P_1, P_2, P_3, P_4 at each pixel point as its representation(see Figure 2 for an example). We call these as 3D Local Binary Patterns (3DLBP). For matching, 3DLBP are first transformed into four maps according to P_1, P_2, P_3, P_4 respectively: 3DLBPMap1(equal to original LBPMap), 3DLBPMap2, 3DLBPMap3, 3DLBPMap4. Then histograms of local regions of the four maps are concatenated as local statistics of correlative features (LSCF) for matching.
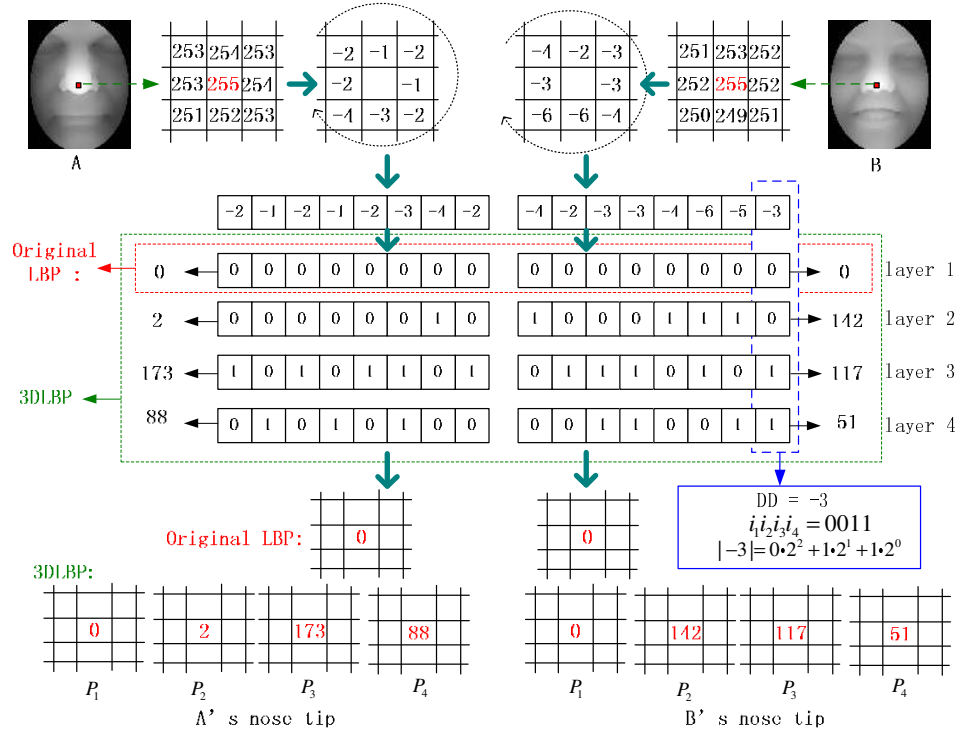

Figure 2: An example of 3DLBP and its comparison to LBP.

4 Experimental results and discussions

We designed three experiments on the full set of 3D data in FRGC 2.0 database to verify our proposed method. First, performance of global statistics of different geometrical

feature are tested and analyzed; then 3DLBP are examined and compared with other methods. Finally, we test the performance of the whole scheme by combining GSGF and LSCF. The experimental results are reported as follows.

4.1 Database description and preprocessing

FRGC 2.0 [13] is a benchmark database released last year. To our best knowledge, the 3D face dataset of FRGC 2.0 is the largest 3D face database till now. It consists of six experiments. Our experiments belong to Experiment 3 which measures the performance of 3D face recognition. The 3D dataset is divided to training and validation partitions. Validation dataset contains 4007 shape data of 466 different persons. All the 3D data were acquired by a Minolta Vivid 900/910 series sensor in three semesters in two years. The 3D shape data are given as raw point clouds, which contain spikes and holes. And manually labeled coordinates of the eyes, noses, and chin are given.

Registration, filling holes and moving spikes, and normalization were carried out on the raw data in succession as processing. After that, the data points in three dimension coordinates were projected and normalized as range images; the regions of faces are cropped and noses were set in the same position; a mask was also used to remove marginal noises. Figure 3 shows some samples.
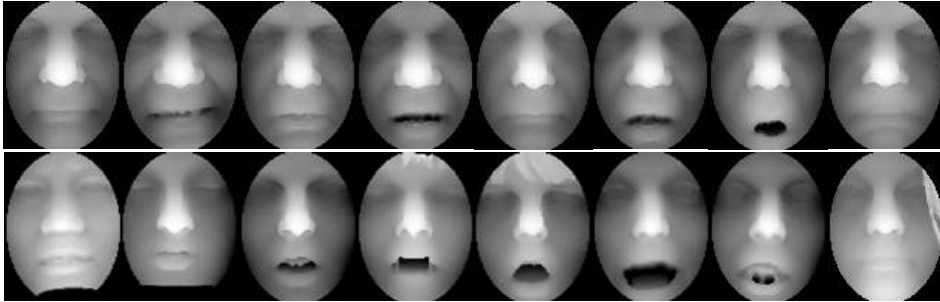

Figure 3: Samples from FRGC 2.0 (after preprocessing). The first row: images from the same person. The second row: some samples of "bad" images suffered from data loss, big holes, and occlusions caused by the hair.

From Figure 3, we can see that FRGC 2.0 is a very challenging database. And some "bad" images like the ones in the second row still exist in the database after the preprocessing step due to the original data and our preprocessing method which is not powerful enough to handle those problems. However, our main focus in this work is not on data preprocessing, but the recognition method. Though those bad images will greatly challenge our final recognition performance, we still carried out all the experiments on the full dataset.

Three ROC curves are required for performance analysis in Experiment 3 of FRGC 2.0. ROC I: gallery and probe are within semester; ROC II: gallery and probe are within year; ROC III: gallery and probe are between semesters. In the following, ROC I curves are reported in the first and second experiments, all the three curves are reported in the third experiment.

4.2 Global statistical analysis of geometrical features

This experiment aims to evaluate the performance of using global statistics of geometrical features (GSGF) for 3D face recognition. Some conclusions about geometrical features of facial surface have been drawn in other researchers' previous work. Here, we analyze them from a new point of view as the theory explanated in Section 2. Global histograms are used as the global statistics of several common geometrical features on facial surfaces. We also combined some of them because different geometrical features are supposed to be complementary in our opinion. ROC I curve are reported as follows:
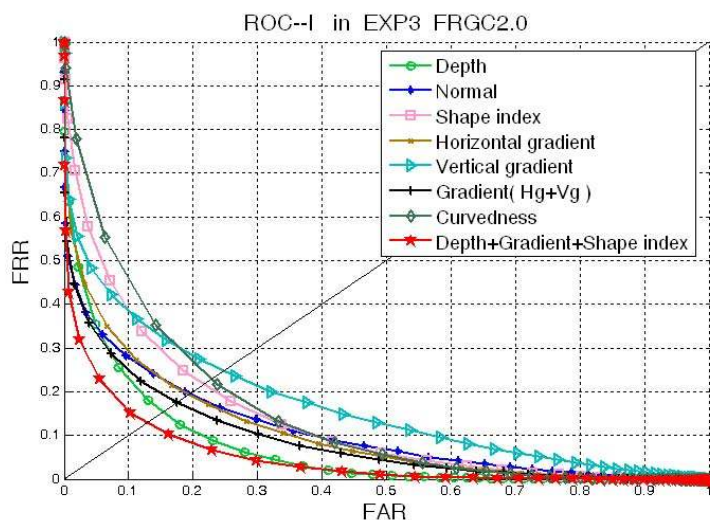

Figure 4: Globally statistical analysis of geometrical features.

From Figure 4, we can see that: 1) as the simplest form of geometrical features, depth provides the lowest EER(Equal Error Rate) among depth, normal, curvature (shape index, curvedness), gradients by using global statistics. The reason lies in two aspects: On one hand, depth images inevitably contain noises since our data are not perfectly acquired. Other geometrical features are calculated from depth images, so that the noises are magnified in obtained features. On the other hand, depth features are easy to neutralize the impact of local variation within person due to their simpleness and straightforwardness. 2) different geometrical features can complement each other to some extent. We tested different combinations. And the best result we obtained is shown as the red curve which is a combination of depth, gradient and shape index. We can see that the combination outperforms all the individual geometrical features. 3) though the combination of different geometrical feature can achieve EER 12.9%, we do not think this result is good enough. Correlative features should also be explored besides these geometrical features.

4.3 Performance evaluation of 3DLBP

This experiment is designed to evaluate the performance of our proposed 3DLBP operator. Each 3DLBP map was cropped into five parts as forehead, eyes, nose and both of its

sides, mouth, and chin. Local histograms of each part are concatenated for matching. The parameters (P, R) were selected as (8, 1), (8, 2), (16, 2), and the best result was reported. Also other three methods were compared with 3DLBP in the experiment. See Figure 5.
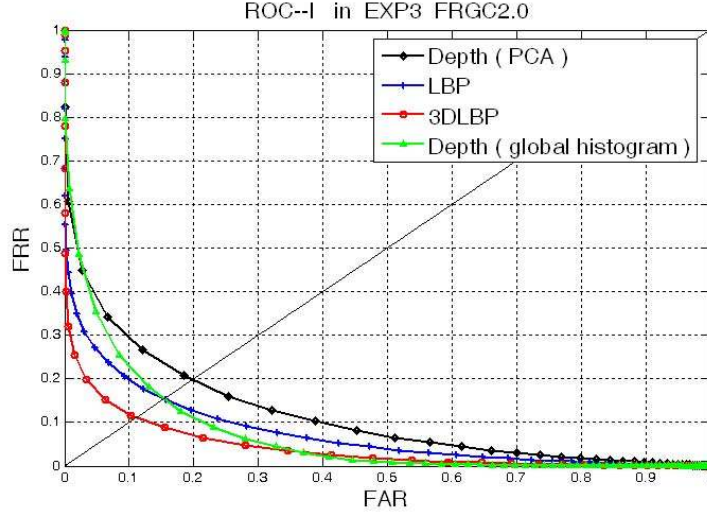

Figure 5: ROC I curves of 3DLBP, LBP, depth+PCA, and depth+global histogram.

From Figure 5, the red one of the curves denotes the performance of 3DLBP. We can see that it significantly outperforms other three methods.

Pixel depth is a typical kind of point feature, which has been demonstrated to be superior to other geometrical features in the first experiment, while 3DLBP is a kind of correlative feature. That 3DLBP outperforms depth supports our supposition in Section 1, which is that relationships between points and their neighbors on the facial surface should own more potential power to describe the intrinsic structure of faces than the individual points.

That 3DLBP outperforms the original LBP method proves 3DLBP's superiority for 3D face recognition than LBP. By encoding the depth differences besides the signs, 3DLBP is able to encode more 3D information of facial surface than the original LBP. This property of 3DLBP makes it especially suitable for 3D face recognition.

4.4 Combination of GSGF and 3DLBP

We have analyzed two different kinds of features in the previous two experiments: geometrical features of individual points, and correlative features between points. Though correlative features have been demonstrated to outperform geometrical features, they are supposed to be indispensable for describing 3D faces. Due to their different properties, they may be able to complement each other to some extent. So the combination of the two kinds of features should provide richer information for facial surfaces than each of them. To verify this hypothesis, we combined global statistics of geometrical features (GSGF) and 3DLBP (LSCF) by concatenating their feature histograms. Experimental results are shown in all the three ROC curves as Figure 6, 7, 8.
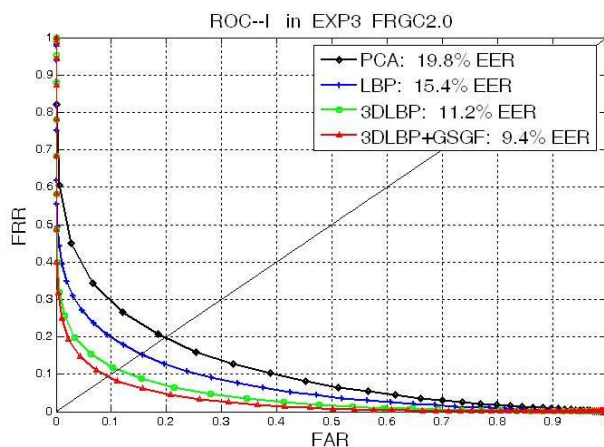
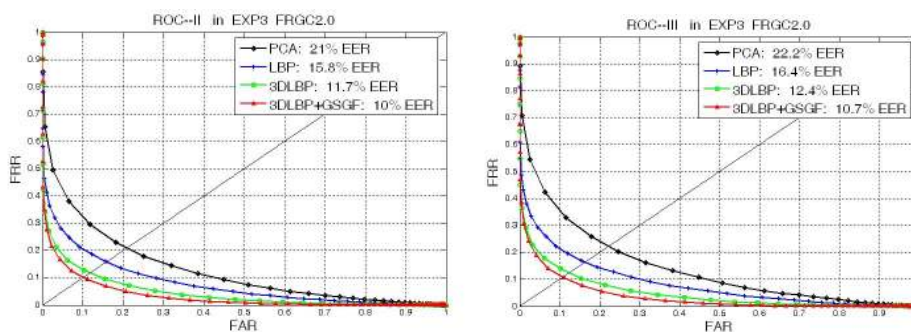Figure 6: ROC I curves of the performance of combining GSGF and LSCF.

Figure 7: ROC II curves of GSGF + LSCF. Figure 8: ROC III curves of GSGF + LSCF

From three ROC curves of experiment 3 in FRGC 2.0, we can see that, by combining 3DLBP and GSGF, we achieved the best results among the four methods. Significant improvement should attribute to the complementary property of GSGF and LSCF, and the combination of them provides more distinct information for recognition. So we can draw a conclusion now: our method: GSGF+LSCF(3DLBP) is a good representation of facial surfaces for 3D face recognition.

5 Conclusion

In this paper, we proposed a new scheme for 3D face recognition based on range images. Our contributions include: 1) We evaluated several common geometrical features by using global histograms. The results show that global statistics can be a good representation of geometrical features for 3D face recognition. 2) We proposed a new method (3DLBP) to encode correlative features of points on facial surface, which has been proved to perform much better than geometrical features. 3) By combining GSGF and 3DLBP (LSCF), we

got a powerful scheme to encode the intrinsic structure information of facial surfaces for recognition.

All experiments in this work were performed on the full set of 3D face data in FRGC 2.0. Encouraging results have demonstrated the effectiveness of our proposed methods. What's more, a superior preprocessing method and other better classification methods, such as SVM, can be adopted to further improve the final performance of our method.

6 Acknowledgement

References

[1] K. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition", *Computer Vision and Image Understanding*, 2005.

[2] Gaile Gordon, "Face Recognition from Depth Maps and Surface Curvature", *Proc. of SPIE, Geometric Methods in Computer Vision*, San Diego, July 1991. Vol. 1570.

[3] C. S. Chua, F. Han, and Y. K. Ho, "3d human face recognition using point signature", *Proc. International Conf. on Automatic Face and Gesture Recognition*, March 2000, pp. 233–238.

[4] G.Medioni, R. Waupotitsch, "Face recognition and modeling in 3D", *IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, Octorber 2003, pp. 232-233.

[5] Xiaoguang Lu, Dirk Colbry, Anil K. Jain, "Three-Dimensional Model Based Face Recognition", *17th International Conference on Pattern Recognition (ICPR'04)*, Volume 1, 2004.

[6] C. Xu, Y. Wang, T. Tan, and L. Quan, "Automatic 3d face recognition combining global geometric features with local shape variation information", *FGR*, pages 308-313,2004.

[7] A. Bronstein, M. Bronstein, and R. Kimmel, "Expressioninvariant 3D face recognition", *Proc. Audio and Videobased Biometric Person Authentication*, 2003, pp. 62–69.

[8] B. Achermann, X. Jiang, and H. Bunke, "Face recognition using range images", *Proceedings of International Conference on Virtual Systems and MultiMedia'97*, pages 129–136, Sept.1997.

[9] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multi resolution gray-scale and rotation invariant texture classification with local binary patterns", *PAMI*, Volume 24 , Issue 7 (July 2002).

[10] A. Timo, H. Abdenour, and P. Matti, "Face recognition with Local Binary Patterns", *ECCV2004*, pages 469-481.

[11] Guenter Hetzel, Bastian Leibe, Paul Levi, Bernt Schiele, "3D Object Recognition from Range Images using Local Feature Histograms", *CVPR*, 2001. Vol. 2, pages 394-399,.

[12] Chitra Dorai, Anil K. Jain, "COSMOS A Representation Scheme for 3D Free-Form Objects ", *IEEE Trans. on PAMI*, volume 19, pages 1115-1130, 1997.

[13] 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. of CVPR*, 2005.