2016 1st International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia

Expression and Occlusion invariant 3D face recognition based on region classifier

Koushik Dutta
Department of Computer Science and Engineering
Jadavpur University
Kolkata, India
koushik.it.22@gmail.com

Debotosh Bhattacharjee
Department of Computer Science and Engineering
Jadavpur University
Kolkata, India
debotoshb@hotmail.com

Mita Nasipuri
Department of Computer Science and Engineering
Jadavpur University
Kolkata, India
mnasipuri@cse.jdvu.ac.in

***Abstract*— In recent year, 3D face images play a major role in face recognition over corresponding 2D images. In this work, the 3D range images are used for face recognition based on region classifiers. Three different regions: eye, nose, and mouth separately classify a face image locally. At first, local binary patterns (LBP) are calculated for all pixels on face images. A new image formed with these LBP values are cropped and then divided into three horizontal regions, namely eye, nose, and mouth. For each such region histogram of oriented gradient (HOG) is used for feature vector creation. Two databases: Frav3D of 106 different subjects and our database consisting of 102 various individuals are used for recognition. Only frontal images with occlusion, expression and neutral images from those databases are utilized for this proposed system. Two fold cross validation technique with a nearest centroid-based classifier is used for classification. In the case of decision level fusion, recognition accuracy is 88.86% on Frav3D database and 77.5% for our newly created database. On the other hand, score level fusion has shown 78.5% recognition accuracy for FRAV3D database and 65.55% for our database.**

Keywords—3D range image; Region classifier; Decision level fusion; Score level fusion; Local binary pattern; Histogram of the oriented gradient.

I. INTRODUCTION

Security is one of the important issues in modern life. To solve the problem of safety, a biometric-based security system can be developed. There are various biometric traits, such as the face, iris, signature and hand geometry can be used for the safety or authentication purpose; out of those face is the more advantageous one than others. During the past few years, various techniques have been used for recognizing faces automatically as in [1]. Till now, face recognition is a challenging task in computer vision and pattern recognition. The process of face recognition matches a test face with faces stored in the database to find a possible candidate. With the technological advancements, face recognition can also be done using various applications in smart devices like tablet, i-phone, etc. In various security based systems, like mobile banking systems, mobile based face recognition is used. Face recognition can be possible using both 2D and 3D face images [2]. Face recognition in 3D is useful for solving various issues like pose variation, which is not possible in 2D. Also, the change in illumination does not affect 3D face recognition. Structured light scanner (SLS), Laser scanner or Kinect are used to capture the 3D face images. After capturing, the face data are stored as 3D point clouds, from which range image or 2.5D image can be generated.

In our work, face recognition is done based on region classifier, which is a new approach in 3D face recognition. In the case of a region, the classification process is done locally. Various regions of a particular 3D face are used to classify the face locally to a particular class. Then decision level [3] and score level [4] fusions are employed to combine the outputs of the different region classifiers for global classification of the face image. Working with a 3D face, four categories of information related to facial geometry, namely, depth, gradient, curvature and surface normal can easily be extracted. In our work, local binary pattern (LBP) is applied on the depth matrix. After that, the LBP matrix is divided into three different regions: eye, nose, and mouth. Then the histogram of oriented gradient (HOG) has been calculated for the various regions for computing the feature vectors. Here occlusions of eye and mouth of the input faces are considered in the proposed algorithm.

The rest of the paper has been organized as follows: Section II describes the review work on different techniques of 3D face recognition. Our proposed system has been explained in section III. Section IV discusses the experimental results. A comparative study of other works has been presented in section V. Finally, Section VI draws the conclusion.

II. REVIEW WORK

During the past two decades, various works have already been done on 3D face recognition, of which some significant works are discussed here. The 3D face recognition can be categorized into three different approaches: (1) Holistic approach, (2) Surface matching-based approach and (3) Local feature based approach.

In the first category, techniques use the information of the whole 3D face image represented as a 3D point cloud or range image. These types of techniques reduce the dimension of input data. In the case of recognition using 3D range image, these techniques work in the same way as with 2D intensity image. Some examples of holistic approach are principle component analysis (PCA) [5-7], linear discriminant analysis (LDA) [8, 9], and independent component analysis (ICA) [10].

In the surface matching-based approach, 3D faces are recognized using curvatures, surface normal and also some other 3D descriptors. In [11], authors have introduced a new approach to 3D face recognition based on curvature analysis of 3D range image. Gaussian, mean and principle curvatures are extracted at each pixel position from gradient values. Singular Value Decomposition (SVD) is used to calculate feature vector for further classification by an Artificial Neural Network (ANN). Highest 86.51% of recognition rate has been obtained on Frav3D database. In [12], authors have proposed ICP based registration followed by feature extraction based on the variation of depth values of the surface normal and also KPCA. They have used three different databases: Frav3D, Gavab, and Bosphorous for testing their technique. They have acquired 96.92%, 96.25% and 92.25% of recognition rate in case of GavabDB, Bosphorous, and Frav3D database.

In the last case, features can be extracted locally from any position on the face. In [13], geometrically localized feature is used for 3D face recognition. Authors extract three curvatures, eight invariant facial feature points, and their related features. Depth based DP (Dynamic Programming) and Feature based Support Vector Machine (SVM) classifiers are used for face recognition algorithm. They achieved a recognition rate of 96% on a dataset consisting of face images of 100 people. In [14] an illumination- and expression-invariant 3D face recognition system has been proposed based on three local descriptors: local phase quantization (LPQ), three-patch local binary pattern (TPLBP) and four-patch local binary pattern (FPLBP). Further, the histogram is calculated on the matrix of three different descriptors. Side by side PCA and LDA are used for feature reduction. Finally, all features are passed through SVM and produce maximum 98% recognition rate on Casia3D database.

III. Overview of Proposed System

Our proposed system has been divided into four main parts: 3D Image Acquisition, Pre-processing, Feature Extraction, and Classification. Figure 1 represents the flow diagram of our proposed system.

A. *3D Face Image Acquisition*

In our proposed system 3D range images of Frav3D [15] database and the 3D face database created in our laboratory are used as input. 2.5D image or range image are the matrix representation of 3D point clouds; where the 3D image is stored as a 2D array, each element (x,y) of the array contains the z-value of a point (x,y,z) in the point cloud. Further, the depth matrix has been normalized in the range 0 to 255; where 0 represent lowest depth and 255 represent the highest depth. 3D images can also be represented in another form like mesh, triangular mesh. Figure 2 represents different forms of a 3D face image along with the corresponding 2D image.

In our work, frontal 3D images with expression variation available in the Frav3D database are considered. Minolta Vivid-700 red laser strip scanner is used to capture the 3D image data of the Frav3D database. The Frav3D database has all total 106 different subjects; where each subject contains 16 different faces of pose, expression and illumination variation. Here, six face images of frontal pose consist of neutral and other expression variations are considered.

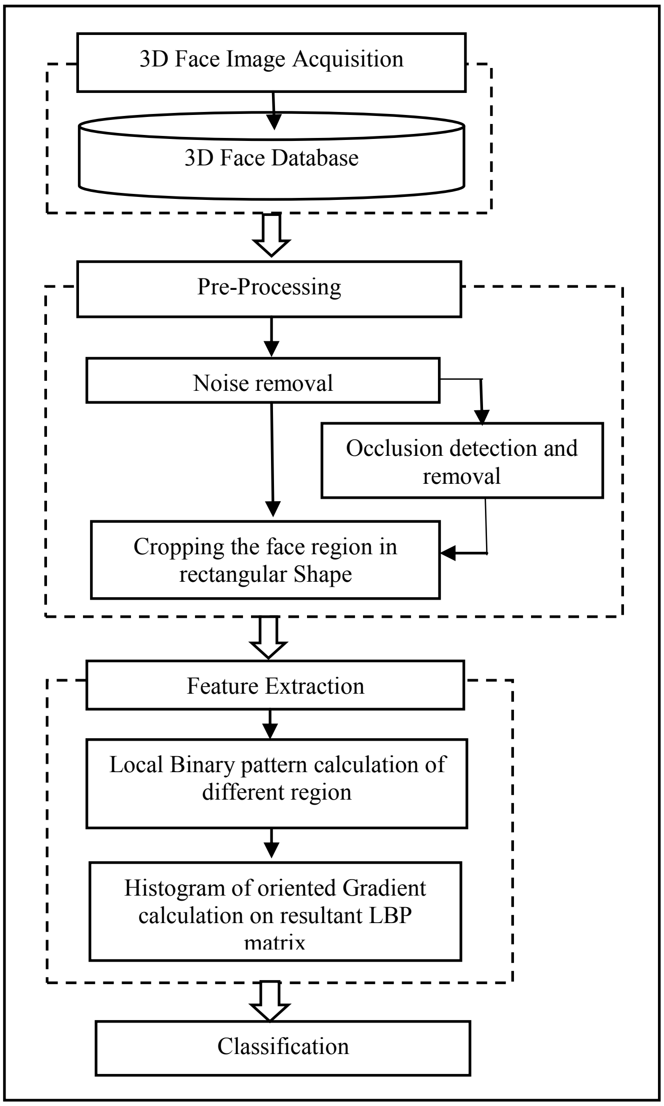


Fig. 1. Flow diagram of proposed work

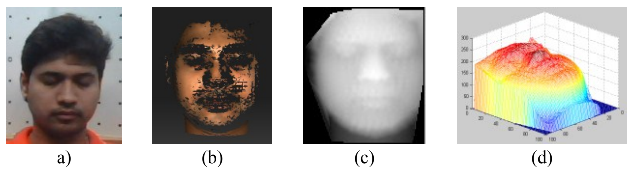
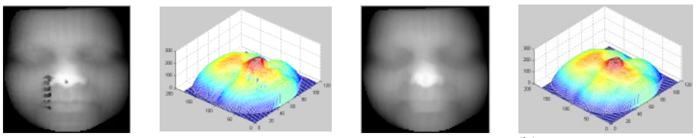


Fig. 2. Different form of face image: (a) 2D image (b) 3D point clouds (c) 2.5D image (d) 3D mesh

On the other side, our database has 102 various subjects; where 3D face images of each subject contain variations in pose, expression, occlusion variation and their combinations. The 3D face image of our database has been captured using Kinect sensor; where 24 different patterns are projected on the object for scanning 3D data. Here, in our proposed work, 12 different frontal faces are considered consisting of various expressions, occlusion, and neutral frontal face images. There are six different expressions in the frontal pose: Anger, surprise, disgust, fear, happy and sad; four types of occlusions, namely, occlusion of mouth by hand, occlusion by hair, finally two neutral pose.



(a) (b)

Fig. 3. Range image and mesh form (a)before smoothing (b) after smoothing

B. Preprocessing

In the preprocessing stage, various noises like holes, spikes are removed from raw input 3D image. The various smoothing technique can eliminate the noise. Here weighted median filter [16] is used for smoothing the 3D surface. In the weighted median filter, at first 9×3 filter mask, is employed in the depth image by considering weight value equals to 1. The pixel values within that mask are sorted, and the middle element (14^{th} element) is used to replace the value of the pixel location coinciding with the center of the mask. Figure 3 (a), (b) illustrated the smoothing of a raw 3D image.

1) For Occlusion Detection and Removal: There are some occluded images in our database. They contain four types of occlusions: mouth occluded by hand, eyes occluded by handkerchief and power free glass, and occlusion by a hair. At first, the occluded portion has been detected and after that we remove the occluded portion.

Here, we use fuzzy C-means clustering [17] for removing the mouth occlusion by hand. Fuzzy C-means clustering have segmented an image into different clusters. Now, for calculating the cluster numbers, we use the Davies-Bouldin index [18] concept. Here, according to the Davies-Bouldin index we consider cluster value equal to 4. Figure 4(a) shows input occluded range image, and 4 (b)-(e) illustrate different clustered image.

Now from these clusters, one cluster is chosen by considering the maximum center value. After choosing that cluster, shape index (SI) is calculated on the selected cluster and finally that portion is selected where SI value becomes greater than 0.625. From figure 4-(d), it may be seen that cluster-3 contains a maximum center value. After detection of occluded portion, those are removed from the face. Figure 5(a)-(d) shows the occlusion detection and removal of the occluded portion of the face.

Now in the case of eye occlusion by handkerchief and glass, depth related issue does not happen here. Here nose tip is detected based on highest depth. After that according to the distance between nose tip and the lowermost portion of the eye, an approximate region containing the eyes is cropped for removing the occluded eye region. Here the distance between nose tip and starting of eye position from the bottom of the eye and the size of the eye are considered as 6 pixels and as 20

(a) (b) (c) (d) (e)

Fig. 4. (a) Input range image (b) Cluster 1-4

(a) (b) (c) (d)

Fig. 5. (a) Maximum center value cluster (b) SI >0.625 on that cluster portion (c) Occluded portion detection (d) Removal of occluded portion

pixels respectively. The figure 6 and 7 (a)-(d) illustrates the procedure of detection and removal of eye occlusion. Next for occlusion by a hair in not a problem in the selected regions.

2) Rectangular cropping of the face region and Region Division: After smoothing of the raw input 3D face image, here we cropped the face rectangular for considering the most significant region of any particular face. So cropping is based on nose tip detection. Nose tip is detected based on the highest depth of a normalized range image. Figure 8 shows (a) input 3D range image (b) nose tip detection and (c) rectangular cropped image. In some of the cases of occlusion, nose tip cannot be detected based on highest depth. In those cases, occlusion is removed according to the discussion in the previous section followed by nose tip detection.

After rectangular cropping we have divided the resultant cropped image into three regions, where the first region contains the eye portion, the second region contains the nose portion, and the third region contains mouth portion.

C. Feature Extraction

Features are extracted by calculating the local binary patterns of the resultant cropped region image followed by the histogram of gradient calculation in a different region.

1) Local Binary Pattern calculation: In this proposed work, the fundamental operation of LBP [19] is applied on depth image. In general, LBP is applied on 2D images, whereas depth values of range image are used for LBP operation in our work. Here 3×3 mask is employed in a clock-

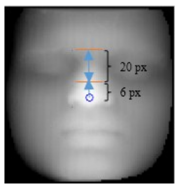
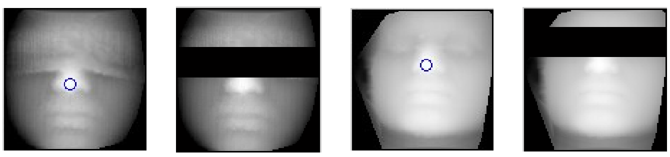


Fig. 6. Length calculation for occlusion detection



(a) (b) (c) (d)

Fig. 7. (a) Eye occlude by handkerchief (b) Removal of occluded portion (c) Eye occlude by handkerchief (d) Removal of occluded portion

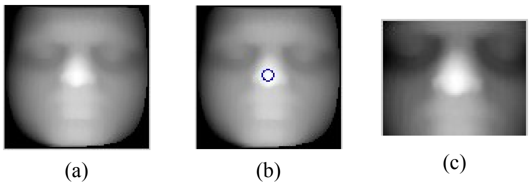


Fig. 8. (a) Input 3D range image (b) Nose tip detection (c) Rectangular cropped image

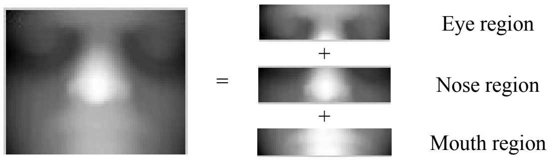


Fig. 9. Region Division of resultant cropped image

-wise direction; mask is shifted along the whole resultant image, then replace the center value of the mask by decimal conversion of corresponding binary LBP code. Figure 10 (a) – (d) illustrate the LBP operation on input range image and followed by region division.

2) Histogram of oriented gradient calculation: After getting the LBP image, a histogram of oriented gradient has been calculated in a different region of LBP matrix. HOG is calculated based on orientation histogram of edge intensity of a picture. HOG feature is introduced by N. Dalal and B. Triggs [20]. The first step of HOG is the computation of gradient values in horizontal and vertical direction. Next, a 1D mask is used for gradient calculation in horizontal and vertical direction, shown in (1) and (2).

$$F_x = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \tag{1}$$

$$F_y = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \tag{2}$$

In step 2, the convolution operation is done on the original image. Let image is I, the convolution operation in x and y-direction is shown in (3) and (4).

$$I_x = I.F_x \tag{3}$$

$$I_y = I.F_y \tag{4}$$

Further, we have calculated the magnitude and orientation of gradient, which is represented in (5) and (6).

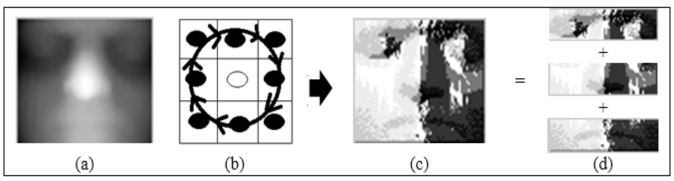


Fig. 10. (a) Input 3D range image (b) 3x3 clockwise LBP mask (c) LBP form (d) Region Division of LBP matrix

$$|G| = \sqrt{I_x^2 + I_y^2} \tag{5}$$

$$\theta = \tan^{-1} \frac{I_y}{I_x} \tag{6}$$

After that, histogram on magnitude matrix have been calculated with the orientation of different region.

D. Classification

In this stage, to classify the resultant HOG feature vector, we have used Euclidean distance based centroid classifier. Minimum distance has been calculated based on mean data of different classes and test data. Here two steps are used for classification; region based local classification followed by decision level global classification. At the first step, the classification process is applied to different regions then the final determination of the identity of the face is made by fusion of classification decisions of the individual region classifiers. The input images are of size 100×100 pixels, which are then cropped to retain only the face region. After cropping 57×71 pixels, each of the cropped face images is horizontally divided into three bands, each of size 19×71 pixels. In our experiment, 2-fold cross validation technique is applied for classification, where total image set is split into two separate disjoint sets, one is treated as training, and another one as a test set.

IV. Experimental Result and Discussion

Experiments are conducted on Frav3D database and our database. Six frontal images including expression variation of Frav3D database and 12 frontal images including expression, occlusion variation of our existing database are used for the experiment. In 2-fold cross validation technique, the whole set of the image is divided into two disjoint sets. Let X is treated as training set and Y is treated as the test set and the vice-versa. In our proposed work, this technique is applied to different regions of the input face image. Table I and II illustrates the recognition accuracy of three different parts of Frav3D database and our proposed database.

A graphical representation of the results shown in Table I and II is presented in figure 11. Five different symbols: diamond, rectangle, triangle, multiplication, star, and the circle is used to represent the experimental result of Training set X and Test set Y, Training set Y, and Test set X and average of previous two experiment of Frav3D and our existing database.

TABLE I. Recognition Accuracy of Three Different Regions on FRAV3D Database

Region	Recognition Rate (%)		
	Experiment 1: Training set: X and Test set: Y	***Experiment 2: Training set: Y and Test set: X***	***Average of experiment 1 and 2***
Eye	88.32	89.4	88.86
Nose	77.4	78.67	78.03
Mouth	69.75	67.23	68.49

TABLE II. RECOGNITION ACCURACY OF THREE DIFFERENT REGIONS ON PROPOSED DATABASE

Region	Recognition Rate (%)		
	Experiment 1: Training set: X and Test set: Y	*Experiment 2: Training set: Y and Test set: X*	*Average of experiment 1 and 2*
Eye	76.4	78.59	77.5
Nose	70.48	72.53	71.5
Mouth	46.96	48.34	47.65

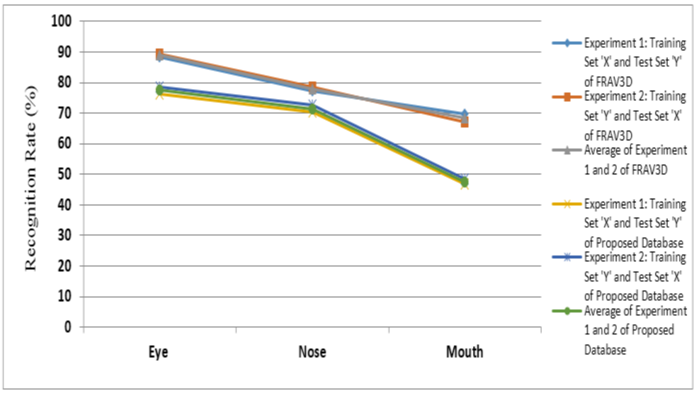


Fig. 11. Graphical representation of recognition accuracy of three regions on two input databases

Now from the above Table I and II, we have calculated overall recognition rate of two different databases using the concept of decision level fusion and score level fusion technique. Table III given below represent recognition rate based on two distinct techniques.

From the above experimental analysis, it can be observed that region-based classification is very useful for recognition and sometimes it gives better accuracy than a simpler approach. It can also be inferred that eye region gives maximum accuracy than other portions of the face. A person can easily be identified based on eye region when other information are unknown to us.

V. COMPARISON WITH OTHER WORKS

In our proposed work, Frav3D and our database are used as input database. Some 3D face recognition techniques are available in the literature, which reported their results on Frav3D database. Here in this section, we have compared our method with other techniques of Frav3D database. Previously Ganguly et al.[11] proposed a new technique for recognition of whole face region on the same database. Authors used curvature based analysis followed by SVD for recognition purpose; further, calculate the recognition accuracy 86.51% using the artificial neural network classifier. Another work with the same database by Belghini et al.[21], proposed a method for recognition of 3D depth face images using the computation of Gaussian-Hermite Moments on each. They have used back propagation neural network for computing the recognition and achieved a recognition accuracy of 89%. Hajati et al.[22] proposed a new Geodesic texture warping(GTW) based technique for recognizing the pose-

TABLE III. RECOGNITION ACCURACY OF THREE DIFFERENT REGIONS ON PROPOSED DATABASE

Database	Eye Region	Nose Region	Mouth Region	Decision level Fusion	Score level Fusion
FRAV3D	88.86	78.035	68.49	88.86	78.46
Proposed Database	77.5	71.5	47.65	77.5	65.55

invariant 3D face. They have achieved 90.3% recognition accuracy using Euclidean distance based classification.

Now, the recognition accuracy of our proposed technique is 88.86%. All other works, mentioned above, are based on entire face area, whereas our proposed method recognizes a given face image only based on a particular region of the face. On that basis, the proposed system is more advantageous in the case of partial occlusion. Table IV given below represents comparison result with other techniques on Frav3D database in ascending order of recognition rate.

TABLE IV. COMPARISON OF RECOGNITION ACCURACY OF DIFFERENT METHOD ON FRAV3D DATABASE

Methods	Recognition rate (%)
Curvature analysis + SVD + ANN (Classification on whole face)[11]	86.51
Proposed Method (Region based classification)	88.86
Gaussian-Hermite Moments computation + Neural network (Classification on the whole face)[21]	89
Geodesic texture wrapping + Euclidean based classification(Classification on whole face)[22]	90.3

VI. CONCLUSION & FUTURE WORK

In this work, a region based classifier is used on the 3D face images. Our proposed system is robust due to the reason that it recognizes a face based on information from some individual local regions like eye, nose, and mouth of the face. Two databases: Frav3D and the 3D face database developed in our laboratory are used in this work. Only frontal faces with occlusion and expression variations along with neutral poses are considered for the system. Limitation of the present technique is that pose variations are not considered here. In future, we will consider pose variation into our system by registration of the face images. On the other side, occlusion of the eye by glass and handkerchief is not detected in an automatic way. Overall it can be concluded that the simple algorithm of this system is computationally efficient and robust for expression and occlusion handling.

Acknowledgment

Authors are thankful for a project supported by DeitY (Letter No.: 12(12)/2012-ESD), MCIT, Govt. of India, at Department of Computer Science and Engineering, Jadavpur University, India for providing the necessary infrastructure for this work.

References

[1] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," ACM Computing Surveys, vol. 35, issue 4, December 2003.

[2] A. F. Abate, M. Nappi, D. Riccio, and G.Sabatino, "2D and 3D face recognition: A survey," Image Information Control, vol. 28, issue 14, pp. 1885-1906, October 2007.

[3] T. Faltemier, K. W. Bowyer, and P. Flynn. "3d face recognition with region committee voting," Intl. Symp. on 3D Data Processing, Visualization, and Transmission, 2006.

[4] J. Cook, M. Cox, V. Chandran, and S. Sridharan. "Robust 3d face recognition from expression categorization," Intl. Conf. on Biometrics, pp. 271–280, 2007.

[5] O. Gervei, A. Ayatollahi, and N. Gervei, " 3D Face Recognition Using Modified PCA Methods," World Academy of Science, Engineering & Technology; vol. 4, issue 39, pp. 264, March 2010.

[6] T. Russ, C. Boehnen, and T. Peters, "3D face recognition using 3D alignment for PCA," Proceedings of the CVPR, pp. 1391–1398, 2006.

[7] A. B. Moreno, A. Sanchez, J. F. Velez, and F. J. Diaz, "Face recognition using 3D local geometrical features: PCA vs SVM," Proceedings of the ISPA, pp. 185–190, 2005.

[8] T. Heseltine, N. Pears, and J. Austin, "Three-dimensional face recognition: A Fisher surface approach," Proceedings of the ICIAR, pp. 684–691, 2008.

[9] C. B. Abdelkader, and P. A. Griffin, "Comparing and combining depth and texture cues for face recognition," Image and Vision Computing, vol. 23, pp. 339–352, 2005.

[10] C. Hesher, A. Srivastava, and G. Erlebacher, "A novel technique for face recognition using range imaging," Proceedings of the Seventh International Symposium on Signal Processing and Its Applications, pp. 201–204, 2003.

[11] S. Ganguly, D. Bhattacharjee, and M. Nasipuri, "3d face recognition from range images based on curvature analysis," ICTACT Journal on image and video processing. vol. 4, issue 03, pp 748-753, 2014.

[12] P. Bagchi, D. Bhattacharjee and M. Nasipuri, "3D Face Recognition using Surface Normals," TENCON 2015 - 2015 IEEE Region 10 Conference, pp. 1-6, 2015.

[13] Y. Lee, H. Song, U. Yang, H. Shin, and K. Shon, "Local Feature Based 3D Face Recognition," 5th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA), vol. 3546, pp. 909-918, 2005.

[14] A. Chouchane, M. Belahcene, A. Ouamane, and S. Bourannanae, "3D face recognition based on histogram of local descriptors," 4th International Conference on Image Processing Theory, Tools and Applications (IPTA), pp. 1-5, October 2014.

[15] FRAV3D Database, http://www.frav.es/databases/FRAV3d/, 2004.

[16] H. Yagou, Y. Ohtake, and A. G. Belyaev , "Mesh Smoothing via Mean and Median Filtering Applied to Face Normals," Geometric Modeling and Processing, Geometric Modeling and Processing, pp. 124–131, 2002.

[17] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms," Springer Publishers, New York, 1981.

[18] F. Kovacs, C. Legany, and A. Babos, "Cluster Validity Measurement Techniques," 5th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases, pp. 388-393, 2006.

[19] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971-987, 2002.

[20] N. Dalal, and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 886–893, June 2005.

[21] N. Belghini, A. Zarghili, and J. Kharroubi, " 3D Face Recognition using Gaussian Hermite Moments," Special Issue of International Journal of Computer Applications on Software Engineering, Databases and Expert Systems – SEDEXS, September 2012.

[22] F. Hajati, and Y. Gao, "Pose-Invariant 2.5D Face Recognition using Geodesic Texture Warping," 11th. International Conference on Control, Automation, Robotics and Vision Singapore, pp. 1837–1841, December 2010.