

A New Algorithm for 3D Object Representation and its Application for Human Face Verification

Pranali Dhane, Ankita Jain, Krishnan K. Kutty

Center for Research in Engineering Sciences & Technology (CREST)

KPIT Cummins Infosystems Ltd.

Pune, India

Abstract—There are many papers published on the use of 2D images for object representation. The fundamental issue with these approaches lies in the fact that a conventional image maps 3D real world objects into a 2D plane. This causes loss of 3D information - which means loss of 3D features as well for further analysis. There have been attempts using 3D and 2.5D scanners in order to access range data from objects for further analysis. However, these systems require special scanners and associated hardware, and many a times the complete information that is obtained from such systems is not required for further analysis. In this paper, we propose a simple system that projects a light beam over an object of interest and using a simple optical camera a video is recorded. This enables us to have 2D data as acquired from the camera along with the 3D data that is derived out of the deviations in the scan pattern because of the non-planarity of the object of interest. We have limited our research to using 3D features only for object representation and extended the same logic for face verification. The results that we have obtained show that in spite of using only 3D features, the system is able to verify faces with varied expressions correctly on 75% of the test data. The system is robust to slight tilt in head angle and normal human expressions like smile, frown, grin etc.

Keywords—*Depth map; Fourier descriptor; Centroid signature; Mahalanobis distance; Face Verification*

I. INTRODUCTION

2D object representation is a well researched subject and has found many applications. The challenge with conventional image based shapes is that all 3D real world objects are mapped into a 2D plane - thereby losing out on the depth/range information. The 3D view is an interesting topic of current research.

3D features provide the depth information which does not change much with translation and changes in illumination [3]. Most matching algorithms employ either the 2D or 3D data but not both simultaneously. However, most existing range sensors provide 2D colour intensity information and 3D shape information in the respective grid; thus, one could take the advantage of both to accomplish greater performance and robustness [4]. Multiple methods have been suggested and explored in literature for estimating depths from stereo images as well as from monocular images. However, there is a limitation with respect to exact range/depth estimation. Moreover, these techniques work best for macro objects - and seldom fail to explain the intricate depth variations on an object. The stereo vision based approach is well defined in [2]. There are different methods in literature for generating 3D

model of an object which use combination of depth map and texture information [5].

In this paper, we propose a simple yet robust technique for estimating relative 3D depths of all points in an object. This is realized by projecting a light beam over an object of interest and using a simple optical camera to capture the complete video while the scan is on. The proposed method is a simple, low cost, vision based system for obtaining range information for an object of interest and further probes into using this data for object verification.

The 3D shape of an object doesn't change much with its scaling or illumination changes when seen by an optical camera. The shape of an object using a local shape descriptor is a unique feature using which we could distinguish between various objects.

In our proposed method, we use Fourier Descriptors (FD) to represent shapes after the scan is complete. The first few low frequency terms in Fourier descriptors capture global shape features, while higher frequency terms capture finer features of the shape under consideration. The other advantage of FD is that it can be easily normalized to make scale, translation and rotation invariant [1]. We have also tried to expand our research for face verification using the same method by using only the range information, which will be explained in the later part in this paper. B. H. Shekhar et.al. used significant points based SULD descriptor for 3D face recognition. The significant points are detected using Hessian based detector [6].

The rest of the paper is organized as follows: Section II gives a brief of the proposed method and describes how the method performs for simple symmetric 3D objects. Section III describes how we have used the scan line data in order to extract range information for the object of interest. Section IV describes representing objects using a depth map. IN Section V, we describe the use of Fourier descriptors for extracting 3D shape information and object representation. Section VI extends the proposed approach for face verification. Section VII details our simple experimental set up while Section VIII shows the results for face verification. Section IX details the conclusion and scope for future work. In this paper we have tried to use the 3D features and FDs for face verification. There are various methods of face verification using 2D features.

II. PROPOSED METHOD

In this paper, we project a colour light beam over an object of interest and a simple optical camera set up is used to capture

the video while the scan is on. This colored beam gets distorted due to the variations in the object depth with respect to the camera plane. The deviation of the scan line on the object with respect to the background line (undistorted) is used to calculate the distortion at every point on the object of interest. This distortion in the line, as seen by the camera, provides a good estimate of the relative depth of the object for each point on the object surface. The distortion of the projected beam as seen by the camera is shown in Fig.1 (a).

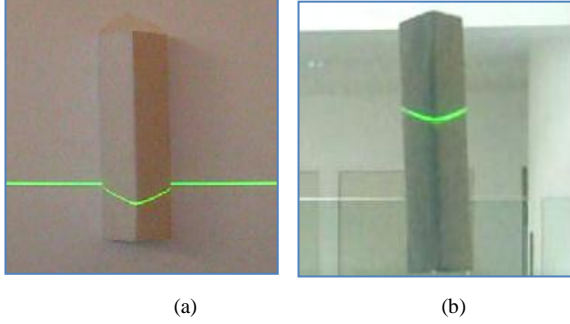


Figure 1. Distorted line due to object depth (a) With Background (b) Without background

However, it is possible that there is no background available for calculating the distortion. Fig. 1(b) shows an example of one such case where the object of interest was leaning against a background made of solid glass – because of which there was no projection of the beam on the glass. In order to cater to these scenarios, in the proposed method, we developed a method that eliminated the need of having a background. The algorithm for calculating the relative depth with and without background is discussed in Section III.

III. DEPTH CALULATION

A. With Background

Consider any two points on the background line as reference. The equation of a line through two given points (x_1, y_1) and (x_2, y_2) is given in (1).

$$(y - y_1) = \frac{(y_2 - y_1)}{(x_2 - x_1)}(x - x_1) \quad (1)$$

We convert this equation into the format of the standard line equation given in (2).

$$ax + by = c \quad (2)$$

Where a,b,c are given in (3),(4) and (5) respectively,

$$a = (y_2 - y_1) \quad (3)$$

$$b = -(x_2 - x_1) \quad (4)$$

$$c = -((x_2 - x_1) * y_1) - ((y_2 - y_1) * x_1) \quad (5)$$

The distortion of any point (u,v) on the object with respect to the background is calculated by using the formula given in (6).

$$d = \frac{|au + bv + c|}{\sqrt{a^2 + b^2}} \quad (6)$$

Equation (6) denotes that the distortion is calculated by finding the normal distance of every point (u,v) from the

background line. The depth map for Fig.1 (a) that has been calculated using (6) is as shown in Fig.2 (a).

B. Without Background

In this case, since there is no line for a reference, consider any point on the object as a reference. Let consider (X_r, Y_r) a point on an object as a reference point and (X_i, Y_i) be any other point on the object. The distortion at a point (X_i, Y_i) is given in (7).

$$\text{Distortion} = Y_i - Y_r \quad (7)$$

Using (7) find distortion for every point on the object. All these distortion values are stored in a matrix format, which consist of both positive and negative values. In order to eliminate negative values in this matrix, find the minimum value in this matrix and add that value to every entry of the matrix. This matrix with all positive values is indicative of the relative depth values for all points on the surface of the object. The depth map obtained for Fig.1 (b) is as shown in Fig.2 (b).

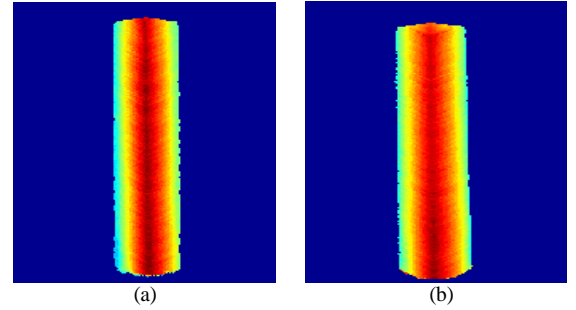
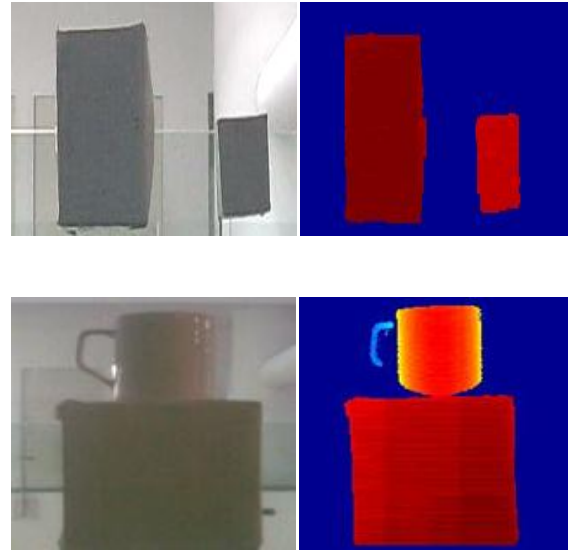


Figure 2. (a) and (b) are depth maps of Fig. 1(a) and (b) respectively

IV. REPRESENTING OBJECTS USING DEPTHPMAP

After The depth maps, as obtained by the proposed method, of some symmetrical and smooth objects are shown in Fig. 3.



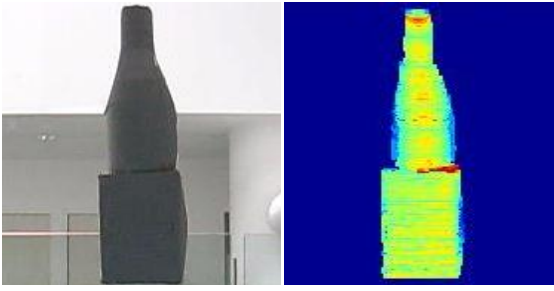


Figure 3. Depth map of Different Objects

The performance of the proposed algorithm on objects that have subtle variation on the object surface is shown in Fig 4. This represents the depth map as estimated by the proposed method for a toy-mask.

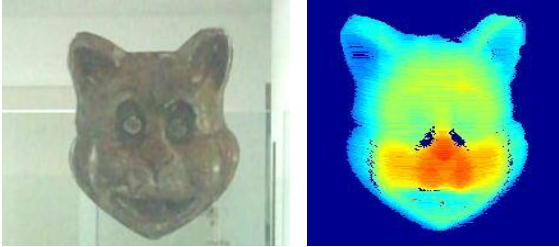


Figure 4. Depth map of a mask

V. OBJECT VERIFICATION USING 3D FEATURES

In order to perform object verification, it is imperative to generate features that are robust and invariant to scaling, translation and rotation to the extent possible. Shape is a very prominent feature to distinguish between different objects. There are many shape representation and retrieval techniques available in literature. In this paper, Fourier descriptors (FD) were used for representing shapes from the 3D profiles.

Shape retrieval using FD has many variants that use one of the following: centroid distance, complex coordinates, curvature function, and cumulative angles. Centroid distance is a shape representation between contour based representation and region based representation. Shape retrieval using FDs derived from centroid distance signature is significantly better than that using FDs derived from the other three signatures [1]. We have therefore used the centroid distance signature for computing the FD.

In order to calculate the Fourier descriptor, the first step is to convert the image into binary form. This is followed by a denoising process that eliminates isolated pixels or small segments and provides only the boundary of the object of interest.

To calculate the centroid distance signature, assume any point $(x(t), y(t))$ on the traced boundary, where $t=1, 2, \dots, N$. N denotes the total number of points on the perimeter of the traced contour. Centroid distance is the distance between every point on boundary and the centroid of that shape. Centroid distance is given by (8)

$$r(t) = \sqrt{(x(t) - xc)^2 + (y(t) - yc)^2} \quad (8)$$

Where $r(t)$ is the centroid distance, (xc, yc) is the centroid of shape which is given by (9)

$$xc = \frac{\sum_{t=1}^N x(t)}{N}, \quad yc = \frac{\sum_{t=1}^N y(t)}{N} \quad (9)$$

Thereafter, the contour is sampled with a fixed number of points and then the Fourier transform is applied on the sampled points by using (10).

$$u(n) = \frac{\sum_{t=1}^L r(t) e^{\frac{-j2\pi n t}{L}}}{L} \quad (10)$$

Where L is number of sampled points, n varies from 1, 2 ... L and coefficients $u(n)$ are FDs of the shape, denoted by FD_L [7].

FD using centroid distance signature is invariant to translation and rotation. To normalise FD using centroid distance, only half of the coefficients of FD are considered. To make it scale invariant, the magnitude values of the first half coefficients are divided with the DC component [1].

$$f = \frac{|FD2|}{|FD1|}, \frac{|FD3|}{|FD1|}, \dots, \frac{|FD(\frac{L}{2})|}{|FD1|}$$

Fourier descriptors of some objects are shown in Fig. 5:

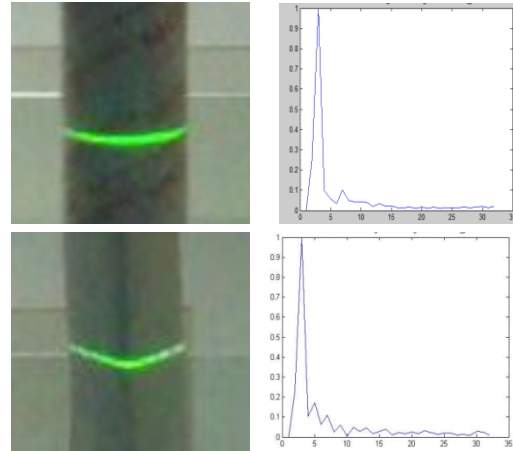
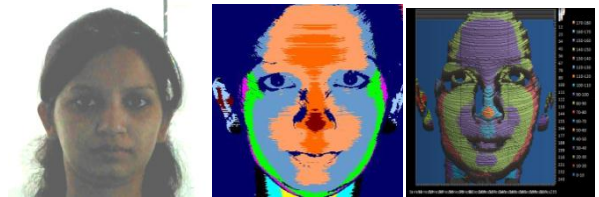


Figure 5. Centroid signature of objects

As seen from Fig.5, Fourier descriptor is different for objects of different shape.

VI. APPLICATION TO FACE VERIFICATION

The proposed method was extended to face verification by exploiting the 3D features that have been extracted. For sample faces without any facial expression and head tilt, the associated depth maps are as shown in Fig. 6.



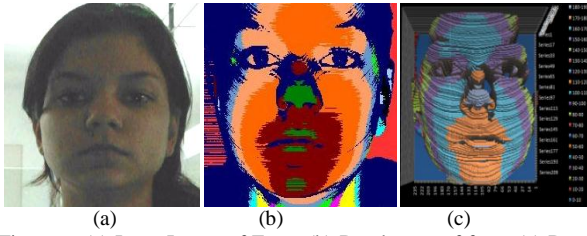


Figure 6. (a) Input Image of Faces (b) Depth map of faces (c) Depth map with different range of colour map

Fig.6 (a) shows the input images of faces which are scanned by projecting a light beam; and Fig.6 (b) shows its corresponding depth map. Fig.6 (c) represents the same depth map with a modified range for colour-map. The depth map changes with various head poses and facial expressions. It is also evident, as seen from Fig. 6 (c), that the nose tip is the point having highest depth value and also it is a distinct point in the depth map.

It was also noticed based on our experiments that the nose tip was insensitive to subtle changes in facial expression. For ideal condition, the nose tip is a point having highest value in Z – direction. However, the highest point may change with change in head pose. If the subject is looking upwards, it is possible that a point on chin might be a highest point. Likewise, if the subject is looking downward a point on the forehead may become the highest point. Therefore, in order to locate the nose tip exactly, we find the location of the maximum depth value in each row as shown in Fig 7.



Figure 7. Maximum depth value points along each row

Thereafter, we select one of the highest peaks in each row and consider corresponding row profile. Let cp denote the column index of the peak. For each row profile, the metric sg is computed by (11).

$$sg = \sum_{c \in R} |Zcp - Zc| \quad (11)$$

Where, R denotes the columns around cp . The metric sg measures the variations of the row profile with respect to each peak [9]. The largest value of sg denotes the row having the highest variation along Z-axis, the position of the peak for this row is selected as a nose tip.

The results for extraction of nose tip for various head poses are as shown in Fig. 8.



Figure 8. Determination of nose tip for different head poses

It was also noticed that the nose tip is very distinct for different people. The Fourier descriptor of the nose tip frame for some distinct faces is shown in Fig. 9.

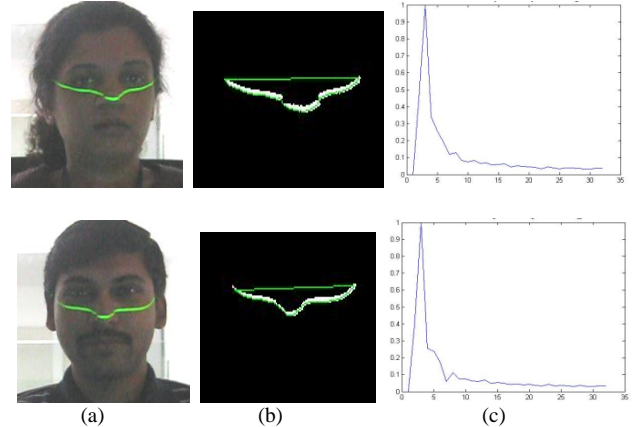


Figure 9. (a) Input nose tip frame (b) Boundary trace of nose tip frame (c) Centroid distance signature of nose tip frame.

As shown in Fig.9, Fourier descriptor for the nose tip frame is different for different person. Thus, we take advantage of the Fourier descriptor, in order to perform a face verification based on range features. The steps included in proposed face verification method are as shown in Fig.10. In the proposed approach, we have considered 40 scan lines to either side of the line corresponding to the nose tip for verification purposes.

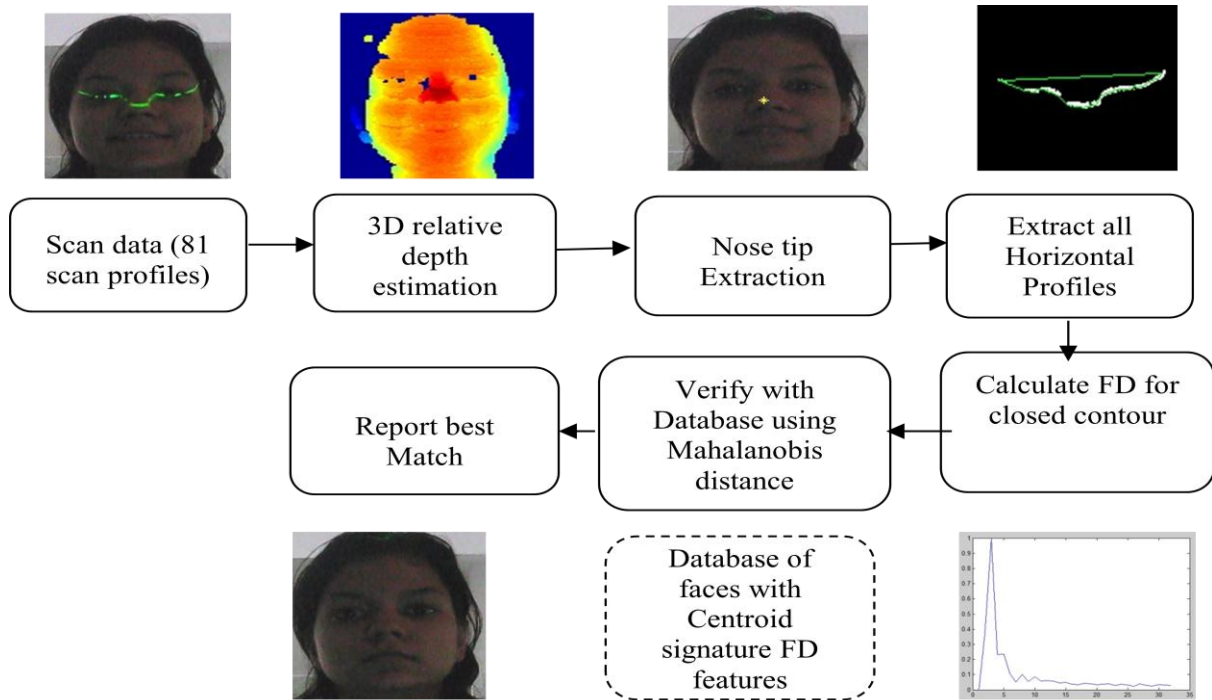


Figure 10. Proposed algorithm for face verification

VII. EXPERIMENTAL SETUP

The set up for the proposed method is as shown in Fig.11.

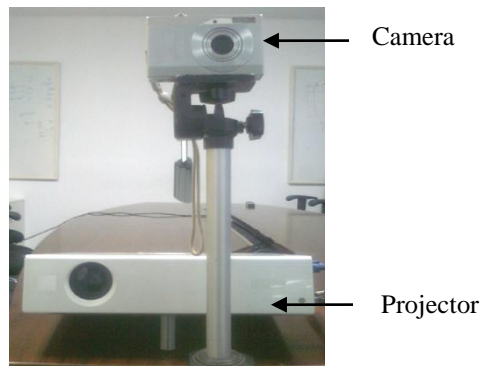


Figure 11. Setup for obtaining 3D view

The light source is kept in such a way that it scans the whole object with a colored beam of light. The color of the light beam should be distinct so that it can be detected very easily on the object. The camera is fixed above the light source such that it focuses on the object.

VIII. RESULTS AND DISCUSSION

For experimentation purposes, we have considered a database (3D) of 50 people. Some sample images from the database are as shown in Fig.12.

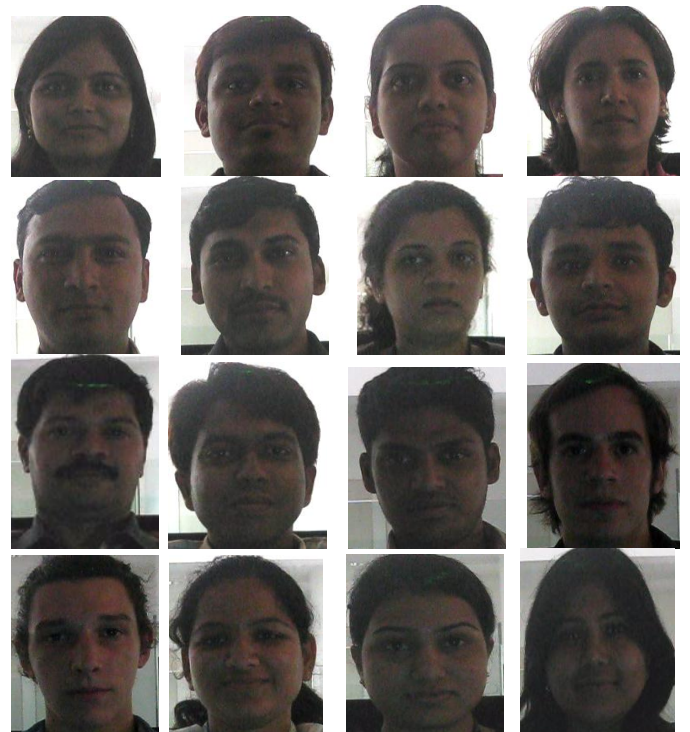


Figure 12. Sample faces from database

In order to maintain consistency in the database, we have considered centroid signature FD of all features for everyone. These FDs of features are different for each person. The pattern of these FDs is similar for various individuals but not

same. In addition, we have considered several features for better verification.

For comparison of FDs, we have used the Mahalanobis distance metric. Mahalanobis distance is based on correlations between variables by which different data sets can be identified and analysed. It is a useful way of determining similarity between an unknown sample set and a known one. It is different from Euclidean distance as it takes into account the correlations of the data set. The formula for calculation of Mahalanobis distance is:

$$d_M = ((\bar{X} - \bar{Y})C^{-1}(\bar{X} - \bar{Y})^T)^{1/2} \quad (12)$$

Where, d_M is Mahalanobis distance of each observation in Y from the reference sample in matrix X . \bar{X} and \bar{Y} are mean of X and Y respectively. C is the covariance of reference matrix X . For calculating Mahalanobis distance X and Y must have same number of columns [8].

For verification we have taken different facial expressions for each face in database. In all, we have tested the proposed method with 112 videos. The FD of the given test subject is compared with each face in the database using Mahalanobis distance. The face that has the minimum Mahalanobis distance is considered as the best match for sample test face. Some sample faces with some facial expressions are as shown in Fig. 13.



Figure 13. Faces with different expressions

Among tested videos 60% faces matched accurately using proposed method. It was noticed that for some cases, the correct match was the second best match. This was because of missing data while the scan was on or because of considerable changes in the head pose while testing. In order to correct for these scenarios, we further consider the top 4 matches from the data and recalculate the Mahalanobis distances for the top half of the face only. This is done because the features above the nose tip are less affected by changes in facial expression. By incorporating this logic, the accuracy of the proposed method increases to 75%.

IX. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a method that uses a simple set up to provide detailed information of the depth map for all points on an object of interest. Fourier descriptors are used for extracting features from the depth map. The proposed method was extended for human face verification using depth map. In order to compare test subjects with that in the database, we have used the Mahalanobis distance metric. With a limited database, our proposed method works well and provides verification accuracy of 75%. It is noteworthy that nowhere in the proposed method, we have used any pixel related information that is inherently captured by the camera when the scan was on. The accuracy numbers can be further improved by complimenting the existing feature vector with features that can be extracted using the captured 2D image.

ACKNOWLEDGMENT

The authors would sincerely like to thank all the employees of KPIT Cummins Infosystems Ltd. for providing an excellent platform, necessary equipment and valuable guidance for conducting research.

REFERENCES

- [1] Dengsheng Zhang and Guojun Lu, "A Comparative Study on Shape Retrieval Using Fourier Descriptors with Different Shape Signatures", *Journal of Visual Communication and Image Representation*, pp.- 41-60, 2003.
- [2] Daniel Scharstein, Richard Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms", *International Journal of Computer Vision*, Volume 47 Issue 1-3, pp. 7 - 42, April-June 2002.
- [3] Xiaoguang Lu and Anil K. Jain, "Automatic Feature Extraction for Multiview 3D Face Recognition", *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, Southsampton, UK, pp.- 585-590, April 2006.
- [4] Chris Boehnen, Trina Russ, "A Fast Multi-Modal Approach to Facial Feature Detection", *Proceedings of the Seventh IEEE Workshops on Application of Computer Vision (WACV/MOTION'05) - Volume 1*, pp. 135-142, 2005
- [5] Philipp Fechteler, Peter Eisert and J'urgen Rurainsky, "Fast And High Resolution 3d Face Scanning" *International Conference on Image Processing - ICIP*, pp. 81-84, 2007.
- [6] B H Shekar, N Harivinod, M Sharmila Kumari, K Raghurama Holla, "3D Face Recognition using Significant Point based SULD Descriptor", *IEEE-International Conference on Recent Trends in Information Technology, ICRTIT, MIT, Anna University, Chennai*, pp- 981-986, June 3-5, 2011.
- [7] Dengsheng Zhang and Guojun Lu, "An Integrated Approach to Shape Based Image Retrieval", *The 5th Asian Conference on Computer Vision*, Melbourne, Australia, pp.-1-6, 23-25 January 2002.
- [8] <http://people.revoledu.com/kardi/tutorial/Similarity/MahalanobisDistance.html>
- [9] Xiaoguang Lu and Anil K. Jain, "Multimodal Facial Feature Extraction for Automatic 3D Face Recognition", *Technical Report MSU-CSE, Computer Science and Engineering, Michigan State University*, pp.- 05-22, August 2005.