

# 3D Face Recognition Using Wiener Filter and DFT Based On Optimized Directional Faces

Shilpa S. Nair, Naveen S., Moni R.S

**Abstract**— Traditional 2D face recognition methods based on intensity or color images, face challenges in dealing with pose variations or illumination changes. The face recognition based on combination of 3D shape information and 2D intensity/color information is a novel approach, which provides an opportunity to improve the face recognition performance. This paper proposes an efficient multimodal face recognition method by combining the textural as well as depth features, extracted from directional faces of input image. To overcome problems occurred due to low quality image, pre-processing is done before extracting features from the image. The directional faces captured using Local Polynomial Approximation (LPA) filters are adaptively optimized. The modified LBP (mLBP) is used for the feature extraction from these directional faces. The spectral transformation of the concatenated block histogram of mLBP feature image acts as the robust face descriptor. Discrete Fourier Transform (DFT) is used as the transformation tool. The fusion of both modalities is performed at score level. The experimental results shows that the proposed method gives better performance than single modality.

**Index Terms**— DFT, MLBP, multimodal, ODF, Wiener filter.

## I. INTRODUCTION

3D face recognition is one of the fast growing research area in pattern recognition and computer vision. The main challenges in 2D face recognition are illumination and pose variations which affect the accuracy of the recognition system. To improve the performance some additional information needs to be considered along with the texture information. Such types of recognition system where multiple features are used for identification purpose are called multi modal face recognition systems. Depth is an important information that is invariant to physical changes and provides recognition in different view angles. If the depth information is also made available along with texture information, fusing these two features promise a better performance than both used separately. Textural data is the 2D (x, y) representation of intensity values while depth image or range images contains at most one depth value (z direction) for every point in the (x, y) plane. In this paper, the objective is to develop an accurate and robust multimodal face recognition system, by fusing the depth and intensity information effectively. Intensity information is more reliable than depth information in the context of expression

variations while depth information is more robust than intensity information in the context of pose variations. Thus, their combination is helpful in improving the recognition performance. To enhance the input image, preprocessing is carried out using Wiener filter. The texture and depth information are extracted from directional faces at multiple levels [1]. Direction is a low level feature that enhances edges and boundaries, can be considered for face recognition because a large amount of information in face is aligned in specific directions. For capturing the directional features, derivatives of directional filters designed using Local Polynomial Approximation (LPA) is used. The features are extracted at local and global level using mLBP from the directional faces. The LPA-mLBP map is then divided into blocks and histogram is evaluated in each block and concatenated. The spectral representation of the histogram is obtained by applying Discrete Fourier transform (DFT). The spectral representation, thus obtained serves as the face descriptor. Finally PCA is applied for dimensionality reduction of the descriptor and classification is done.

Most of the early works on face recognition concentrates on global approach. In this method the whole image is considered and its features are extracted at global level. The significant methods in this category are Eigen face method and Fisher face techniques which are based on the Principal Component Analysis (PCA) [2] and Linear Discriminant Analysis (LDA) [3] respectively. Principal Component Analysis proposed by Matthew A. Turk and et.al encodes the variations between face images characterized by a set of Eigen vectors (Face space). The weights of the new face image are calculated by projecting the same into the face space. The main drawback of this method is that, it is highly affected by illumination and pose changes. C. Podilchuk et.al introduced DCT based feature vector extraction [4] as the DCT coefficients reflect the important frequencies that are present in the face image. Ahonen et al. applied LBP to face recognition [5] by partitioning the face image into non-overlapping rectangular blocks. The operator assigns a label to every pixel of an image by thresholding the 3x3-neighborhood of each pixel with the center pixel value and considering the result as a binary number. The accuracy of 2D recognition system is influenced by illumination and pose changes. To overcome these drawbacks, the 3D information which is available from face image is combined along with the 2D information. A comparative study on combining depth and texture information is carried out by Ben Abdelkader et al [6] and found that a significant improvement in face of the recognition is achieved by multimodality. As by [7] Antonio Rama et.al presents a novel face recognition approach which uses only partial information in the recognition stage. The algorithm is based on an extension of the classical PCA called as P2CA, and it implements a mixed 2D-3D method.

Manuscript published on 30 August 2015.

\* Correspondence Author (s)

**Shilpa S. Nair**, Department of Electronics and Communication, LBS Institute of Technology for women, Thiruvananthapuram, India.

**Naveen S.**, Department of Electronics and Communication, Kerala LBS Institute of Technology for women, Thiruvananthapuram, India.

**Dr R.S Moni**, Professor, Department of ECE, Marian Engineering College, Trivandrum, Kerala, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

## II. PROPOSED METHOD

The detailed explanation of proposed algorithm is explained in this section. The database considered is FRAV3D [8]. It is a multimodal database which contains texture information (2D image), face model in range data format (2.5D image) and VRML file (3D image). The texture information is available as RGB image of size 400x400 while 3D data in form of point cloud as a matrix array of  $M \times 3$ . The value of  $M$  denotes the number of points in the 3D space. The point cloud data in 3D space is projected to an X-Y grid to get 2.5D (2.5D image is the depth map itself) image using standard projection formula. This depth data will be having the pixel value as the Z- Coordinate of Point Cloud data. The first step is preprocessing, which aims an improvement in the input image data by suppressing unwanted distortions or for enhancing some image features which is important for further processing. Then the directional faces of the input face are extracted by convolving it with LPA filter [9]. The directional faces so obtained at multiple scales are adaptively optimized using Intersection of Confidence Interval (ICI)

[10]. A modified form of Local Binary Pattern (LBP) is used as the feature extraction tool which aids the feature extraction at both global and local level. The spectral transformation of the feature vector is performed and is fed to the classifier. The fusion of both modalities is performed in score level. Fig.1 shows the block diagram of the proposed method.

### A. Preprocessing

The main objective of preprocessing stage is to enhance the input image prior to computational processing. Preprocessing of images commonly involves removal of low-frequency background noise, normalizing the intensity of the individual particles images, removing reflections, and masking portions of images. Wiener filtering is one of the optimal ways of filtering off the noisy components so as to give the best reconstruction of the original signal. Wiener filter produces an estimate of desired or target random process by linear time-invariant filtering of an observed noisy process,

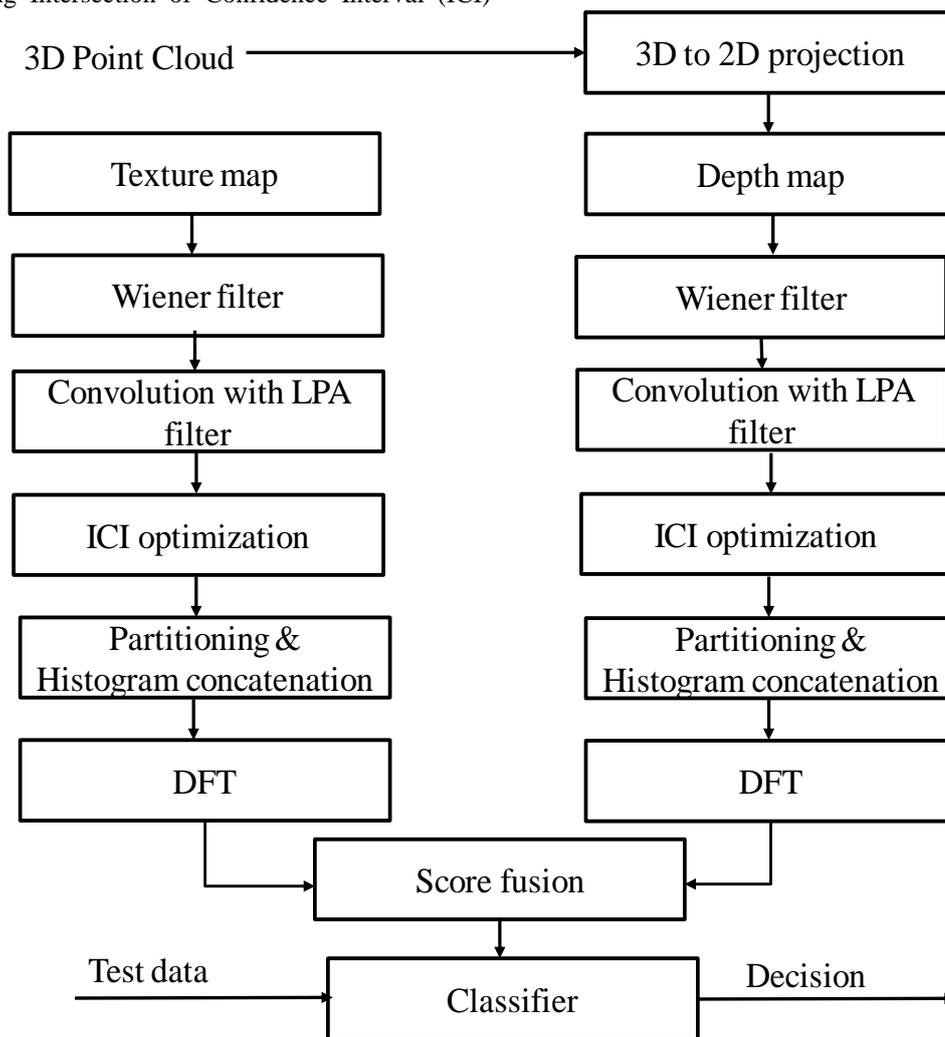


Figure 1. Block diagram of the proposed system

assuming known stationary signal and noise spectra, and additive noise. In short Wiener filter produces an estimate of the signal of interest by filtering out noise from it. Wiener filter operation in 2D does the adaptive noise removal filtering. It acts like a low pass filter for an intensity image that has been degraded by constant power additive noise by using a pixel-wise adaptive Wiener method. It is based on statistics

estimated from a local neighborhood of each pixel, which can also estimate the additive noise power before doing the filtering.

### B. LPA filters and optimized directional faces

Face contains information that are aligned in varying directions. For example, the features like eye brows, eyes are aligned in horizontal direction while nose or face contour are aligned in vertical direction. In order to capture these specialties of the faces and to incorporate them for the recognition purpose, directional filters designed using LPA is used. LPA is a polynomial fitting technique generally used in designing of directional filters at multiple scales. Edges and boundaries are those regions where there is a sharp transition of intensity values, which can be captured by using derivatives of directional filters. The first order LPA derivative filter ( $g_{s,\theta}$ ) for direction  $\theta$  and scale  $s$  is defined as follows:

$$g_{s,\theta}^1(x-v) = -w_{s,\theta}(x-v) \Phi^T(x-v) \Phi_s^{-1} \Phi^1(0) \quad (1)$$

$$\Phi_s = \Sigma w_{s,\theta}(x-v) \Phi(x-v) \Phi^T(x-v) \quad (2)$$

where  $v \in R^2$  is a domain of coordinates,  $x \in R^2$  is the center of the LPA,  $w_{s,\theta}$  is a window function used for a fitting in the neighborhood of the center  $x$ , the scale  $s$  determines the size of the neighborhood, and  $\Phi$  is the vector polynomial. For a specified scale, the LPA kernels are obtained by rotating the kernels in required angles. The directional faces are generated by convolving the input face image with directional filter kernels. For the face image  $F(x)$ , the directional face ( $Ds_{\theta}$ ) is defined as

$$Ds_{\theta}(x) = F(x) * g_{s,\theta}^1(x) \quad (3)$$

The number of directional faces thus obtained is equal to the product of number of directions and scales used in the filter design. Scale  $s$  is the support size of filter, which determines the number of pixels selected to estimate the directional derivative at a point. In this work, number of scales and directions is set to be 4 in order to reduce the complexity. Hence here 16 directional faces are generated. But it is not easy to deal with this large number of directional faces as it results higher complexity in feature extraction. As a solution, the directional faces are adaptively optimized using Intersection of Confidence Interval (ICI) rule. The ICI adaptively selects the pixel in images for each scale such that mean square error (MMSE) is minimized. The directional faces obtained after optimization is called optimized Directional Faces (ODFs). Direction of filter is represented as  $\theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$  and corresponding ODFs as  $\{O_{\theta_j}\}_{j=1}^4$ . Fig.2 shows the 16 directional faces and 4 Optimized Directional Faces for 4 directions and 4 scales. The textural and depth information are extracted from these directional faces simultaneously.

### C. Feature extraction

Feature extraction which is also called representation, is the crucial step that influence the accuracy of recognition system. Feature extraction deals with the extraction of specific features that efficiently represents the input face. Because of its computational simplicity and effectiveness, LBP operator has become a popular approach in recognition applications. LBP operator generates a series of binary codes by thresholding the neighborhood of each pixel. For keeping the low dimensionality of the descriptor, a modified form of

LBP called mLBP is used for a window of suitable size. For a 3x3 window it is defined as:

$$MLBP(t_c) = \sum_0^3 f(t_i - t_{i+4}) 2^i + f(t_c - \bar{t}) 2^4 \quad (4)$$

where  $f(t)$  denotes unit step function,  $t_c$  is the intensity value of the center pixel,  $t_i$  are the neighbors around the center pixel. mLBP operator is applied to each ODF separately and the extracted features are concatenated. Initially the mLBP map is divided into blocks and then histogram of each blocks is calculated and concatenated. Fig.3 shows the 3 level partitioning and histogram calculation of an ODF. The number of blocks depends on the level of partitioning. For a  $k^{\text{th}}$  level partitioning, the image is divided into  $4^k$  ( $k=0,1..K$ ) non-overlapping blocks. At level  $k=0$  the image is considered as a whole and global features are obtained. Histogram obtained from ODF  $O_{\theta_j}$  at each level  $k$  is concatenated together to get  $h_{jk}$ . Then the histograms for different levels (or  $k$ ) are concatenated to get the histogram of that ODF represented as  $H_{\theta_j}$ . The same procedure is repeated for all the four ODFs and the feature vectors are combined to get the final face descriptor represented as  $H$ .

$$H = [H_{\theta_1} H_{\theta_2} H_{\theta_3} H_{\theta_4}]$$

Since the face contains redundant information, removing them adds the performance of the system. Spectral transformation is one of the simplest way to achieve this. Discrete Fourier Transform (DFT) is used as the spectral transformation tool, using equation (5)

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi nk/N} \quad (5)$$

The Fourier spectrum is a plot of energy against spatial frequencies, where spatial frequencies relate to spatial relations of intensity in image. The better energy compaction of DFT and its rotation invariance property adds the accuracy rate. By spectral transformation, the redundancy in data vectors can be reduced considerably as well as it enables the better representation of the feature vector. The Fourier transform converts the input vector into its complex valued output spectrum consists of magnitude and phase. If we want to reconstruct the complex spectrum back into its original spatial domain, both the phase and magnitude needs to be preserved. Most of the information is available in the magnitude of the spectrum and there is no need of a retransformation into the spatial domain in any of the further stages. Therefore, only the magnitude of the complex spectrum is preserved. The same procedure is followed for both texture and depth data simultaneously to get their spectral transformation which acts as the face descriptor. Even though the transformation reduces the redundancy, the multilevel directional, texture and depth feature extraction results in higher dimensionality. To compensate this higher dimensionality of the feature vector and to reduce the complexity, PCA [2] technique is applied.

### D. Dimensionality reduction

The higher level partitioning of the mLBP map leads to the higher dimensionality of the face descriptor. For a 4 level ( $k=0,1,2,3$ ) partitioning of ODF having mLBP of dimension 32, results in a face descriptor of dimensionality  $4 * 32 * \sum_{l=0}^3 4^l = 10880$ .



To reduce the complexity due to higher dimensionality of the feature vector Principal Component Analysis (PCA) [2] based technique is applied. The idea behind PCA is that a high dimensional feature vector can be represented by a set of Eigen vectors that characterizes the variations in face mage. Let  $F_i$  be the concatenated histogram of the  $i^{th}$  person, it is grouped as an  $M \times N$  matrix  $F = [F_1 F_2 \dots F_N]$ , where  $N$  is the number of face samples under consideration. Mean vector is calculated as follows

$$F_m = \frac{1}{N} \sum_{i=1}^N F_i \quad (6)$$

Standard deviation will be calculated

$$F_{SD} = \sum_{i=1}^N (F_i - F_m) \quad (7)$$

Covariance matrix is calculated

$$F_{COV} = F_{SD} * F_{SD}^T \quad (8)$$

This is a matrix of size  $M \times M$ , which is of very large dimension. Also it gives  $M$  Eigen values and  $M$  Eigen vectors which are very large in number to process. To reduce high dimensionality the covariance matrix is changed as,

$$F_{COV} = F_{SD}^T * F_{SD} \quad (9)$$

The result is a matrix of size  $N \times N$ , where  $N$  is the number of faces under consideration. It gives  $N$  Eigen values and  $N$  Eigen vectors. The Eigen values are sorted in descending order and will select the first  $N_n$  largest Eigen values and corresponding Eigen vectors. Eigen vectors in  $N_n$  dimension is transformed to the higher dimension of  $M$  by multiplying with Standard deviation Matrix. Now the test data is projected to this lower dimension space to get the corresponding weight vectors.

#### E. Fusion and Face Classification

After collecting the information at multiple levels, next step is fusion of individual information and classification. Classification is the decision making stage. The function of the classifier is to group the unknown incoming face image to one of the known labeled class. Classification is performed using Euclidean distance method, which calculates the distance between test data and trained data and the unlabeled face image is classified into one with minimum error or distance



**Figure 2. The face image is convolved with LPA filters to get directional faces.Optimised using ICI in each direction to obtain Optimised Directional Face in each direction**

One of the important factors in multimodal system is how to fuse the information captured from different sources. Mainly there are 3 types of fusion schemes – feature level fusion, score level fusion and decision level fusion [13]. This paper focuses on score level fusion. Score level fusion combines the matching scores of two modalities and decision is taken based on this combined score. In first step the classifier is allowed to estimate the distance between the training vector and testing vector independently for intensity map and depth map. In next step we combined the individual error scores from depth and texture as a single error value and find the minimum distance of the test image feature vector from the features of other images in the training samples.

### III. EXPERIMENTAL RESULTS

The proposed system is evaluated on a Windows7 system with 4GB RAM and Intel i3 processor using

MATLAB R2012b. For testing and analysis of the proposed system FRAV3D database is used. It contains 16 different poses of 106 persons. Of those 75 subjects is considered for evaluation. There is a significant improvement in recognition rate when fusion scheme is employed. First the recognition accuracy is obtained using texture and depth separately. Accuracy is checked by varying number of persons from 10 to 75 with 16 samples of each person. The performance of the fused scheme is compared with the individual performance. There is a significant improvement in recognition rate when fusion scheme is employed.

Table I and II shows the individual results by varying number of training images from one to three. Texture and depth features are extracted separately and are evaluated. Texture information is more robust than depth in terms of pose variations; on the other hand depth is more pertinent in illumination variations. The results obtained by fusion of texture and depth are tabulated in Table III. Fusion of the matching scores proves that the recognition accuracy can be improved significantly by fusion of scores of multiple representations. From the analysis of tables it is clear that there is some improvement in the accuracy obtained by fusing two modalities. The recognition rate can be further improved by including more number of training images. The spectral transformation using DFT reduces the redundancy in the feature vector thus adds the efficiency. Table IV shows the comparison of recognition rate after applying DFT. Fig. 4 shows the graphical representation of same. From the analysis of table V, it is clear that the preprocessing of input images prior to computations improve the overall performance of the system.

Table VI is the tabulation of complete comparison of the proposed system with existing methods. From the analysis of table it is clear that there is a minimum improvement of 10% using the proposed method. For large number of testing images there is an improvement upto 20%. Fig.5 is the graphical comparison of the proposed system.

**TABLE I. ACCURACY USING TEXTURE ONLY**

No: of persons	No: of samples	Accuracy (%)		
		No. of Training Images		
		1	2	3
10	160	80.23	82.45	90.58
15	240	70.83	77.25	82.50
20	320	64.68	74.58	80.75
25	400	63.50	70.25	77.27
50	800	51.52	57.75	63.57
75	1200	45.75	53.00	57.67

**TABLE II. ACCURACY USING DEPTH ONLY**

No: of persons	No: of samples	Accuracy (%)		
		No. of Training Images		
		1	2	3
10	160	45.2	52.5	64.37
15	240	43.33	51.67	62.91
20	320	41.78	47.5	56.87
25	400	38.5	44.5	52.75
50	800	30.75	36	44.75
75	1200	27.41	33.5	40.75

**TABLE III. ACCURACY USING TEXTURE AND DEPTH WITH DFT**

No: of persons	No: of samples	Accuracy (%)		
		No. of Training Images		
		1	2	3
10	160	77.50	86.75	91.88
15	240	70.42	76.67	81.25
20	320	66.56	74.06	78.44
25	400	61.25	69.75	75.25
50	800	50.88	58.50	63.13
75	1200	46.25	53.42	58.50

**TABLE IV. COMPARISON OF FUSION**

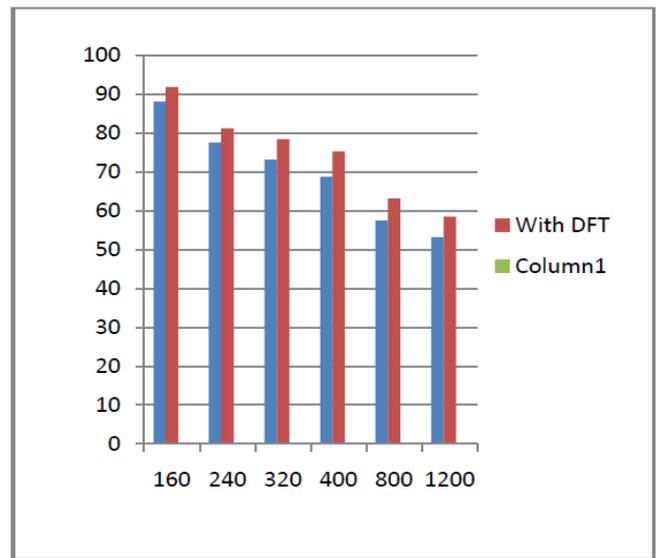
No: of persons	No: of samples	Accuracy (%)	
		Without DFT	With DFT
10	160	88.12	91.88
15	240	77.50	81.25
20	320	73.13	78.44
25	400	68.75	75.25
50	800	57.50	63.13
75	1200	53.25	58.50

**TABLE V ACCURACY WITH PREPROCESSING USING WEINER FILTER**

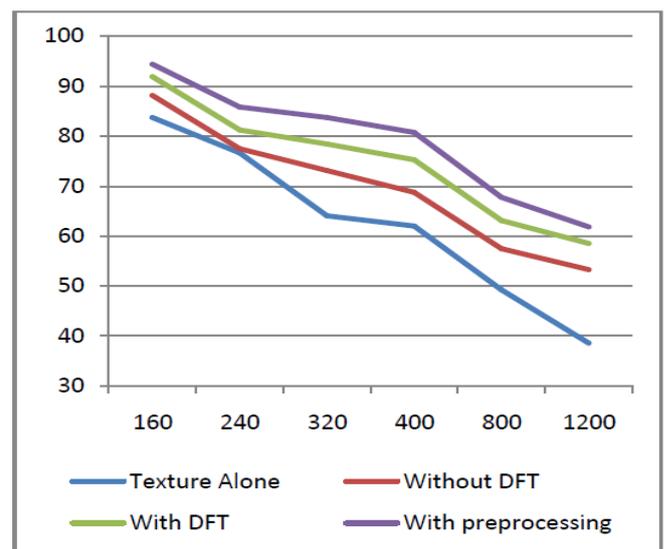
No: of persons	No: of samples	Accuracy (%)		
		No. of Training Images		
		1	2	3
10	160	81.25	85.65	94.38
15	240	75.83	79.17	85.83
20	320	72.82	79.37	83.75
25	400	68.75	74.58	80.75
50	800	57.75	63.24	67.75
75	1200	51.35	57.58	61.83

**TABLE VI COMPARISON OF SINGLE MODAL SYSTEM WITH PROPOSED MULTI MODAL SYSTEM**

No. of persons	No. of samples	Accuracy (%)			
		Texture alone	Texture and depth		
			Without DFT	With DFT	With preprocessing
10	160	83.75	88.12	91.88	94.38
15	240	76.67	77.50	81.25	85.83
20	320	64.06	73.13	78.44	83.75
25	400	62.00	68.75	75.25	80.75
50	800	49.25	57.50	63.13	67.75
75	1200	38.55	53.25	58.50	61.83



**Figure.4. Graphical comparison**



**Figure.5. Graphical comparison of single modal system with proposed multi modal system**

#### IV. CONCLUSION

An efficient multimodal face recognition method using combined textural and depth information is proposed in this work. The information is extracted from Optimized Directional Faces captured at multiple directions. LPA filters are utilized to extract the directional information from face images at multiple scales, which are then optimized using ICI. The texture and depth features are captured from these directional faces using mLBP tool. The feature vector thus obtained is robust and is invariant to physical variations. Texture information is pertinent than depth information in the context of expression variations and depth information is more reliable than texture information in the case of pose variations. Thus, their combination is helpful in improving recognition rate and performance. Besides, spectral transformation of feature vector and preprocessing of face images using Wiener filter improves the recognition accuracy of the system. Further improvement in recognition can be achieved by including more number of training images and with a better classifier.



## REFERENCES

1. R. Mehta n, Jirui Yuan, Karen Egiazarian —, Face recognition using scale-adaptive directional and textural features, Pattern Recognition 47(2014) 1846–1858
2. Mathew Turk and Alex Pent land, —Eigen faces for Recognition, Journal of Cognitive Neuroscience Volume 3, Number 1(1991).
3. P.N. Belhumeur, J.P. Hespanha, D.J.Kriegman, —Eigen faces vs. Fisher faces : recognition using class specific linear projection | IEEE Trans. Pattern Anal. Mach. Intell.(1997)
4. C.Podilchuk, X.Zhang, —Face recognition using DCT based feature vectors, IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP 96, vol.4, 1996 .
5. Chiraz BenAbdelkadera, Paul A. Griffin, Comparing and combining depth and texture cues for face recognition, Image and Vision Computing, Elsevier, (2005) pp.339-352
6. Rinky B, Payal Mondal, K Manikantan., S Ramachandran , —DWT based feature extraction using edge tracked scale normalization for enhanced face recognition |, Procedia Technology 6 ( 2012 ) 344 – 353
7. Antonio Rama, Francesc Tarres, P2CA: A New Face Recognition Scheme Combining 2D And 3D Information, IEEE Transactions, Image Processing, Vol.3 (2005) pp.776-779
8. <http://www.frav.es/databases/FRAV3D>
9. V.Katkovnik, Multiresolution local polynomial regression : a new approach to pointwise spatial adaptation, Digital Signal Process.(2005)73–116.
10. Katkovnik., K. Egiazarian, J.Astola, —Adaptive window size image de-noising based on intersection of confidence intervals(ICI) rule, J.Math. Imaging Vision (2002)223–235
11. P.Liang, S.Li, J.Qin ,Multi resolution local binary patterns for image classification, in: Proceedings of the Conference on Wavelet Analysis and Pattern Recognition, 2010, pp.164–169.
12. Naveen S. and Dr. R.S Moni, —Multimodal Approach for face recognition using 3D-2D face feature fusion, International Journal of Image Processing Vol.8, No.3 (2014)
13. P. S. Hiremath and Manjunatha Hiremath, —3D Face Recognition Based on Depth and Intensity Gabor Features using Symbolic PCA and AdaBoost, International Journal of Signal Processing Vol.6, No.5 (2013)

**Shilpa S Nair**, is currently pursuing M.Tech. Degree in Signal Processing with the Department of Electronics and Communication Engineering, LBS Institute of Technology for women, Poojappura, Trivandrum, Kerala. She received B. Tech degree from M.G University, Kottayam, in 2012 in Electronics and Communication Engineering.

**Naveen S**, currently working as Assistant Professor, in department of Electronics and Communication Engineering at LBS Institute of Technology for Women, Thiruvananthapuram, Kerala, India. Has Obtained B.Tech and M.Tech degrees from University of Kerala and is currently pursuing research in the area of pattern recognition.

**Dr. Moni.R.S.** Received the PhD degree in Electrical Engineering from Indian Institute of Technology, Chennai, India in 1983. He worked as the Principal, College of Engineering Thiruvananthapuram, Kerala and retired as Director of Technical Education, Kerala state. After the retirement from Government service, he is working as visiting faculty in engineering colleges. His present assignment is as Professor in Marian Engineering College, Thiruvananthapuram, Kerala, India. Has several national and international publications to his credit. His area of research includes signal processing; Image processing, communication and instrumentation