Hand Gesture Recognition by Walsh Transformation

Tanuja K. Sarode,
PhD
Thadomal Shahani
Engineeringg, College,
Affiliated to Mumbai,
Univiersity, India.

Priyanka Agrawal
Thadomal Shahani
Engineering College,
Affliated to Mumbai
University, India.

Gaurav Deshpande
Thadomal Shahani
Engineering College,
Affliated to Mumbai
University, India.

Anjali Jogeshwar
Thadomal Shahani
Engineering College,
Affliated to Mumbai
University, India.

ABSTRACT
Modern day computers are shifting from pointer based devices to systems based on human actions and interaction methods. Thus, gesture and voice recognition forms an integral part of research. In gesture recognition, hand gestures of natural languages have a great potential in development of vision based hand interaction. The paper presents Hand Gesture Recognition method based on Walsh Transform and also presents its comparative study with other popular transforms DCT and Haar. The approach consists of three important phases 1) Image Pre-processing 2) Image Transformation 3) Image Recognition. Best accuracy was found to be 100per cent

General Terms
Pattern Recognition

Keywords
Hand Gesture Recognition, Walsh Transform,

1. INTRODUCTION
Gesture and voice recognition thus forms an integral part of modern day computers and other systems. Hand gesture recognition is an important aspect of virtual reality and human computer interaction. Modern day computers are shifting from pointer and key based devices like mouse, keyboard, touch screens etc. to systems that go along with the natural human actions and interaction methods. Hand gesture methods based on natural languages have a great potential in development on vision based hand interaction method.

Current methods of gesture recognition have been concentrating on improving the accuracy and the speed of recognition. Jagdish Raheja et al[1] proposed Principle Component Analysis method for hand recognition. In his system, hand gestures of 10 different gestures give 90% accuracy in proper illumination. Yuehai Wang et al[2] used Morphological Method for gesture recognition based on skin color extraction, dilation, erosion and subtraction for hand gesture recognition. Deg-Yuan Huang et al[3] used gabor filter and support vector machine, principle component method to reduce the dimensionality of the feature space.

This paper presents hand gesture recognition based on Walsh transform[4]. The paper also presents comparative study of the algorithm used under DCT[5] and Haar Transform. The transforms are applied under four different methods of scaled image, cornered image, row and column mean method to study accuracy under them.
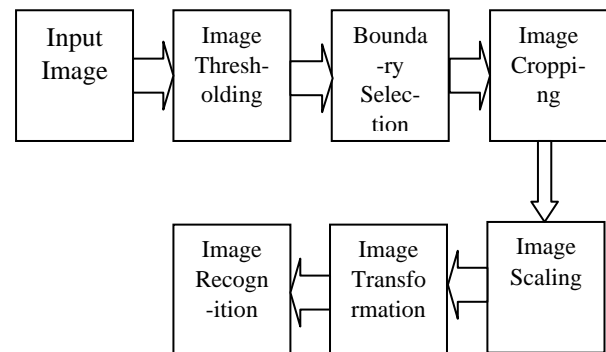
2. SYSTEM DESCRIPTION

Fig.1: Block Diagram of the System

This section consists of three major components such as pre-processing, feature extraction and pattern matching. The block diagram of the system is as shown in the Fig.1.

2.1 Image Pre-processing
Image pre-processing is used to discard the unwanted and unimportant part in the database image. Image Pre-processing rids the gesture image of its background and the object that aren't used in the recognition of the gesture[6].

The image pre-processing steps are as follows:

1. Repeat step 2-7 for every image in the database.

2. Convert the image to grayscale.

3. Determine the global threshold level of the image.

4. Convert the selected grayscale image into binary image using this global threshold level.

5. Now to extract the object of interest, from the binary image search for the first and the last occurrence of the white pixel while traversing the image horizontally and vertically. This determines the cropping point for the image.

6. Crop the image using the cropping point determined previously.

7. The image is then resized to a standard size i.e., 128x128.

2.1.1 Image Thresholding
It is a segmentation technique that partitions the image directly into regions based on the intensity values.

$$g(x, y) = 0 \quad iff \text{ x ,y } > T$$
$$1 \quad iff \text{ x, y } \leq T, \qquad (1)$$

Where T is global threshold level

The algorithm assumes that the image contains two classes of pixels following a bimodal graph(background and foreground). Focuses on minimizing intra-class variance and maximize inter-class variance.

2.2 Image Transforms

On the basis of type of technique (column mean, row mean, cornered image etc.) used the feature vector is three or two dimensional. In three dimensional array (x,y,z) the x-plane and y-plane defines the transformed image, while in two dimensional array(used in column and row mean technique) the column or row defines the transformed image. The transform technique applied in the system are:

1. Discrete Cosine Transform
2. Haar Transform
3. Walsh Transform

2.2.1 Discrete Cosine Transform

The discrete cosine transform(DCT)[7] is closely related to the Discrete Fourier Transform. The definition of the two dimensional Discrete Cosine Transform can be written as follows, in terms of pixel values f(x, y) for x, y = 0, 1, N-1 and the frequency domain transform coefficient F(u, v),

$$F(u, v) = \alpha(u)\alpha(v) \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} f(x,y)$$
$$cos\left(\frac{(2x+1)u\pi}{2N}\right) cos\left(\frac{(2y+1)v\pi}{2N}\right) \qquad (2)$$

For u, v = 0, 1, 2 … N-1 and

$$f(x, \qquad y) =$$
$$\sum_{u=0}^{N-1} \sum_{v=0}^{N-1} \alpha(u)\alpha(v)F(u,v) cos\left(\frac{(2x+1)u\pi}{2N}\right) cos\left(\frac{(2y+1)v\pi}{2N}\right)$$
(3)

The N x N Cosine Transform Matrix C = {C(u, v)} us defined as

$$C(u, \qquad v) =$$
$$\begin{cases} \sqrt{1}/N & u = 0, 0 \leq v \leq N-1 \\ \sqrt{2}/N \cos[\pi(2v+1)u/2N] & 1 \leq u \leq N-1, \ 0 \leq v \leq N-1 \end{cases}$$
(4)

The Cosine Matrix is real and orthogonal but not symmetric,

$$C = C^* \qquad C\text{-}1 = C'$$

$$C.C' = I \qquad (5)$$

The 2D DCT of an image can be generated using equation,

$$F = C f C' \qquad (6)$$

If DCT is applied to real data, the result is also real. The DCT tends to concentrate information, making it useful for image compression applications and also helping in minimizing feature vector size for different applications. For full 2-Dimensional DCT for an NXN image the number of multiplications required are N2 (2N) and number of additions required are 2N2 (N-1)[8].

2.2.2 Haar Transform

This sequence was proposed in 1909 by Alfréd Haar. Haar functions are used to give an example of a countable orthonormal system for the space of square-integral functions on the real line. The Disadvantage of the Haar wavelet is that it is not continuous, and therefore not differentiable. The Haar Transform[8] is derived from Haar matrix. The Haar transform like most of the other transform is separable and can be expressed in the matrix form.

$$F = HfH' \qquad (7)$$

Where f is an NxN image, H is a NxN matrix and F is resultant NxN transform. The transform H contains the Haar basis function $h_k(x)$ which are defined over the continuous closed interval x[0,1]

Haar basis functions are,

When, k = 0 the Haar function is defined as a constant

$$h0(t) = 1 / \sqrt{N} \qquad (8)$$

When, k>0 the Haar function is defined by

$$Hk(t) = 1 / \sqrt{N}$$
$$\begin{cases} 2^{p/2} \ (q-1)/2^p & \leq t < (q-0.5)/2^p \\ -2^{p/2} \ (q-0.5)/ & \leq t < q/2^p \\ 0 & otherwise \end{cases}$$
(9)

From the definition, it can be seen that p determines the amplitude and width of the non-zero part of the function, while q determines the position of the non-zero part of the function.

2.2.3 Walsh Transform

The Walsh functions[9] consist of trains of square pulses (with the allowed states being −1 and 1) such that transitions may only occur at fixed intervals of a unit time step, the initial state is always **+1**, and the functions satisfy certain other orthogonality relations. In particular, the 2^n Walsh functions of order *n* are given by the rows of the Hadamard matrix H_{2^n} when arranged in so-called "sequency" order. There are 2^n Walsh functions of length 2^n.

Walsh transform matrix [7] is defined as a set of N rows, denoted W_j, for j = 0, 1, ... , N - 1, which have the following properties:

W_j takes on the values +1 and -1.

$W_j[0] = 1$ for all j.

$W_j \times W_k^T = 0$, for j≠ k and $W_j \times W_k^T = N$, for j=k.

W_j has exactly j zero crossings, for j = 0, 1, ..., N-1.

Each row W_j is even or odd with respect to its midpoint.

Walsh transform matrix is defined using a hadamard matrix of order N. The row of Walsh transform matrix is considered as the row of the hadamard matrix. It specified by the Walsh code index, which must be an integer in the range [0, ..., N - 1]. For the Walsh code index equal to an integer j, the respective Hadamard output code has exactly j zero crossings, for j = 0, 1, ..., N - 1.

For the full 2-Dimensional Walsh transform applied to image of size NXN, the number of additions required are $2N^2$ (N-1) and absolutely no multiplications are needed in Walsh transform [7].

3. PROPOSED ALGORITHM

3.1 Techniques

Four basic techniques have been applied namely row mean, column mean, corner image (i.e., cropping the image and pasting it on the upper left corner of a 128x128 pixels black image) and scaled image (cropping the image and scaling it to 128x128).

3.1.1. Scale image

Here, the hand gesture is cropped and scaled to size 128x128. This is done for all the images from the set of training images and stored as a .mat file. Then, an image is picked (one by one from the testing set), cropped, scaled to size 128x128 and compared for errors. If the error is least for the same alphabet, it is a hit else it's a miss. Table 4.4 shows the results for varying number of training and test images. Walsh has been applied on every greyscale image and its threshold image. The accuracy is evaluated on fully and partial transformed co-efficient i.e., 128x128, 64x64, 32x32 and 16x16.

3.1.2 Row mean

Here the mean of every row of the image from the set of training images is taken and stored as a .mat file. Then, an image is picked (one by one from the testing set), its row mean calculated and compared for errors. If the error is least for the same alphabet, it is a hit else it's a miss. Table 1 shows the results for varying number of training and test images. Walsh has been applied on every greyscale image and its threshold image. The accuracy is evaluated on fully and partial transformed coefficient i.e., 128x128, 64x64, 32x32 and 16x16.

3.1.3 Column mean

Here the mean of every row of the image from the set of training images is taken and stored as a .mat file. Then, an image is picked (one by one from the testing set), its row mean calculated and compared for errors. If the error is least for the same alphabet, it is a hit else it's a miss. Table 2 shows the results for varying number of training and test images. Walsh has been applied on every greyscale image and its threshold image. The accuracy is evaluated on fully and partial transformed coefficient i.e., 128x128, 64x64, 32x32 and 16x16.

3.1.4 Corner image

Here, the hand gesture is cropped and pasted on the left corner of a black image. This is done for all the images from the set of training images and stored as a .mat file. Then, an image is picked (one by one from the testing set), cropped, pasted on black image of size 128x128 and compared for errors. If the error is least for the same alphabet, it is a hit else it's a miss. Table 4.3 shows the results for varying number of training and test images. Walsh has been applied on every greyscale image and its threshold image. The accuracy is evaluated on fully and partial transformed co-efficient i.e., 128x128, 64x64, 32x32 and 16x16.

Fig 2: Cropped Image at the corner of a 128x128 image

3.2 The Algorithm To Generate The Feature Vector Database Gesture Recognition Of Training Images

1. Repeat step 2-4 for every image in the database.
2. Apply image-pre-processing techniques to the image.
3. Apply the full / row mean / column mean transformations (DCT[8] / Haar / Walsh) to the pre-processed image.
4. Store the transformed coefficient in the feature vector database.

3.3 The algorithm for matching hand gesture recognition of testing images

1. Repeat step 2-7 for every image in the test database.
2. Apply image-pre-processing techniques to the image.
3. Apply the full / row mean / column mean transformations (DCT / Haar / Walsh) to the pre-processed image.
4. Compute the Squared Error method of the feature vector of the current test image with the feature vector database.
5. Store the error in an array.
6. Determine the index of the minimum error from the array.
7. From the index determine the corresponding gesture.

3.4 Squared Error

If $\hat{Y}$ is a vector of n predictions, and Y the vector of observed values corresponding to the inputs to the function which generated the predictions, then the SE of the predictor can be estimated by,

$$SE = \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2 \qquad (10)$$

4. RESULT AND DISCUSSION

The implementation of the proposed algorithm is done in MATLAB 10 using a computer with Intel Core i3 processor. The proposed algorithm is tested on 48 different gestures representing American Sign Languages [10]. The hand gestures for training were collected with the plane background by same hand gesture 5 times; each time from a different angle and position. The dataset contains 120 (=24*5) images of each hand gestures. For testing 48(=24*2) images have been used.

Table 1 shows the accuracy of full, row mean and column mean using DCT transform for scaled image to 128x128 and cornered image on binary and grey scale images for varying feature vector size from 128x128 to 16x16.

Table 2 shows the accuracy of full, row mean and column mean using Haar transform for scaled image to 128x128 and cornered image on binary and grey scale images for varying feature vector size from 128x128 to 16x16.

Table 3 shows the accuracy of full, row mean and column mean using Walsh transform for scaled image to 128x128 and cornered image on binary and grey scale images for varying feature vector size from 128x128 to 16x16.

Fig 3 shows the comparative study of column mean, row mean, cornered and scaled images using DCT, Haar and Walsh transformation for 30 training and 10 test binary images.

Table 1: Accuracy of full, row mean and column mean using DCT transform for scaled image to 128x128 and cornered image on binary and grey scale images for varying feature vector size from 128x128 to 16x16.

DCT Transform							
Method	Dimension (pixels)	Binary Image			Grey Scale Image		
		30 10	25 15	20 20	30 10	25 15	20 20
Column Mean	128x128	53.33%	28.06%	54.12%	25.83%	65.56%	59.38%
	64x64	53.75%	28.06%	4.17%	4.17%	65.56%	4.17%
	32x32	51.25%	27.50%	4.17%	4.17%	65.00%	4.17%
	16x16	49.17%	26.94%	4.17%	4.17%	61.39%	4.17%
Row Mean	128x128	47.08%	49.17%	45.00%	42.20%	41.11%	54.17%
	64x64	47.50%	49.17%	52.92%	42.08%	41.67%	4.17%
	32x32	46.67%	50.28%	49.17%	43.33%	41.94%	4.17%
	16x16	46.67%	47.22%	47.71%	44.17%	44.17%	4.17%
Cornered Image	128x128	**93.33%**	**91.39%**	81.25%	**96.67%**	**95.00%**	40.21%
	64x64	**93.75%**	**91.67%**	4.17%	4.17%	**95.00%**	4.17%
	32x32	**95.42%**	**91.94%**	4.17%	4.17%	**95.28%**	4.17%
	16x16	**95.42%**	**94.44%**	4.17%	4.17%	**96.11%**	4.17%
Scaled Image	128x128	58.33%	60.56%	54.38%	63.33%	67.22%	79.58%
	64x64	58.33%	60.83%	4.17%	4.17%	67.22%	4.17%
	32x32	57.92%	61.67%	4.17%	4.17%	67.50%	4.17%
	16x16	61.25%	62.50%	4.17%	4.17%	67.78%	4.17%

Table 2: Accuracy of full, row mean and column mean using Haar transform for scaled image to 128x128 and cornered image on binary and grey scale images for varying feature vector size from 128x128 to 16x16.

Haar Transform							
Method	Dimension (pixels)	Binary Image			Grey Scale Image		
		30 10	25 15	20 20	30 10	25 15	20 20
Column Mean	128x128	52.50%	53.33%	54.17%	59.58%	65.00%	49.38%
	64x64	35.42%	39.72%	44.79%	4.17%	46.11%	48.75%
	32x32	24.58%	23.06%	21.25%	4.17%	19.44%	20.00%
	16x16	22.08%	21.94%	16.88%	4.17%	11.67%	14.17%
Row Mean	128x128	46.67%	49.17%	45.21%	4.17%	41.39%	40.42%
	64x64	52.50%	51.67%	45.63%	4.17%	49.72%	41.88%
	32x32	53.33%	49.44%	38.96%	4.17%	51.11%	45.83%
	16x16	52.50%	44.44%	34.79%	4.17%	41.94%	39.79%
Cornered Image	128x128	**94.58%**	**91.67%**	83.71%	4.17%	**95.28%**	74.58%
	64x64	**94.58%**	**91.67%**	80.21%	4.17%	**95.28%**	74.58%
	32x32	87.50%	84.17%	78.33%	4.17%	88.06%	74.79%
	16x16	66.67%	60.56%	53.96%	4.17%	72.78%	74.58%
Scaled Image	128x128	57.92%	61.11%	54.38%	4.17%	67.50%	52.08%
	64x64	75.83%	71.67%	65.21%	4.17%	75.00%	61.67%
	32x32	60.83%	63.33%	54.38%	4.17%	67.78%	30.83%
	16x16	56.67%	55%	51.04%	4.17%	55.28%	14.17%

Table 3: Accuracy of full, row mean and column mean using Walsh transform for scaled image to 128x128 and cornered image on binary and grey scale images for varying feature vector size from 128x128 to 16x16.

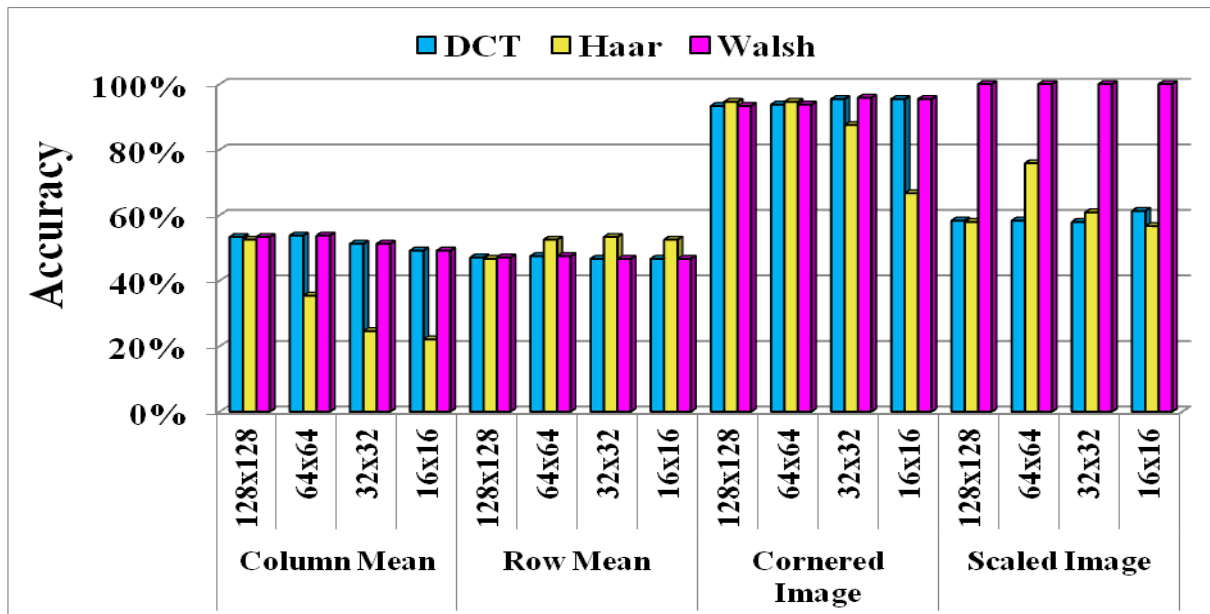| Walsh Transform							
Method	Dimension (in pixles)	Binary Image			Grey Scale Image		
		30 10	25 15	20 20	30 10	25 15	20 20
Column Mean	128x128	53.33%	3.06%	4.17%	59.58%	3.61%	5.21%
	64x64	53.75%	3.06%	4.17%	4.17%	1.39%	4.79%
	32x32	51.25%	3.06%	4.17%	4.17%	2.50%	4.79%
	16x16	49.17%	3.06%	4.17%	4.17%	3.61%	5.21%
Row Mean	128x128	47.08%	2.78%	45.00%	36.67%	3.61%	5.21%
	64x64	47.50%	2.78%	52.20%	4.17%	1.39%	4.79%
	32x32	46.67%	2.78%	49.17%	4.17%	2.50%	4.79%
	16x16	46.67%	2.22%	47.71%	4.17%	3.61%	5.21%
Cornered Image	128x128	**93.33%**	**95.00%**	30.63%	4.17%	**91.39%**	79.58%
	64x64	**93.75%**	**95.28%**	4.17%	4.17%	**91.11%**	81.67%
	32x32	**95.83%**	**96.11%**	4.17%	4.17%	**91.39%**	82.71%
	16x16	**95.42%**	**96.39%**	4.17%	4.17%	**91.94%**	81.67%
Scaled Image	128x128	**100.00%**	60.56%	54.38%	63.33%	35.28%	37.08%
	64x64	**100.00%**	60.83%	4.17%	4.17%	35.28%	37.08%
	32x32	**100.00%**	62.50%	4.17%	4.17%	35.28%	37.29%
	16x16	**100.00%**	62.50%	4.17%	4.17%	35.28%	36.67%

Fig 3: Comparative study of column mean, row mean, cornered and scaled images using DCT, Haar and Walsh transformation for 30 training and 10 test binary images.

5. CONCLUSION

In modern computing world, more and more complex systems are being developed. These systems are aimed at assisting humans in various complex tasks. Using pointer based devices and other input devices that are not based on conventional human interaction methods limit the interaction capability of the systems. To improve these systems, devices based on conventional human interaction methods are necessary. Hand

gesture being a prominent interaction method, systems based on them should have speed and accuracy.

In the presented system, using Walsh transform 100% accuracy has been obtained, using DCT 96.67% accuracy has been obtained and using Haar transform 95.28% accuracy has been obtained. The presented system's approach is simple and accurate.

Future Scope for the project will focus on complex gestures, with complex background and also improve on speed of image recognition. It will also work on working on images with higher pixel density.

6. REFERENCES

[1] Jagdish Lal Raheja, Radhey Shyam, Umesh Kumar, P Bhanu Prasad, " Real-Time Robotic Hand Control using Hand gestures" proceeding of International Conference on Machine Learning and Computing (ICMLC) of IEEE, pp.12-16, Feb 2010.

[2] Yuehai Wang, JianfeiLi "Entertainment Robot Hand Gesture Recognition" proceeding of International Conference on Database Technology and applications (DBTA), pp. 27-28, Nov. 2010.

[3] Yikai Fang , Kongqiao Wang ,Jian cheng,Hanqing Lu "A Real-Time Hand Gesture recognition Method" Proceeding of International conference on Multimedia and Expo. IEEE pp.995-998, July 2007.

[4] H. B. Kekre, Tanuja Sarode, Meena Ugale, "Performance Comparison of Image Classifier Using Discrete Cosine Transform and Walsh Transform" Proceeding of 2nd International Conference and workshop on Emerging Trends in Technology (ICWET) 2011.

[5] H. B. Kekre, Sudeep D. Thepade and Akshay Maloo "Performance Comparison of Image Retrieval Using Fractional Coefficients of Transformed Image Using DCT, Walsh, Haar and Kekre's transform" Proceeding of International Journal of Image Processing (IJIP) .

[6] Tanuja Sarode, Vaishali D. Sakhpal, "PerformanceComparison of Transforms Using Row Mean and Column Mean for Hand Gesture Recogniton" Proceeding International Journal of Computer Applications pp.0975-8887, September 2013.

[7] H.B. Kekre, Tanuja Sarode, Pratik Bhatia, Sandhya Nayak, Dhreej Nagpal, "Iris recognition using Partial Coefficients by applying Discrete Cosine Transform", International Journal of Computer Applications(IJCA), vol 32, No. 6 October, 2011.

[8] H.B. Kekre, Sudeep Thepade, Juhi Jain, Naman Agrawal, "IRIS Recognition using Texture Features Extracted from Haarlet Pyramid", International Journal of Computer Applications(IJCA), vol 11, No. 12 December, 2010.

[9] H.B. Kekre, Tanuja Sarode, Sudeep Thepade,Sonal Shroff "Instigation of Orthogonal Wavelet Transforms using Walsh, Cosine, Hartley, Kekre Transforms and their use in Image Compression", International Journal of Computer Science and Information Security(IJCSIS), vol 9, No. 6, 2011.

[10] Thomas Moeslund's Gesture Recognition database http://www-Prima.inrialpes.fr/FGnet/data/12-Moeslund Gesture / database.html.

7. AUTHOR PROFILE

Dr. Tanuja K. Sarode has received BSc.(Mathematics) from Mumbai University in 1996, BSc. Tech(Computer Technology) from Mumbai University in 2004, PhD. Mukesh Patel School of Technology, Management and Engineering, SVKM's NMIMS University, Vile Parle (W), Mumbai, INDIA. She has more than 17 years of experience in teaching and currently working as Associate Professor in Dept. Of Computer Engineering at Thadomal Shahani Engineering College, Mumbai. She is life member of IETE and ISTE. Her areas of interest are Image Processing, Signal Processing and Computer Graphics. She has more than 160 papers in National/International Conferences/journal to her credit.

Priyanka Agrawal is currently pursuing B.E. in Computer Engineering from Thadomal Shahani Engineering College. She has keen interest in coding and teaching. She aims to complete her B.E. degree then wishes to pursue a Masters degree in the same field.

Gaurav Deshpande is currently pursuing B.E. in Computer Engineering from Thadomal Shahani Engineering College. He has keen interest in designing algorithms and programming. He aims to complete his B.E. degree and then wishing to pursue a Masters degree in the same field.

Anjali Jogeshwar is currently pursuing B.E. in Computer Engineering from Thadomal Shahani Engineering College. She has keen interest in Computer Science and Image Processing. She aims to complete his B.E. degree and then wishing to pursue a Masters degree in the same field.