A review on Face Detection and study of Viola Jones method

Monali Chaudhari[#1], Shanta sondur[*2], Gauresh Vanjare[$3]

[#]*Assistant Professor, Dept. of Elec. & Telecom, V.E.S.I.T, Mumbai, INDIA.*
[*] *Professor, Dept. of Information. & Technology, V.E.S.I.T, Mumbai, INDIA.*
[$]*B.E student, Dept. of Elec. & Telecom, V.E.S.I.T, Mumbai, INDIA.*

Abstract— *This paper describes the necessity and adopted methods to detect a human face. Since the data is computed by the computer, many algorithms are developed to detect a face. Some of the key challenges for the process of face detection are discussed. Four general face detection methods that are universally used are elaborated with their capabilities, advantages and disadvantages. A rapid approach to detect face developed by viola and jones is explained in brief. The 4 main concepts involved in the viola jones method such as haar features, integral image, Adaboost and classifier cascade are demonstrated.*

Keywords— *Adaboost, Face detection, challenges, haar cascade.*

I. INTRODUCTION

In the past 10 years, there has been an exponential development in the field of image processing. New generation computers are becoming smarter and faster to process Terabytes of data. Object recognition is a field which has gained attention because of its utility in industries such as manufacturing, packing etc. A common example is supermarket that uses barcode to identify a product. A manufacturing unit employs a device that needs to accurately find the position of an object. For example, a bottle filler robotic arm needs to know the relative position of the bottle in which the liquid is filled. Object recognition needs computer vision to detect the structural nature of the object. Computer vision allows the computer to sense the object and process the desired information at blazing speeds. Just like eyes are the source for human, a camera is the undisputed source for computer.

Face detection is also a part of object detection [1].Face detection can be classified into two classes (face and non-face). Most applications are based on face recognition and tracking. These applications needs to locate the position of the face in the image or video. Moreover, face detection has added a much needed aspect of security in the recent years. Biometric systems, a front sided camera (Selfie) of a smart phone, human presence detection are some of the key implementations of face detection. Basically face detection senses the presence of the face in a 2D frame. Fig 1. shows human faces that has a close resemblance with a uniform structures. Hence the detection of human face is possible [2]. Several methods and approaches are developed for face detection.

Fig 1. The extended Yale face database B.

The results generated by the computer depends upon the data feed into it, the correctness and the reliability of the algorithm used. To evaluate the performance of the face detectors some parameters are used as a standard. Some of them are [3]:

a)Learning time: time required to adapt and improve the reliability to differentiate between a face and non-face,

b)Samples needed for training: more the number of samples more is the accuracy of the result obtained sacrificing the speed of detection,

c) The ratio between detection rate and false alarm, **d) Execution time**.
In section 3, some factor affecting the face detection algorithm are discussed. Section 4 describes the commonly used face detection approaches. Section 5 throws light on viola jones method with adaptive boost learning.

II. LITERATURE REVIEW

Over the years many contribution were done to the field of face detection and recognition. G. Yang came up with Multiresolution rule method. This knowledge based method used the structural nature of the face for detection [4]. Feature based method uses the facial features [5][6], skin color [7][8]and combined multiple features [9]of the face for better accuracy and detection speed. In order to increase the detection speed, the accuracy is sacrificed. For this, a steady and uniformly scaled images using template matching method was employed. Predefined face templates [10] and deformable templates [11] were incorporated which was completely based on the

template (a predefined structure) without using learning. Appearance based methods gives faster detection speeds, more accurate results and adaptive nature that could distinguish a face from a non-face in any environmental conditions. Neural networks [12] is commonly used model for getting the desired results.

III. CHALLENGES

The challenges associated with face detection can be attributed to the following factors:

A. Head pose:

As the human head is somewhat like a sphere in 3D space, hence it is difficult to get the image of the face facing towards the viewer and at the center position all time as shown in Fig 2. Due to this, sometimes an eye, partial nose would be blocked [13]. Hence the face may appear in half profile and would make it difficult to extract the features that would rather make it easy to detect a face.

Fig 2. Head rotated in horizontal direction.8

B. Facial expression:

Sometimes in face recognition, a prerecorded face pattern may not match with the current facial expression. This ambiguity is due to slight variation in the facial expression. Fig 3. represents some facial expressions from the database collected from University of Minnesota.

Fig 3. Facial expression of a person.

C. Image orientation:

An image may be upside-down, rotated or mirror image of original image. This foils the process of face detection. Fig 4. Shows the orientation of the face in a 3D space.

Fig 4. Same face in different pose.

D. Obstruction in front of the face:

Viola jones method uses haar features to classify that detected object is a face or not based on the difference in the intensity level of the different part of the face. The presence of another object in front of the face hinders the detection of the required feature to possibly detect a face. Hence more features with more data points are incorporated in the basic face detection method to increase the efficiency. Fig 5. Shows some of the occlusions such as a sunglass on the face, mouth covered, eyes coved. Apart from these factors hairs in front of the face or else an object in front of the face can also affect the face detection.

Fig 5. Simple occluded face images.

E. Illumination:

It is the most crucial factor in deciding the quality of an image and also decides the evaluation time to detect a face. A bright image impede the detection of certain set of features of the face. Similarly, a dark image produces low contrast image making it difficult to detect the variation in the intensity level along the face as shown in Fig 6.

Fig 6. Harvard database demonstrating the wide range of illumination variation.

F. The computational time and speed:

Face detection method must be robust and accurate to detect a face. At the same time it must be fast enough to be applied in the real time application like security and industry application. To increase the accuracy and detection speed neural networks are used.

IV. FACE DETECTION METHODS

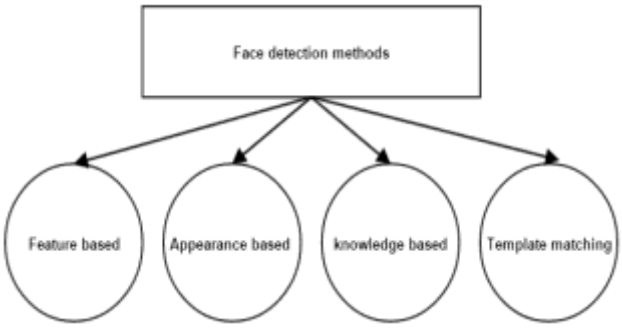


Fig 7. Commonly used face detection methods.

A. Feature based

Feature based approach [14] makes use of the unique elements of face. Elements such as eyes, nose and lips are used to detect a face. The structure and size of these elements are different from the non- face object. Hence Feature based approach is reliable. Also these elements are always on the face of the subject, their position is related to each other. Nowadays, skin is used to detect the face as a human skin has unique texture and color than that of non-face objects. Moreover, edges also play an important role in highlighting the desired region. Blobs and contours are used for this purpose. Feature based approach needs to localize the face on image [15]. This method finds the skin region and validates by searching at least one eye in the frame. The next step is identification of facial features (fiducial points). The H-plane of the HSV color space is used to detect the eye pupil from eye detected region [3].

The eye look for area is minimized by assuming the eyes expected position to be at the upper portion of the face. Haar-like features cascade is used for the eye detection. It locates the rectangular regions which containsthe eyes. In a given ROI of eyes, an algorithm is developed to locate the eye pupil by taking Hue information of the eye image. The hue image is threshold and contour is detected in the threshold image. Centroid of the contour is detected as the eye pupil. Next, the nose is detected using haar-like features. Having known the eyes center, and the position of the nose, an approach is proposed based on the facial geometry for mouth location estimation. Based on the uniform distance maintained between the eye and the nose tip for a detected face location of the other element of the face can be known.

Advantage:

- Features are invariant to pose and orientation change.

Disadvantage:

- Hardship to locate facial features because of several corruption (illumination, noise, occlusion).
- Difficult to detect features because complex background.

B. Knowledge based

Knowledge based approach is also known as Top-down approach. This method uses the geometrical nature of the face. A human Face has two eyes which are symmetrical to each other and one nose at the center point of the face[22].

G. Yang and T.S. Huang [4] presented an algorithm for rule based face detection in frontal views for multiresolution image.A face can be represented using a set of human-coded rules Fig 8. The general steps are

a) The center part of face has uniform intensity values.
b) The difference between the average intensity values of the center part and the upper part is significant.
c) A face often appears with two eyes that are symmetric to each other, a nose and a mouth.

Fig 8. A typical face used in knowledge-based top-down methods.

There are three levels in this methods:

Level 1 (lowest resolution):

Apply the rule "the center part of the face has 4 cells with a basically uniform intensity" to search for candidates.

Level 2: local histogram equalization followed by edge equalization followed by edge detection.

Level 3: search for eye and mouth features for validation.

Fig 9 : (a) n = 1, original image. (b) n = 4. (c) n = 8. (d) n = 16. Original and corresponding low resolution images. Each square cell consists of n x n pixels in which the intensity of each pixel is replaced by the average intensity of the pixels in that cell.

This test was carried out on 60 images. The method detected face in 50 images with 28 as false face detection.One attractive feature of this method is that a coarse-to-fine or focus-of-attention strategy is used to reduce the required computation

C. Appearance based

It rely on techniques from statistical analysis and machine learning to find relevant features of face and non-face. The learned characteristics are in the form of distributed models or discriminant functions that are consequently used for face detection [16]. Dimensionality of the image is reduced for better computational efficiency. This method uses classification and uses the features in the search window. Modified census Transform and Haar-like features are commonly used. The statistical method needs learning between the enormous instances. For instance, adaptive boosting [17] combines some weak classifiers to create a saccade classifier of multistage. An image containing face has some strong classifiers and weak classifiers. The search window detects the weak classifier and if they are above certain threshold then it passes the weak classifier to strong classifier. This methods of this method is the non-face window is easily rejected which increases the speed and accuracy.Other examples are multilayer perceptron neural network, support vector machine, Hidden markov model, principal component analysis.
Methodology is explained in Fig 10. At first preprocessing of the input image is done. Preprocessing treatment includes rotation, scaling and quantizing. Then the classifiers are used for labelling and distinguishing the needed data. Post processing contours around the detected face.

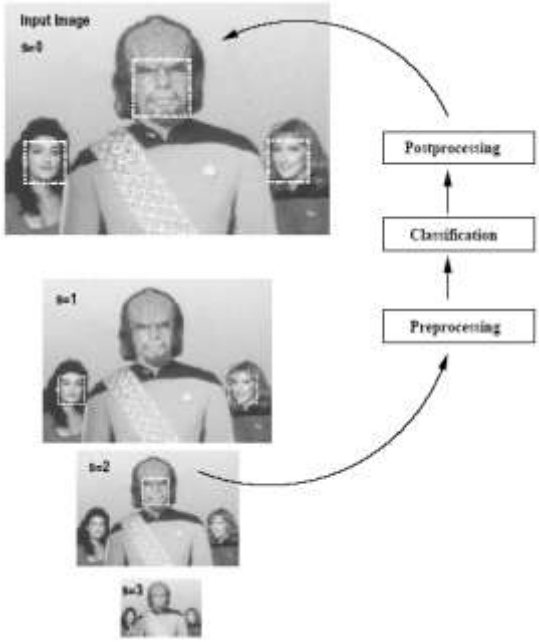


Fig 10. Facial feature recognition for classification

Advantages:

- Use powerful machine learning algorithms.
- Has demonstrated good empirical results.
- Fast and fairly robust.
- Extended to detect faces in different pose and orientation.

Disadvantages:

- Usually needs to search over space and scale.
- Need lots of positive and negative examples.
- Limited view-based approach.

D. Template matching

It is commonly used in the systems where there is high possibility of getting a human face. A template is predefined structure of a uniform size and shape that makes detection of desired object easy just by comparing the template with the objects. In case of

face detection, the template matching finds the relation between the input image or video and the face patterns or the features. Fig 11. shows a template for face detection.

Fig 11. A template of human face shape oriented in vertical and rotated form.

Template matching method is deformable and based on the facial contours. Unlike the appearance based method which uses neural network, templates are hand coded (not learned) and uses correlation to locate the faces.

Advantages

- Simple method
- Include less amount of data points for face detection

Disadvantages

- For frontal views, face must be having no occlusion
- Face(s) must of same size as that of the template
- This method is dependent to size, scale and rotation
- Computational efficiency is less
- To cover more views of the face, more number of templates are needed and hence needs more time to detect a face.

V. VIOLA JONES APPROACH

A very fast and accurate approach to detect an object was devised by viola and Jones[18] in the year 2001.

Nowadays, this method is used in cell phone cameras, security perimeters and list goes on. Due to the use of Haar features and adaboost machine learning computational speed increased. And within a millisecond a face can be detected in a frame. Further improvements were done by Lienhart and Maydt [19] in the year 2002.In this method, firstly, the value of all pixels in greyscale images which are in black accumulated. Then, they subtracted from the total of white boxes. Finally, the result will be compared to the defined threshold and if the criteria is met, the feature considers a hit.

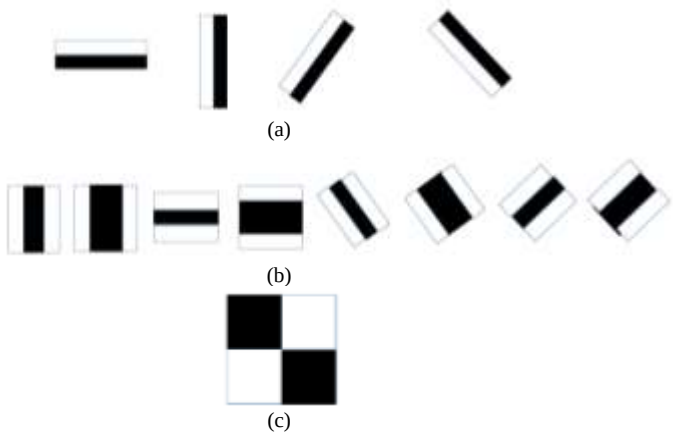
This approach to detecting objects combines four key concepts:

1. Simple rectangular features, called Haar-like features.
2. Integral image for rapid features detection
3. AdaBoost machine-learning method
4. Cascade classifier to combine many features efficiently

A. Haar like features

Haar like features are used to detect variation in the black and light portion of the image. This computation forms a single rectangle around the detected face. Based on the color shade near nose or forehead a contour is formed. Some commonly used Haar features are:

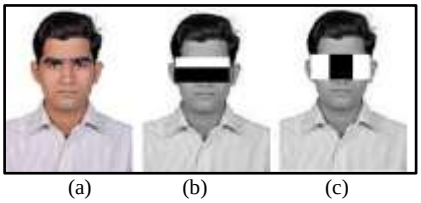
1. Two rectangle feature.
2. Three rectangle feature.
3. Four rectangle feature.



(a) (b) (c)

Fig.12. (a) Two Rectangle Feature, (b) Three Rectangle Feature, (c) Four Rectangle Feature

The value of two rectangle feature is the difference between the sums of the pixels within two rectangle regions as sown in Fig 12. In three rectangles, the value is center rectangle subtracted by the addition of the two surrounding rectangles. Whereas four rectangle features computes the difference between the diagonal pairs of the rectangles [20].Fig13 shows two features of Haar.



(a) (b) (c)

Fig. 13. (a) is the original unaltered image, (b) shows the feature extraction using the difference in the intensity between eyes and the upper Cheeks as the eye region is anytime more dark. (c) Compares the intensity in the eye region and across the bridge of the nose

Haar-like feature can be calculated with the following equation[13]:

Feature = $\sum$ *ie{1..N}wi.RecSum (x, y,w,h)* (1)

Where *RecSum (x,y,w,h)* is the summation of intensity in any given upright or rotated rectangle enclosed in a detection window and x,y,w,h are for coordinates , dimensions , and rotation of that rectangle, respectively. Haar Wavelets represented as box classifier which is used to extract face features by using integral image which is described in the next section.

B. Integral Images

They are also known as summed area tables. Integral image is used to felicitate quick feature detection. The meaning of integral image is the outline of the pixel values in the original images.

The integral image at location (x, y) contains the sum of the pixels above and to the left of (x, y) inclusive

$$II(x, y) = \sum_{x' \le x, y' \le y} i(x, y)$$

(2)

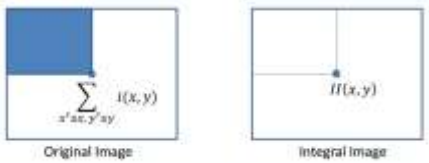


Fig 14 Demonstrating the concept of integral image

As can be seen from Fig 14, each location of x and y in the integral image is the sum of pixel values in above and left location of x and y [21].

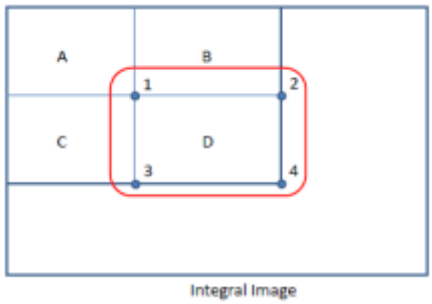


Fig 15 Finding the sum of the shaded rectangular area

For instance in Fig 15, let 1,2,3,4 be the values of the integral image at the corner of a rectangle, next the sum of original image values within the rectangle can be computed as the below equation and only 3 additions are required for any size of the rectangle.

SUM = 4-2-3+1

Value = $\sum$ (Pixels in White area) – $\sum$ (Pixels in Black area)

(3)

C. Adaboost machine learning method

It uses an important concept of Bagging that is procedure for combining different classifiers constructed using the same data set.It is an acronym for bootstrap aggregating, a motivation of combining classifiers is to improve an unstable classifier and an unstable classifier is one where a small change in the learning set/classification parameters produces a large change in the classifier.

AdaBoost algorithm helps to select small features from the face that facilitates fast and easy computation [18].Unlike other methods, AdaBoost algorithm gives desired region of the object discarding unnecessary background. The working model can in interpreted by using neural networks [18].

- Given image is in the form (x_1, y_1)……. (x_n, y_n)
- y_i =0,1 for negative and positive examples.
- Initialize the weights $w_{1,i}$ =$\frac{1}{2m}$, $\frac{1}{2l}$ for y_i =0,1 respectively, where m and l are number of positives and negatives respectively.
- For t=1,..,T:
  1. Normalize the weights,
     $w_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,i}}$
     (4)
     $w_t$is the probability distribution
  2. For each feature j, train a classifier h_j which is restricted to use a single feature.The error is evaluated with respect to w_t, E_t
     $=\sum_i w_i |h_j(x_i), y_i|$
     (5)
  3. Choose the classifier h_t with lowest error E_t
  4. Update the weights
     $w_{t+1,i} = w_{t,i} B_t^{1-e_i}$
     (6)

Where e_i =0 if examples x_i is classified correctly.

e_i =1 otherwise

And $B_t = \frac{e_t}{1-e_t}$

- The final strong classifier is:

$$h(x) = 1 \quad \sum_{t-=1}^{T} a_t h_t(x) \geq \frac{1}{2}\sum_{t=1}^{T} a_t \quad (7)$$

where $a_t = \log\frac{1}{Bt}$

AdaBoost learning process is fast and gives more number of desired data. This data can be classified into classifier. A classifier contains small features the face. It is commonly employed for pattern detection.This method has high accuracy and detection speed with about 1% false detection [18] but requires more time to train.

D. Cascade classifier

The Viola and Jones face detection algorithm eliminates face candidates quickly using a cascade of stages. The cascade eliminates candidates by making stricter requirements in each stage with later stages being much more difficult for a candidate to pass. Candidates exit the cascade if they pass all stages or fail any stage. A face is detected if a candidate passes all stages. This process is shown in Fig. 16.

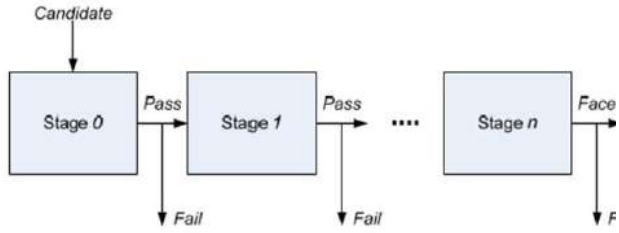


Fig 16.Cascade of stages. Candidate must pass all stages in the cascade to be concluded as a face.

VI. CONCLUSION

In this paper, we have discussed about the commonly used face detection methods. Each of these methods signify the importance and utility for different applications. Since these methods are progressive, more and more advancements are made every day to achieve accurate and true face detection. For applications such as employee details, member details and criminal record uses the frontal views of the face. Hence feature based and knowledge based methods are used. For identifying faces in videos or images those are occluded and oriented other than frontal view, template matching network and appearance method are used.

A rapid approach using machine learning and haar-like features are used recently for fast detection and reduce the likelihood of computing huge amount of data. This approach was designed by viola and jones to increase computational efficiency despite of having false face detection.

REFERENCES

[1] Hsu, Rein-Lien, Mohamed Abdel-Mottaleb, and Anil K. Jain. "Face detection in color images." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.5 (2002): 696-706.

[2] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, *From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell*. 23 (6) (2001) 643–660.

[3] Mayank Chauha and Mukesh Sakle. "*Study & Analysis of Different Face Detection Techniques*." International Journal of Computer Science and Information Technologies, Vol. 5 (2), 2014, 1615-1618.

[4] G. Yang and T. S. Huang, "Human Face Detection in Complex Background," Pattern Recognition, vol. 27, no. 1, pp. 53-63, 1994.

[5] T.K. Leung, M.C. Burl, and P. Perona, "Finding Faces in Cluttered Scenes Using Random Labeled Graph Matching," Proc. Fifth IEEE Int'l Conf. Computer Vision, pp. 637-644, 1995.

[6] K.C. Yow and R. Cipolla, "Feature-Based Human Face Detection," Image and Vision Computing, vol. 15, no. 9, pp. 713-735, 1997.

[7] J. Yang and A. Waibel, "A Real-Time Face Tracker," Proc. Third Workshop Applications of Computer Vision, pp. 142-147, 1996.

[8] S. McKenna, S. Gong, and Y. Raja, "Modelling Facial Colour and Identity with Gaussian Mixtures," Pattern Recognition, vol. 31, no. 12, pp. 1883-1892, 1998

[9] R. Kjeldsen and J. Kender, "Finding Skin in Color Images," Proc. Second Int'l Conf. Automatic Face and Gesture Recognition, pp. 312- 317, 1996.

[10] I. Craw, D. Tock, and A. Bennett, "Finding Face Features," Proc. Second European Conf. Computer Vision, pp. 92-96, 1992

[11] A. Lanitis, C.J. Taylor, and T.F. Cootes, "An Automatic Face Identification System Using Flexible Appearance Models," Image and Vision Computing, vol. 13, no. 5, pp. 393-401, 1995.

[12] H. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 1, pp. 23-38, Jan. 1998.

[13] Sharifara, Ali, et al. "A general review of human face detection including a study of neural networks and Haar feature-based cascade classifier in face detection." *Biometrics and Security Technologies (ISBAST), 2014 International Symposium on*. IEEE, 2014.

[14] Zhengming Li; Lijie Xue; Fei Tan, "Face detection in complex background based on skin color features and improved AdaBoost algorithms," Progress in Informatics and Computing (PIC), 2010 IEEE International Conference on , vol.2, no., pp.723,727, 10-12 Dec. 2010.

[15] Campadelli, Paola, Raffaella Lanzarotti, and Chiara Savazzi. "A feature-based face recognition system." *Image Analysis*

and Processing, 2003. Proceedings. 12th International Conference on. IEEE, 2003.

[16] Yang, Ming-Hsuan, David J. Kriegman, and Narendra Ahuja. "Detecting faces in images: A survey." *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.1 (2002): 34-58.

[17] Xiaowei Zhao, Xiujuan Chai ,"Context Constrained Facial Landmark Localization Based on Discontinuous Haar-like Feature" International Conference on Computer Vision (ICCV2013),2013.

[18] Paul Viola, Micheal Jones, "Rapid object detection using a Boosted Cascade of Simple features" CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2001

[19] Lienhart and J. Maydt. An Extended Set of Haar-like Features for Rapid Object Detection. IEEE ICIP 2002

[20] Monali Chaudhari, Gauresh Vanjare, Dhairya Thakkar, Malay Shah and Amit Kadam," Intelligent Surveillance and Security System", Vol. 3, Issue 3, March 2015, pp. 2291-2299.

[21] Cordiner, A.; Ogunbona, P.; Wanqing Li, "Face detection using generalised integral image features," Image Processing (ICIP), 2009 16th IEEE International Conference on , vol., no., pp.1229,1232, 7- 10 Nov. 2009.

[22] Ms.Drashti H. Bhatt , Mr.Kirit R. Rathod , Mr.Shardul J. Agravat. Article: A Study of Local Binary Pattern Method for Facial Expression Detection. International Journal of Computer Trends and Technology (IJCTT) 7(3):151-153, January 2014. Published by Seventh Sense Research Group.

BIOGRAPHY

Mrs. MonaliChaudhariis an assistant professor at Electronics and Telecommunication Department of VESIT, Mumbai University, with 9 years of teaching experience. She has done Masters in Engineering from Mumbai University. Her area of interest includes Image Processing and Wireless Technology.

Dr. Shanta Sondur is a professor at Information and Technology Department of VESIT, Mumbai University, with 15 years of teaching experience. She has done Ph.D from I.I.T, Bombay. Her area of interest includes Instrumentation &control and Image Processing.

Gauresh Vanjare is currently pursuing B.E. (Final Year) in Electronics and Telecommunication (Mumbai University) from V.E.S. Institute of Technology, Mumbai. Council member of ISTE-VESIT.His area of interest includes Image Processing, Artificial Intelligence and Embedded Systems.