

Performance Analyses and Comparison of Eye Detection Techniques

Vijayalaxmi¹ and D. Elizabeth Rani²

¹Vignan Institute of Technology & Science
Vignan Hills, Nalgonda District, Hyderabad, Andhra Pradesh, India

²Gitam Institute of Technology,
Gitam University, Vishakapatnam, Andhra Pradesh, India

¹laxmi81181@gmail.com

Abstract

A robust and accurate real time eye tracking system has been a challenging task for many computer vision applications. Different researchers working world wide have tried various approaches to solve this problem. Although many different algorithms exist to perform eye detection, each has its own weaknesses and strengths. But so far no system / technique exists which has shown satisfactory results in all circumstances. This research work is a comparative study on the performances of algorithms – Template Matching, Skin Segmentation, Artificial Neural Network and Haar Cascade Classifier for eye recognition. All the algorithms are developed on OpenCV platform and tested on images from Mathworks Video, GTAV, Face Expression and VITS database in the laboratory. The comparison is done based on the success rate i.e. total number of images with eyes detected to the total number of input images. The comparison results show that Haar Cascade Classifier has satisfactory results on images under different conditions such as tilted head position, closed eyes, occluded face, etc.,. The purpose of this research work is to develop a Non-intrusive Driver's Drowsiness detection system based on eye blink rate for preventing accidents on road.

Keywords: Haar, Neural Network, OpenCV, Skin, Template

1. Introduction

The motivation behind writing this paper is to compare developed eye detection algorithm for efficient eye tracking of a person from video image sequences. The efficient method can be used to study different eye-related behaviour analyses, e.g., fatigue test for drivers, sleep driving, physical-eye related diseases and lie detecting process. Compared to eye-blink detection by using some head-mounted devices [15] and some commercial eye trackers, eye-blink detection from video images may not be as accurate as them. However, this is usually compensated by greater ease of use, non-invasiveness and much lower cost. This paper may help the researchers to know about different eye detection and tracking algorithms which will ease the work. The eye tracking can run in real-time, without any additional hardware (like IR illumination for example) and be capable of operating under varying indoor conditions (typical office environment).

In the past few years lot of research work has been published in the field of eye detection so far. Various techniques have been proposed using template matching, IR based approaches, feature based approach, Skin Segmentation method, Hough transform method, Eigen space method or combination of these for eye detection.

Infrared lighting to capture the physiological properties of eyes and an appearance-base model to represent the eye patterns are the most common approach to achieve eye detection

in

Real-time [18, 21, 1,4]. In the appearance-base approach, eyes are detected based on the intensity distribution of the eyes. This method requires a significant number of training data to enumerate all possible appearances of eyes *i.e.*, representing the eyes of different subjects, under different face orientations, and different illumination conditions. Vezhnevets, *et al.*, [17] focus on several landmark points (eye corners, iris border points), from which the approximate eyelid contours are estimated. The iris centre and radius is detected by looking for a circle separating dark iris and bright sclera. The upper eyelid points are found based on the observation that eye border pixels are significantly darker than surrounding skin and sclera. In Skin Segmentation method, the detection of the skin region is very important to segment eyes from the rest of the image.

Various techniques have been adopted for eye detection which includes wavelets, principal component analysis, fuzzy logic, evolutionary computation and hidden markov models and template matching. Huang and Wechsler [7], used optimal wavelet packets for eye representation and radial basis functions for classification of eye and non-eye regions. Talmi, *et al.*, and Pentland, *et al.*, [8, 2] made use of very few characteristics of human eyes in Principal Component Analysis for eye detection. Torii, A. and Imiya A [16], used Karhunen-Loeve-Transformation to represent the major characteristics of human eyes and are stored as reference patterns for the localization of human eyes in video images.

Hjelms and Wroldsen [6] used combination of Gabor Filters and PCA for eye detection. Li, *et al.*, [19] constructed a template using fuzzy logic to judge the similarity between the input image and eye template. In the template, the eyelid is constructed by a region of adjacent segments along the piecewise boundary. Each segment in the fuzzy template is filled with the darkest intensity value within this segment. This method is highly robust to different illumination and resolution for various input images. The algorithm uses a neural network to scan an input window of pixels across the image, where each gray value in the input window serves as an input for the neural network. The neural network is then trained to give a high response when the input window is centered on the eye. After scanning the entire image, the position with the highest response gives the centre position of the eye in the image.

2. Proposed Methods

The purpose of this research work is to compare and analyze performance of different eye detection techniques which benefits computer vision applications like biometric system for security purpose, driver's drowsiness detection system, home applications and Human Computer Interaction (HCI). Four algorithms for eye detection developed on OpenCV platform and tested in the laboratory on different images from Mathworks Video, GTAV, Face Expression and VITS databases. The four algorithms are: Template matching, Skin Segmentation, Neural Network and Haar Cascade Classifier. The performance analysis of these algorithms is measured by testing them on different types of images available from above mentioned databases. Section 3 discusses about four proposed methods, Section 4 shows the results of all algorithms tested on different images, Comparison of proposed algorithms is done in Section 5 and Conclusions a drawn in Section 6.

2.1 Template Matching Method

As eyes is one of the important features of human face, the success of facial analysis and face recognition lies in detection of eyes. This method is an efficient approach to achieve fast and accurate eye detection in still gray level images with unconstrained background.

Template matching is a method in digital image processing for finding template in an original image .The process of template matching involves moving the template from

point to point in an image to evaluate the similarity of the counterpart using normal cross correlation function. Template is part of an image which is also known as mask or kernel.

Before processing the image, the captured image is preprocessed i.e enhanced to improve the quality of an image. Homomorphic filtering is a method to develop a frequency domain procedure for improving the appearance of an image by simultaneous gray level range compression and contrast enhancement. The enhancement using Homomorphic filtering is summarized in equation 1.

$$\mathbf{f(x,y)} \rightarrow \ln \rightarrow \text{FFT}(\mathbf{u,v}) \rightarrow \mathbf{H(u,v)} \rightarrow [\text{FFT}(\mathbf{u,v})]^{-1} \rightarrow \exp[\mathbf{f(x,y)}] \rightarrow \mathbf{g(x,y)} \text{-----}$$

---(1)

K-Mean Clustering algorithm is applied to remove the background. To separate object from background, image segmentation is done with a threshold value of 128. The low grey level pixel representing the background is set to 255, the intermediate representing the skin is set to 128, and the high grey level pixels representing both the features and other dark pixels of the image is set to 0. After thresholding binary image is obtained with white background and black object.

Template Matching is an efficient method for detecting eyes in still gray level images. The template of eye is used as a robust cue to detect eyes in an input image.

2.2 Skin Segmentation Method

Head position, background and intensity of the image always decrease the performance of eye detection system. It is observed that eye region in images are characterized by low illumination and high contrast as compared to other parts of the face. In Skin Segmentation method, the face is first extracted from the background based on Skin range value. The face is extracted using HSV colour model, while the eyes are detected using LAB colour space thus eliminating unwanted areas using Region of Interest (ROI) and Labelling process. The skin range in an image is identified using HSV colour space due to its advantage of describing pure colour in the image and not influenced by the intensity. Lab colour model is used to improve the colour representation as each colour is designated using specific 'a' and 'b' values and L yields brightness.

In a video simple search techniques will not help to identify the eyes among the other structural components of the face such as beard, moustache, *etc.* The problem is further complicated with live images, such as video, as the time available to search in each image frame is restricted.

The method proposed involves skin segmentation to eliminate background components followed by eye detection. Eliminating the skin region helps in determining the approximate eye position by eliminating a large number of false eye candidates. The HSV provides colour information similar to humans think of colours. "Hue" describes the basic pure colour of the image and "saturation" gives the manner by which this pure colour (hue) is diluted by white light, and "Value" provides an achromatic notion of the intensity of the colour. As proposed in the paper [11] the first two parameters, H and S will provide discriminating information regarding skin. As reported [11] the value of H shall be between 0.01 and 0.1. i.e

$$0.01 < \mathbf{H} < 0.1 \rightarrow \text{skin} \text{-----} (2)$$

It may be noted that, this does not remove all the skin in the image but small portions of it remains in the image. As YCbCr colour space for any image is considered as an advantage. This left out skin areas, are effectively removed using YCbCr colour space. After experimenting on various images with different threshold values, the best results were found by using the adaptive histogram technique. Based on Cb and Cr values, the range for skin is identified using values shown in equation 3.

$$140 < \mathbf{Cb} < 195 \rightarrow \text{Skin and } 140 < \mathbf{Cr} < 165 \rightarrow \text{Skin} \text{-----} (3)$$

After skin range detection, region properties are applied to know maximum area indicating face in the given input image. The Lab colour model is applied to the extracted face using HSV which provides measurement of both hue and density changes. In the

proposed algorithm the Lab colour model is applied subsequently to HSV colour space, to get the exact eye region.

After obtaining L^*a^*b colour space images, the threshold value is selected using adaptive histogram technique 'b' image the obtained threshold value is 135. The L^*a^*b -image is converted into binary image using 135 as threshold. After extracting face from image, the search area for detecting the eyes is reduced by finding region of interest. In this method the region of interest is the region containing eyes. The region of interest is obtained by dividing the face in to four quadrants, as proposed [13], the experiment on different images shows that eyes are found in the second quadrant. The search for the eye will be limited to the Region of interest, this limited area made the search more efficient.

Labelling of a binary image is the operation of assigning a unique value to pixels belonging to the connected region. Connected components labelling scans the image and groups its pixels based on pixel connectivity *i.e.*, all pixels in a connected component share similar pixel intensity values and are in some way connected with each other. Once all groups have been determined, each pixel is labelled with a gray level or a colour (colour labelling) according to the component assignment. 8-pixel connectivity is employed on binary image in the proposed algorithm. The labelling process is applied twice in the algorithm. Initially, it is applied before obtaining ROI, to remove maximum unwanted areas to reduce the noise from the face. Later it is applied after ROI, to eliminate the minor unwanted areas.

2.3 Neural Network Method

Driving support systems, such as car navigation systems are becoming common and they support driver in several aspects. Non-intrusive method of detecting Fatigue and drowsiness based on eye-blink count and eye directed instruction control helps the driver to prevent from collision caused by drowsy driving. Eye detection and tracking under various conditions such as illumination, background, face alignment and facial expression makes the problem complex. In the proposed algorithm, first the Network is trained to reject the non-eye region based on images with features of eyes and the images with features of non-eye using Gabor filter and Support Vector Machines to reduce the dimension and classify efficiently. In the algorithm, first the face is segmented using L^*a^*b transform colour space, later eyes are detected using HSV and Neural Network approach.

Training SVM is the first step in proposed algorithm. A large data of eye and non-eye patterns are collected to train SVM. The images containing eyes of various sizes, orientations, positions, open and close, non-eye-pair images are cropped to train SVM. Selection of proper non-eye-pair images is very important to train SVM because performance of SVM is influenced by what kind of non-eye-pair images are used. In the initial stage of training SVM, the non-eye-pair images are similar to eye pair such as eyebrows, nostrils and other eye-pair like patches as eye pair. The performance of the method is measured in terms of sensitivity rate of 88.6%, specificity of 95.2% and accuracy of 89.2%.

2.4 Haar Cascade Classifier Method

In the past few years' lot of efforts have been made in the field of face and eye detection. Detecting faces using Viola and Jones is accurate and rapid. This technique is used to accurately detect face and eyes in images. But, this requires analyzing the entire image to find the location with the highest probability of containing the feature. The proposed algorithm is to locate and track face, eyes using Haar cascade classifier. The images of size 480x640 are captured through Logitech USB camera at a rate of 30fps. These images are processed to find and track face and eye. The simulation results shows that the algorithm takes less than 5ms time to locate and track face and eye.

The core basis for Haar classifier object detection is the Haar-like features. These features, rather than using the intensity values of a pixel, use the change in contrast values between adjacent rectangular groups of pixels. The contrast variances between the pixel groups are used to determine relative light and dark areas. Two or three adjacent groups with a relative contrast variance form a Haar-like feature. Haar features can easily be scaled by increasing or decreasing the size of the pixel group being examined. This allows features to be used to detect objects of various sizes [12].

The simple rectangular features of an image are calculated using an intermediate representation of an image, called the integral image [20]. The integral image is an array containing the sums of the pixels' intensity values located directly to the left of a pixel and directly above the pixel at location (x, y) inclusive. So if A[x,y] is the original image and AI[x,y] is the integral image then the integral image is using equation 4.

$$AI [x , y] = \sum_{x' \leq x , y' \leq y} A(x' , y') \text{-----} (4)$$

The rotated integral image is calculated by finding the sum of the pixels' intensity values that are located at a forty five degree angle to the left and above for the x value and below for the y value. So, if A[x,y] is the original image and AR[x,y] is the rotated integral image then the integral image is computed as shown in equation 5.

$$AR[x , y] = \sum_{x' \leq x , x' \leq x - y - y'} A (x' , y') \text{-----} (5)$$

3. Results

This Section discusses results of all the algorithms. Figure 1 shows the results of Template Matching method, first image is original image, second shows eye template, third image represents eye detected image and last image shows the failure case. However this method fails to detect the eyes when eyes are closed. The method cannot deal with moderate rotations, wearing glasses and partial face occlusions.

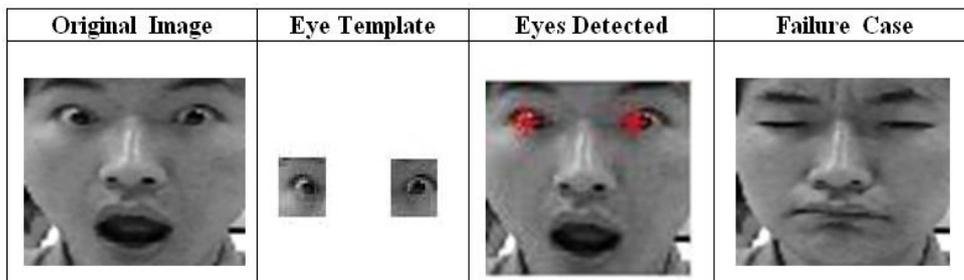


Figure 1. Template Matching Method Images showing Eyes Detected

Figure 2 shows the results of Skin Segmentation method, which shows original image, face extracted image and last image shows eyes marked with red dot. This method works satisfactorily on Mathworks Video database consisting of 126 images and eyes are detected on all 126 images. The success rate of this method is 100% and for other database images it shows an average of 65%. The major problem with this method is to know 'Hue' value for face segmentation and it varies from person to person.



Figure 2. Skin Segmentation Method Images Showing Eyes Detected

Figure 3 shows the results of Neural Network method which shows original images, sample of an image used for training SVM, third image shows eyes marked with red square and last image shows failure case. The images used for training SVM consists of nearly 50 eye and non-eye images, one set of sample are shown in Figure 3. The method fails to detect eye properly when the face is tilted by an angle. The efficiency of the method can be improved by training SVM with more set of positive and negative images.

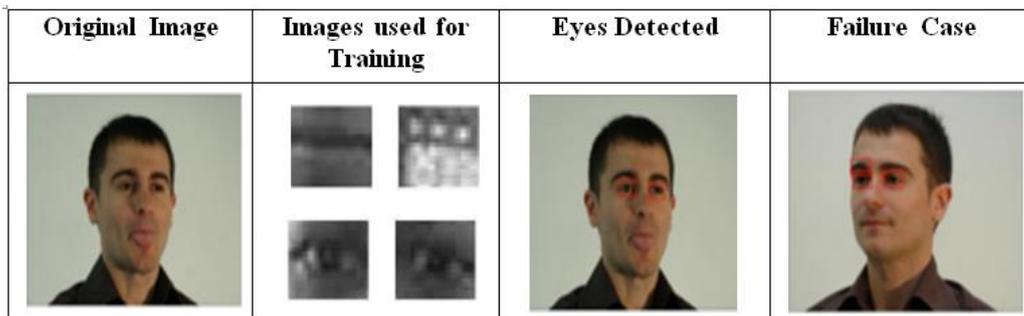


Figure 3. Neural Network Method Images showing Eyes Detected

Figure 4 shows the results of Haar Cascade Classifier, the algorithm is tested on live images captured through Logitech USB camera at a rate of 30fps. The Figure 4 shows original image, Face detected image and last images shows both face and eyes marked with square and rectangle with pink and green colours respectively. The method works satisfactorily on all types of images such as face occlusion, tilted head position, person wearing glasses and closed eyes.

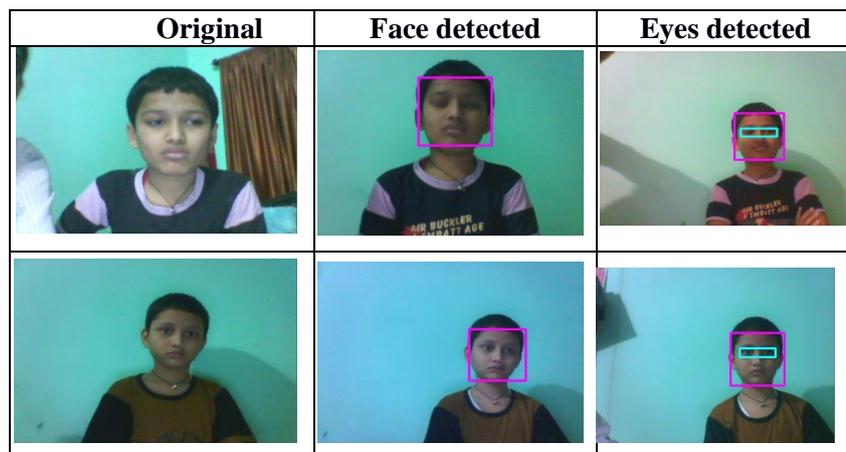


Figure 4. Haar Cascade Classifier Method Images showing Eyes Detected

4. Comparison of Methods

The algorithms Template Matching, Skin Segmentation, Neural Network and Haar Cascade Classifier for eye detection have been tested in the laboratory on various images from Mathworks Video, GTAV, Face Expression and VITS database and the results tabulated in Table 1. The success rate of all algorithms is calculated using equation 5 and is tabulated in Table 1.

$$\text{Success Rate} = (\text{Total number of images successfully detected as eyes}) / (\text{Total number of images used for testing}) * 100 \text{ ----- (5)}$$

Table 1. Comparison Results of four Proposed Methods

| Database Name | No. of images | Template Matching (%) | Skin Segmentation (%) | Neural Network (%) | Haar Cascade Classifier (%) |
|-----------------|---------------|-----------------------|-----------------------|--------------------|-----------------------------|
| Mathworks video | 126 | 65.87 | 100 | 75.4 | 89.6 |
| GTAV | 46 | 50.0 | 76.4 | 60.8 | 84.7 |
| Face Expression | 75 | 82.667 | 78.66 | 70.6 | 74.6 |
| VITS | 200 | 45.5 | 52.3 | 83.2 | 98.6 |

Table 1 comparison result shows that Template Matching method has poor performance when tested on all database images except Face Expression database. The major problem with Template Matching is creating a perfect template, which can search for proper match. The Skin Segmentation method has satisfactory results on Mathworks video, GTAV and Face Expression database images. The problem with Skin Segmentation method is change in skin colour causes error in segmenting face and eye from the background. The Neural Network approach results are satisfactory when tested on all database images. The major problem with this method is training of SVM which requires large number of eye and non-eye images for better performance and also takes more time for eye detection which is undesirable. The fourth method Haar Cascade Classifier performs better with all type of images, the method works better on live video stream and also on still images from all database images. The success rate of this method when tested on all database images is more than 85% except for Face Expression database images and time taken to track face and eye on live video is less than 5ms.

The performance analysis of all proposed methods is shown in terms of graphical representation for better comparison in below Figure 5.

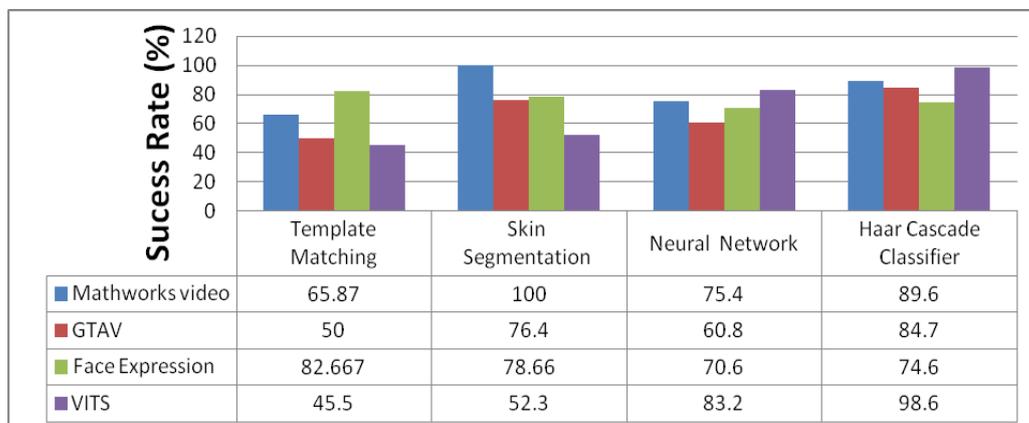


Figure 5. Graphical Representation of Success Rate of Proposed Algorithms

5. Conclusion

The proposed methods Template Matching, Skin Segmentation, Neural Network and Haar Cascade Classifier have been tested on different types of images from databases. From the comparison Table 1 Haar Cascade Classifier is better when compared to other methods. With Template Matching, it is required to create different templates to match the target, the method fails to give satisfactory results. Skin Segmentation when tested on images available at Mathworks yields 100% result but same algorithm shows poor performance on other database images. The reason is, person to person the range of 'Hue' varies, hence it is difficult to segment face from the background. Neural network approach gives better result on all the four database images but it needs more training and the time taken by the method to detect eyes is more. Haar Cascade Classifier method performs better on all database images. It even works better with live video images, images with and without glasses, images with face occlusion, images with tilted head position, *etc.*, Overall, when compared all methods in terms of success rate, time taken to detect eyes and in terms of different types of images Haar Cascade Classifier is a better choice for efficient eye detection. The authors continue to work in the same field and try to improve the efficiency of the proposed algorithm.

References

- [1] A. Haro, M. Flickner and I. Essa, "Detecting and Tracking Eyes By Using Their Physiological Properties, Dynamics, and Appearance", Proceedings IEEE CVPR 2000, Hilton Head Island, South Carolina, (2000) June.
- [2] A. Pentland, B. Moghaddam and T. Starner, "View-based and modular eigen spaces for face recognition", Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, Seattle, WA, (1994), pp. 84-91.
- [3] J. Song, Z. Chi and J. Liu "A Robust Eye Detection Method Using Combined Binary Edge and Intensity Information", Science Direct, Pattern Recognition, vol. 39, (2006), pp. 1110-1125.
- [4] C. Morimoto, D. Koons, A. Amir and M. Flickner, "Real-Time Detection of Eyes and Faces", Proceedings of 1998 Workshop on Perceptual User Interfaces, (1998) November, pp. 117-120, San Francisco, CA.
- [5] R. C. Gonzalez and R. E. Woods, "Digital Image Processing", 2nd edition PHI, Chapter 9, Morphological Image Processing.
- [6] E. Hjelms and J. M. Wroldsen, "Recognizing faces from the eyes only", Proceedings of the 11th Scandinavian Conference on Image Analysis, (1999).
- [7] J. Huang and H. Wechsler, "Eye Detection Using Optimal Wavelet Packets and Radial Basis Functions (RBFs)", International Journal of Pattern Recognition and Artificial Intelligence, vol. 13, no. 7, (1999).
- [8] K. Talmi and J. Liu, "Eye and Gaze Tracking for Visually Controlled Interactive Stereoscopic Displays", Signal Processing: Image Communication, vol. 14, (1999), pp. 799-810, Berlin, Germany.
- [9] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection", IEEE ICIP 2002, vol. 1, (2002) September, pp. 900-903.
- [10] P. Menezes, J. C. Barreto and J. Dias, "Face tracking based on Haar-like features and eigenfaces", 5th IFAC Symposium on Intelligent Autonomous Vehicles, Lisbon, Portugal, (2004) July 5-7.
- [11] "Mustafa UCAK face detection based on colour segmentation and eye region detection".
- [12] "Open Computer Vision Library Reference Manual", Intel Corporation, USA, (2001).
- [13] V. K. Banga and N. Sharma, "Development of a Drowsiness Warning system based on fuzzy logic", International Journal of computer applications, (2010) October.
- [14] S. A. Sirohey and A. Rosenfeld, "Eye detection in a face image using linear and nonlinear filters", Pattern Recognition, vol. 34, (2001), pp. 1367-1391.
- [15] C. Topal, O. N. Gerek and A. Doğan, "A head-mounted sensor-based eye tracking device: eye touch system," Proceedings of the 2008 Symposium on Eye tracking research & applications, (2008), pp. 87-90.
- [16] A. Torii and A. Imiya, "The Randomized Hough Transform Based Method for Great Circle Detection on Sphere", In the Proceedings of Pattern Recognition, (2007), pp. 1186-1192.
- [17] V. Vezhnevets and A. Degtiareva, "Robust and Accurate Eye Contour Extraction", Proc. Graphicon, (2003) September, pp. 81-84, Moscow, Russia.
- [18] X. Liu, F. Xu and K. Fujimura, "Real-Time Eye Detection and Tracking for Driver Observation Under Various Light Conditions", IEEE Intelligent Vehicle Symposium, Versailles, France, (2002) June 18-20.
- [19] Y. Li, X.-L. Qi and Y.-J. Wang, "Eye detection using fuzzy template matching and feature-parameter-based judgement", Pattern Recognition Letters, vol. 22, Issue 10, (2001) August, pp. 1111-1124.

- [20] Z. Zhang and J-S Zhang, "Driver fatigue detection based intelligent vehicle control", in Paper presented at the 18th IEEE International Conference on Pattern Recognition, (2006).
- [21] Z. Zhu, K. Fujimura and Q. Ji, "Real-Time Eye Detection and Tracking Under Various Light Conditions and Face Orientations", ACM SIGCHI Symposium on Eye Tracking Research & Applications, (2002) March 25-27, New Orleans, LA, USA.
- [22] http://gpstsc.upc.es/GTAV/R_ResearchAreas/UPCFaceDatabase/GTAVFaceDatabase.htm.

Authors



Vijayalaxmi, received her B.E (E&CE) degree from GNDEC, Visveswaraiah Technological University, Belgaum, and Karnataka, India in 2003. M.Tech (DE&CS) from JNT University, Hyderabad, Andhra Pradesh, India in 2008. She is pursuing her Doctorate from Gitam Institute of Technology, Gitam University, Vishakapatnam, Andhra Pradesh, India. Her area of research is Digital Image Processing. She has published 2 Journal papers, 3 International Conference papers and 2 National Conference papers. She received Best Paper award in August 2013 in IEEE International conference organized by Saveetha University, Chennai, India. She is the Life Member of ISTE and Member of IEEE.



D. Elizabeth Rani Completed Ph.D Electronics and Communication Engineering in 2003 from Andhra University, India, Masters in Communication Systems in 1984 from Bharathiar University, India, Bachelors in Electronics and Communication Engineering in 1982 from Madurai Kamaraj, University, India. Presently she is Head of the Department of Electronics and Instrumentation branch in Gitam University, Vishakapatnam, India. She has 25 technical publications/ conference papers to her credit. She is a member of MISTE, IETE, SEMCE (I). She has 26 years of teaching experience and 15 years of research experience. Her areas of interest are Signal processing, Communication systems and Image processing.

