

DRIVER ALERTNESS MONITORING USING MACHINE LEARNING AND VISION

- ¹ Dr O. Bhaskar, Associate Professor, Department of Computer Science and Engineering (Artificial Intelligence), Gates Institute of technology, Gooty, Anantapur, Andhra Pradesh, Email id: obhaskaru@gmail.com
- ² G.Anjali ,Student, Department of Computer Science and Engineering (Artificial Intelligence), Gates Institute of Technology-Gooty, Anantapur,Andhra Pradesh, Email id: gotluruanjali77@gmail.com
- ³ N.Lavanya, Student, Department of Computer Science and Engineering (Artificial Intelligence), Gates Institute of technology-Gooty, Anantapur, Andhra Pradesh, Email id: nnlavanya1@gmail.com
- ⁴ D.Prasanth Kumar, Student, Department of Computer Science and Engineering (Artificial Intelligence), Gates Institute of technology-Gooty, Anantapur, Andhra Pradesh,Email id: devarapalliprasanthkumar@gmail.com
- ⁵ C.Sai Kumar, Student, Department of Computer Science and Engineering (Artificial Intelligence),Gates Institute of technology-Gooty, Anantapur, Andhra Pradesh,Email id: saikumarchintha87@gmail.com
- ⁶ C. Tejhu, Student, Department of Computer Science and Engineering (Artificial Intelligence),Gates Institute of Technology - Gooty, Anantapur, Andhra Pradesh, Email id: thejutheju586@gmail.com

Abstract—Drowsy driving is one of the major causes of road accidents and death. Hence, detection of driver's fatigue and its indication is an active research area. Most of the conventional methods are either vehicle based, or behavioural based or physiological based. Few methods are intrusive and distract the driver, some require expensive sensors and data handling. Therefore, in this study, a low cost, real time driver's drowsiness detection system is developed with acceptable accuracy. In the developed system, a webcam records the video and driver's face is detected in each frame employing image processing techniques. Facial landmarks on the detected face are pointed and subsequently the eye aspect ratio, mouth opening ratio and nose length ratio are computed and depending on their values, drowsiness is detected based on developed adaptive thresholding. Machine learning algorithms have been implemented as well in an offline manner. A sensitivity of 95.58% and specificity of 100% has been achieved in Support Vector Machine based classification.

Keywords—*drowsiness detection, visual behaviour, eye aspect ratio, mouth opening ratio, nose length ratio.*

I. INTRODUCTION

Drowsy driving is one of the major causes of deaths occurring in road accidents. The truck drivers who drive for continuous long hours (especially at night), bus drivers of long distance route or overnight buses are more susceptible to this problem. Driver drowsiness is an overcast nightmare to passengers in every country. Every year, a large number of injuries and deaths occur due to fatigue related road accidents. Hence, detection of driver's fatigue and its indication is an active area of research due to its immense practical applicability.

The basic drowsiness detection system has three blocks/modules; acquisition system, processing system and warning system. Here, the video of the driver's frontal face is captured in acquisition system and transferred to the processing block where it is processed online to detect drowsiness. If drowsiness is detected, a warning or alarm is sent to the driver from the warning system.

Generally, the methods to detect drowsy drivers are classified in three types; vehicle based, behavioural based and physiological based. In vehicle based method, a number of metrics like steering wheel movement, accelerator or brake pattern, vehicle speed, lateral acceleration, deviations from lane position etc. are monitored continuously. Detection of any abnormal change in these values is considered as driver drowsiness. This is a nonintrusive measurement as the sensors are not attached on the driver. In behavioural based method [17], the visual behavior of the driver i.e., eye blinking, eye closing, yawn, head bending etc. are analyzed to detect drowsiness. This is also nonintrusive measurement as simple camera is used to detect these features. In physiological based method [8,9], the physiological signals like Electrocardiogram (ECG), Electrooculogram (EOG), Electroencephalogram (EEG), heartbeat, pulse rate etc. are monitored and from these metrics, drowsiness or fatigue level is detected. This is intrusive measurement as the sensors are attached on the driver which will distract the driver. Depending on the sensors used in the system, system cost as well as size will increase. However, inclusion of more parameters/features will increase the accuracy of the system to a certain extent. These factors motivate us to develop a low-cost, real time driver's drowsiness detection system with acceptable accuracy. Hence, we have proposed a webcam based system to detect driver's fatigue from the face image only using image processing and machine learning techniques to make the system low-cost as well as portable.

II. THE PROPOSED SYSTEM AND COMPUTATION OF PARAMETERS

A block diagram of the proposed driver drowsiness monitoring system has been depicted in Fig 1. At first, the video is recorded using a webcam. The camera will be positioned in front of the driver to capture the front face image. From the video, the frames are extracted to obtain 2-D images. Face is detected in the frames using histogram of oriented gradients (HOG) and linear support vector machine (SVM) for object detection [10]. After detecting the face, facial landmarks [11] like positions of eye, nose, and mouth are marked on the images. From the facial landmarks, eye aspect ratio, mouth opening ratio and position of the head are quantified and using these features and machine learning approach, a decision is obtained about the drowsiness of the driver. If drowsiness is detected, an alarm will be sent to the driver to alert him/her. The details of each block are discussed below.

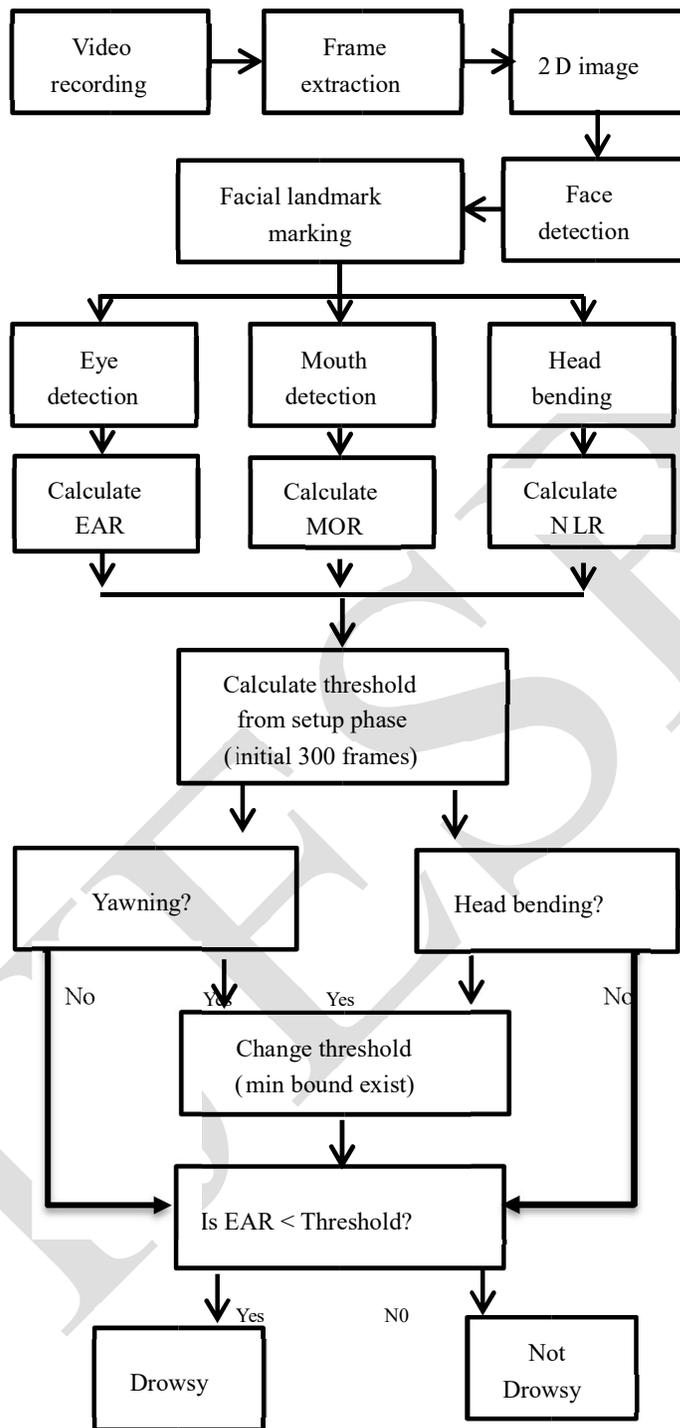


Fig. 1 The block diagram of the proposed drowsiness detection system

A. Data Acquisition

The video is recorded using webcam (Sony CMU-BR300) and the frames are extracted and processed in a laptop. After extracting the frames, image processing techniques are applied on these 2D images. Presently,

synthetic driver data has been generated. The volunteers are asked to look at the webcam with intermittent eye blinking, eye closing, yawning and head bending. The video is captured for 30 minutes duration.

B. Face Detection

After extracting the frames, first the human faces are detected. Numerous online face detection algorithms are there. In this study, histogram of oriented gradients (HOG) and linear SVM method [10] is used. In this method, positive samples of fixed window size are taken from the images and HOG descriptors are computed on them. Subsequently, negative samples (samples that do not contain the required object to be detected i.e., human face here) of same size are taken and HOG descriptors are calculated. Usually the number of negative samples is very greater than number of positive samples. After obtaining the features for both the classes, a linear SVM is trained for the classification task. To improve the accuracy of SVM, hard negative mining is used. In this method, after training, the classifier is tested on the labeled data and the false positive sample feature values are used again for training purpose. For the test image, the fixed size window is translated over the image and the classifier computes the output for each window location. Finally, the maximum value output is considered as the detected face and a bounding box is drawn around the face. This non-maximum suppression step removes the redundant and overlapping bounding boxes.

C. Facial Landmark marking

After detecting the face, the next task is to find the locations of different facial features like the corners of the eyes and mouth, the tip of the nose and so on. Prior to that, the face images should be normalized in order to reduce the effect of distance from the camera, non-uniform illumination and varying image resolution. Therefore, the face image is resized to a width of 500 pixels and converted to grayscale image. After image normalization, ensemble of regression trees [11] is used to estimate the landmark positions on face from a sparse subset of pixel intensities. In this method, the sum of square error loss is optimized using gradient boosting learning. Different priors are used to find different structures. Using this method, the boundary points of eyes, mouth and the central line of the nose are marked and the number of points for eye, mouth and nose are given in Table I. The facial landmarks are shown in Fig 2. The red points are the detected landmarks for further processing.

Table I: Facial landmark points

Parts	Landmark Points
Mouth	[13-24]
Right eye	[1-6]
Left eye	[7-12]
Nose	[25-28]

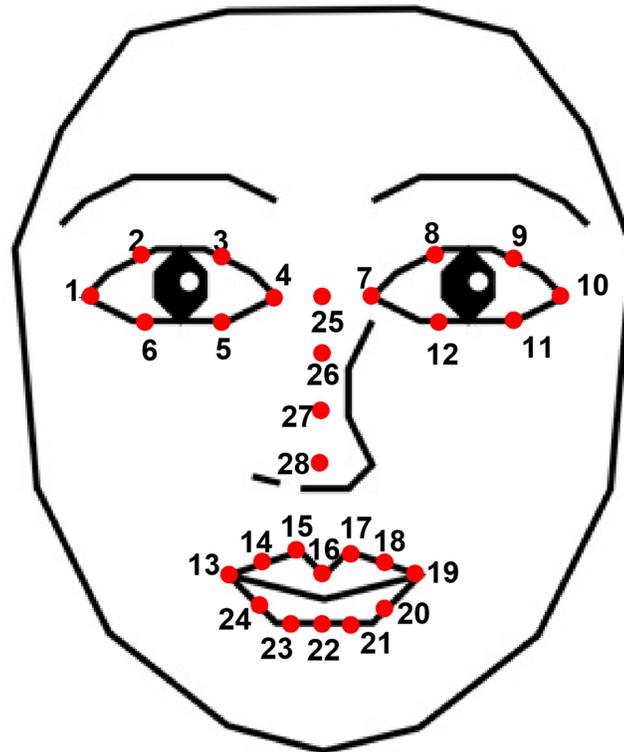


Fig. 2 The facial landmark points

D. Feature Extraction

After detecting the facial landmarks, the features are computed as described below.

Eye aspect ratio (EAR): From the eye corner points, the eye aspect ratio is calculated as the ratio of height and width of the eye as given by

$$EAR = \frac{(p_2 - p_6) + (p_3 - p_5)}{2(p_4 - p_1)}$$

where p_i represents point marked as i in facial landmark and d_{ij} is the distance between points marked as i and j . Therefore, when the eyes are fully open, EAR is high value and as the eyes are closed, EAR value goes towards zero. Thus, monotonically decreasing EAR values indicate gradually closing eyes and it's almost zero for completely closed eyes (eye blink). Consequently, EAR values indicate the drowsiness of the driver as eye blinks occur due to drowsiness.

Mouth opening ratio (MOR): Mouth opening ratio is a parameter to detect yawning during drowsiness. Similar to EAR, it is calculated as

$$MOR = \frac{(p_{15} - p_{23}) + (p_{16} - p_{22}) + (p_{17} - p_{21})}{3(p_{19} - p_{13})}$$

As defined, it increases rapidly when mouth opens due to yawning and remains at that high value for a while due to yawn (indicating that the mouth is open) and again decreases rapidly towards zero. As yawn is one of the characteristics of drowsiness, MOR gives a measure regarding driver drowsiness.

Head Bending: Due to drowsiness, usually driver’s head tilts (forward or backward) with respect to vertical axis. So, from the head bending angle, driver drowsiness can be detected. As the projected length of nose on the camera focal plane is proportional to this bending, it can be used as a measure of head bending. In normal condition, our nose makes an acute angle with respect to focal plane of the camera. This angle increases as the head moves vertically up and decreases on moving down. Therefore, the ratio of nose length to an average nose length while awake is a measure of head bending and if the value is greater or less than a particular range, it indicates head bending as well as drowsiness. From the facial landmarks, the nose length is calculated and it is defined as

$$NLR = \frac{\text{nose length}(p_{28}-p_{25})}{\text{average nose length}}$$

The average nose length is computed during the setup phase of the experiment as described in the next subsection.

E. Classification

After computing all the three features, the next task is to detect drowsiness in the extracted frames. In the beginning, adaptive thresholding is considered for classification. Later, machine learning algorithms are used to classify the data.

For computing the threshold values for each feature, it is assumed that initially the driver is in complete awake state. This is called setup phase. In the setup phase, the EAR values for first three hundred (for 10s at 30 fps) frames are recorded. Out of these three hundred initial frames containing face, average of 150 maximum values is considered as the hard threshold for EAR. The higher values are considered so that no eye closing instances will be present. If the test value is less than this threshold, then eye closing (i.e., drowsiness) is detected. As the size of eye can vary from person to person, this initial setup for each person will reduce this effect. Similarly, for calculating threshold of MOR, since the mouth may not be open to its maximum in initial frames (setup phase) so the threshold is taken experimentally from the observations. If the test value is greater than this threshold then yawn (i.e., drowsiness) is detected. Head bending feature is used to find the angle made by head with respect to vertical axis in terms of ratio of projected nose lengths. Normally, NLR has values from 0.9 to 1.1 for normal upright position of head and it increases or decreases when head bends down or up in the state of drowsiness. The average nose length is computed as the average of the nose lengths in the setup phase assuming that no head bending is there. After computing the threshold values, the system is used for testing. The system detects the drowsiness if in a test frame drowsiness is detected for at least one feature. To make this thresholding more realistic, the decision for each frame depends on the last 75 frames. If at least 70 frames (out of those 75) satisfy drowsiness conditions for at least one feature, then the system gives drowsiness detection indication and the alarm.

To make this thresholding adaptive, another single threshold value is computed which initially depends on EAR threshold value. The average of EAR values is computed as the average of 150 maximum values out of

300 frames in the setup phase. Then offset is determined heuristically and the threshold is obtained as offset subtracted from the average value. Driver safety is at risk when EAR is below this threshold. This EAR threshold value increases slightly with each yawning and head bending upto a certain limit. As each yawning and head bending is distributed over multiple frames, so yawning and head bending of consecutive frames are considered as single yawn and head bending and added once in the adaptive threshold. In a test frame, if EAR value is less than this adaptive threshold value, then drowsiness is detected and an alarm is given to the driver. Sometimes it may happen that when the head is too low due to bending, the system is unable to detect the face. In such situation, previous three frames are considered and if head bending was detected in those three frames, drowsiness alarm will be shown. Table II illustrates this calculation for determining the adaptive threshold.

Table II: Threshold for the computed parameters

EAR from setup phase (average of 150 maximum values out of 300 frames)	0.34
Threshold=EAR-offset	$0.34 - .045 = 0.295$
At Yawning, (MOR > 0.6)	Threshold=Threshold +0.002 *Max bound exist
At Head Bending, (NLR < 0.7 OR NLR > 1.2)	Threshold=Threshold +0.001 *Max bound exist

Apart from using thresholding, the machine learning algorithms are used to detect drowsiness as well. The EAR, MOR and NLR values are stored for the synthetic test data along with actual drowsiness annotation. Prior to classification, statistical analysis of the features has been done. At first, principal component analysis [12] is used to transform the feature space into an independent one. After transforming the feature values, student's t test is used to test whether the features are statistically significant for the two classes. As all the three features are statistically significant at 5% level of significance, all the three features are used for classification using Bayesian classifier [12], Fisher's linear discriminant analysis [12] and Support vector Machine [12].

III. RESULTS AND DISCUSSION

The proposed system has been developed and tested with the generated data. The webcam is connected with the laptop for further processing and classification of the video streaming in an online manner. Subsequently, the feature values are stored for statistical analysis and classification as well. One frame from the normal or awake state is shown in Fig. 3. The feature values for this frame are

0.35, 0.341, 1.003.



Fig. 3 Normal or awake state with facial landmarks



(a). (b), (c), (d)

Fig. 4 Drowsiness detected by the system due to (a) yawn, (b) eye closing, (c) head bending and (d) head too low.

Different drowsy conditions are displayed in Fig. 4. Figure 4(a) illustrates an example of drowsiness alert due to yawn and Fig. 4(b) illustrates an example of same due to eye closing. An example of head bending and

detecting it as drowsiness alert is shown in Fig. 4(c). Figure 4(d) depicts the condition when the head is too low due to bending and drowsiness is detected as described in previous section. Table III illustrates sample values of the parameters for different states.

Table III: Sample values of different parameters for different states

State	EAR	MOR	NLR
Normal	0.35	0.34	1.003
Yawning	0.22	0.77	0.76
Eye Closed	0.15	0.419	0.876
Head Bending	0.15	0.577	0.66

The developed system can also detect drowsiness with persons wearing spectacles as depicted in Fig. 5.



Fig. 5 Detection of eyes in presence of spectacles

The developed algorithm has been tested on INVEDRIFAC dataset [13]. It is a video and image database of faces of invehicle automotive drivers. The developed algorithm has been tested on 6 different driver videos. The performance of the same on this data is with acceptable accuracy. Also, the videos have different illumination conditions which indicate that the algorithm can perform well even at low illumination condition.

Subsequently statistical analysis and classification of the features into two classes have been explored as well. As the features are correlated, principal component analysis has been used to transform the feature space into an independent one. The independent features are statistically significant at 5% level of significance. Bayesian classifier, Fisher’s linear discriminant analysis (FLDA) and Support Vector Machine (SVM) with linear kernel have been used for classification. Presently, this has been done in offline manner on the stored data. Two-third data is used for training and one-third data is used for testing the algorithms. The classifier results are given in Table IV. Sensitivity is calculated as the ratio of correctly classifying drowsy states out of all actual drowsy states and specificity is computed as the ratio of correctly classifying awake states out of all actual awake states. Overall accuracy is computed as the correctly classified states out of all the frames. It is evident that the overall accuracy of FLDA and SVM is better than that of Bayesian whereas Bayesian gives best sensitivity of 97%. However, the specificity of Bayesian is quite low (56%). This may be due to the error in approximating the probability distributions. Due to low specificity, the alarm may ring when actually drowsiness is not there. This can be disturbing to the driver.

Table IV: Accuracy of different classifiers

Method	Sensitivity	Specificity	Overall Accuracy
Bayesian Classifier	0.973	0.561	0.854
FLDA	0.896	1	0.926
SVM	0.956	1	0.958

IV. CONCLUSION

In this paper, a low cost, real time driver drowsiness monitoring system has been proposed based on visual behavior and machine learning. Here, visual behavior features like eye aspect ratio, mouth opening ratio and nose length ratio are computed from the streaming video, captured by a webcam. An adaptive thresholding technique has been developed to detect driver drowsiness in real time. The developed system works accurately with the generated synthetic data. Subsequently, the feature values are stored and machine learning algorithms have been used for classification. Bayesian classifier, FLDA and SVM have been explored here. It has been observed that FLDA and SVM outperform Bayesian classifier. The sensitivity of FLDA and SVM is 0.896 and 0.956 respectively whereas the specificity is 1 for both. As FLDA and SVM give better accuracy, work will be carried out to implement them in the developed system to do the classification (i.e., drowsiness detection) online. Also, the system will be implemented in hardware to make it portable for car system and pilot study on drivers will be carried out to validate the developed system.

REFERENCES

- [1] W. L. Ou, M. H. Shih, C. W. Chang, X. H. Yu, C. P. Fan, "Intelligent Video-Based Drowsy Driver Detection System under Various Illuminations and Embedded Software Implementation", 2015 international Conf. on Consumer Electronics - Taiwan, 2015.
- [2] W. B. Horng, C. Y. Chen, Y. Chang, C. H. Fan, "Driver Fatigue Detection based on Eye Tracking and Dynamic Template Matching", IEEE International Conference on Networking, Sensing and Control, Taipei, Taiwan, March 21-23, 2004.
- [3] S. Singh, N. P. papanikolopoulos, "Monitoring Driver Fatigue using Facial Analysis Techniques", IEEE Conference on Intelligent Transportation System, pp 314-318.
- [4] B. Alshaqqa, A. S. Baquhaizel, M. E. A. Ouis, M. Boumehed, A. Ouamri, M. Keche, "Driver Drowsiness Detection System", IEEE International Workshop on Systems, Signal Processing and their Applications, 2013.
- [5] M. Karchani, A. Mazloumi, G. N. Saraji, A. Nahvi, K. S. Haghighi, B. M. Abadi, A. R. Foroshani, A. Niknezhad, "The Steps of Proposed Drowsiness Detection System Design based on Image Processing in Simulator Driving", International Research Journal of Applied and Basic Sciences, vol. 9(6), pp 878-887, 2015.

- [6] R. Ahmad, and J. N. Borole, "Drowsy Driver Identification Using Eye Blink Detection," *IJSET - International Journal of Computer Science and Information Technologies*, vol. 6, no. 1, pp. 270-274, Jan. 2015.
- [7] A. Abas, J. Mellor, and X. Chen, "Non-intrusive drowsiness detection by employing Support Vector Machine," 2014 20th International Conference on Automation and Computing (ICAC), Bedfordshire, UK, 2014, pp. 188193.
- [8] A. Sengupta, A. Dasgupta, A. Chaudhuri, A. George, A. Routray, R. Guha; "A Multimodal System for Assessing Alertness Levels Due to Cognitive Loading", *IEEE Trans. on Neural Systems and Rehabilitation Engg.*, vol. 25 (7), pp 1037-1046, 2017.
- [9] K. T. Chui, K. F. Tsang, H. R. Chi, B. W. K. Ling, and C. K. Wu, "An accurate ECG based transportation safety drowsiness detection scheme," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 4, pp. 14381452, Aug. 2016.
- [10] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection", *IEEE conf. on CVPR*, 2005.
- [11] V. Kazemi and J. Sullivan; "One millisecond face alignment with an ensemble of regression trees", *IEEE Conf. on Computer Vision and Pattern Recognition*, 23-28 June, 2014, Columbus, OH, USA.
- [12] Richard O. Duda, Peter E. Hart, David G. Stork, "Pattern Classification", Wiley student edition.
- [13] Dataset: <https://sites.google.com/site/invedrifac/>