TEST Engineering & Management

May-June 2020
ISSN: 0193-4120 Page No. 5346-5356

Facial Features and Eyes Closure Dependent Intelligent Method for Driver Drowsiness Identification: Multiple Kernel Learning Based Approach

[1]Abhishek Pratap Singh, [2]Sunilkumar S. Manvi
[1,2]School of Computing and Information Technology, REVA University,
Bengaluru, 560064, India
[1]apsingh95@outlook.com, [2]ssmanvi@reva.edu.in

Article Info
Volume 83
Page Number: 5346-5356
Publication Issue:
May-June 2020

Abstract

Fatigue, lack of concentration or drowsiness of drivers is one of the prominent reasons for roadside accidents world-wide. To address this problem, we have proposed a method to detect drowsiness of the driver based on complete eyes closure including the head position, & facial fatigue as conveyed by yawning and a drowsy face. Eye closure is calculated using landmarks on eyes. Detection of facial expressions of fatigue is divided into three phases. In the first phase, a Curvelet Transform is implemented to transform the input face image into four sub-band images which retain significant facial expression and eye information. The initial image is also sampled to acquire images of different sizes. Based on entropy assessment, each image is further divided into several blocks that are either categorized as informative or non-informative. In the second phase, using Discrete Cosine Transform the characteristics of high variance are selected in zigzag form. In the final phase, the Multiple Kernel Learning classifier is trained and tested to properly classify fatigue expressions of the driver. We combined eye aspect ratio (eye closure) with fatigue facial expressions (yawning and drowsiness) for drowsiness detection of the driver. It is observed that the proposed method has better classification accuracy and low false positive rate in real time detection than the existing methods.

Article History
Article Received:*19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *16 May 2020*

Keywords: *Entropy analysis, facial recognition, low-resolution, drowsiness detection, curvelet transform, eye aspect ratio*

1. Introduction

Sixteen Indians die every hour as a consequence of road accidents. According to the Global Road Safety Report 2015, a total of 141,526 individual died and about five lakh people were harmed as a result of road accidents in India. The National Highway Traffic Safety Administration (NHTSA) estimates that approximately 100,000 accidents happen each year in the United States due to driver drowsiness or fatigue. In 2013, due to driver drowsiness, the NHTSA recorded 72,000 accidents, 44,000 injuries, and 800 deaths.

Drowsiness is a state that decreases the level of attention of a person and thus affects the performance of the driver, increasing the associated risks. The advancement of computer vision-based driving assistance technologies have brought forward advanced transport sector solutions that help drivers gain a better understanding of the road environment and drive safely. Such techniques are also helpful for the identification of driver drowsiness. There are generally three types of drowsy driver detection techniques: vehicle-based, behaviour-based, and physiological-based methods (Ramzan et al. 2019).

In this paper, we consider a behaviour-based method for detecting the drowsiness and facial expressions of a driver. Many of the previous research and commercially available drowsiness detection methods (Bappaditya

May-June 2020
ISSN: 0193-4120 Page No. 5346-5356

Mandal et al 2017, Junli Xu et al 2018) focus primarily on eye closure and not on other facial expressions. Notable results can be achieved by analysing the closure of the eye, including facial motions related to sleepy faces and behavioral changes, which can be more reliable for the detection of drowsiness.

There are several issues in behavioral based drowsiness detection (with camera) methods like 1) noise in the image, 2) illumination effect, 3) image scaling and 4) low lighting condition of an image. These factors affect the performance of recognition systems.

In this paper, we propose a non-intrusive, cost-effective and efficient system for detecting drowsiness based on the driver's eye state and behavior captured through a camera. Decision is made depending on eyes closure, head position (slanting towards left or right) and facial expressions like yawning and drowsiness. Based on the state of an eye like fully closed, partially closed and open, drowsiness level of the driver is analysed. 68 facial landmarks have been considered with use of curvelet transform for image denoising. We applied entropy analysis to divide each image into a number of informative blocks or non-informative blocks. Discrete Cosine Transform (DCT) is applied to the blocks to extract high variance features from the image. Multiple kernel learning (MKL) technique is used for classification that has a high accuracy rate than existing methods.

Our contributions are as follows – c 1) Curvelet transformation is used to reduce noise that has a major benefit over other denoising methods. 2) Multiple kernel learning is used for high accuracy classification compared to other existing models. 3) The proposed method is implemented on the basis of eye closure combined with yawning and drowsy expressions to detect drowsiness in real time. 4) The proposed method is better in terms of accuracy compared to the existing methods.

The paper is divided into 6 sections. Section 2 presents the existing work. Section 3 describes the proposed method. Section 4 deals with the experimental set-up. Section 5 presents results and discussion, and Section 6 finally provides a conclusion

2. Related Work

Maryam Keyvanara et al (2018), suggested a real-time technique using face detection based on Haar wavelets. Eyes were detected using PCA and SVM. Sugur et al. (2013), detected the drowsiness according to the shape measurements of the eye. Authors proposed four main phases (face detection, eye region extraction, eye blink detection, and drowsiness detection.). They included the Eyelid's State Detecting value, where the classification of the eyelid status (closed or opened) was decided by setting a threshold on the measurements. Dalal et al (2005) proposed a non-intrusive approach for face detection that includes the Histogram Oriented Gradient (HOG) descriptor algorithm. The author also presented a comparative performance analysis of HOG descriptors against generalized Haar wavelets, PCA (Principal Component Analysis) SIFT descriptors and shape context descriptors on different datasets. The outcome of this approach is that linear Support Vector Machine (SVM) trained on HOG features classify the facial expression which has higher accuracy than other methods.

M. Omid Yeganeh et al (2011), presented a scheme using the fusion of eye closure and detection of yawning. Alongside yawning and eye behavior, Zhuoni Jie at al (2018) suggested that face touches can be used as a novel clue in automated drowsiness detection.

Lie Zhao et al (2018), suggested a method based on facial dynamic fusion information and a deep belief network (DBN). Sajid Ali Khan et al (2018) introduced a method that is capable of operating with multi-scale images but also capable of overcoming obstacles such as noise, lighting effects, image scaling and redundant data that affect the performance of facial expression recognition systems. Mohammad A. Haque et al (2016) implemented an efficient non-contact system for detecting non-localized physical fatigue from continuous muscle activity using facial videos acquired in a realistic natural lighting environment where participants were allowed to move their heads willingly, change their facial expression, and change their posture.

Choi et al (2013) used the technique of entropy analysis to assess the effectiveness of extracted features that help to reduce dimensionality. Similarly, Wang et al (2015) provided a minimal entropy-based atomic representation (MEEAR) framework for face recognition that works robustly to remove noise.

Discrete cosine transformation (DCT) is another strong transformation used for the selection of discrete coefficients. Wang et al (2016) introduced a novel technique to determine the dominant orientation of multi-orientation instances, called the inherent orientation of DCT, and a new DCT-inspired function transform (DIFT). It first calculates a unique intrinsic DCT orientation in each local region through the DCT matrix and rotates the region accordingly, then describes the rotated region with partial DCT matrix coefficients to produce an optimized low-dimensional descriptor. Liu Cao et al (2015) suggested an effective algorithm for fusing multi-focus pictures or videos using discrete cosine transformation (DCT) norms in a wireless visual sensor network (WVSN). Mehran et al (2014) provided the hashing of Discrete Cosine Transform (DCT) to create index structures for the facial descriptor. Keywords are played with hashes: an index is created and queried to find the most similar images to the query image. Common hash removal is used to enhance recovery efficiency and accuracy.

Some solutions have adopted wavelet transformation (WT) (Dipalee et al 2015) to achieve excellent scarcity for spatially located information such as corners, textures, and singularities. However, WT-based techniques only perform well when representing point singularities because they overlook the geometric characteristics of buildings and do not use edge regularity. Starck et al (2002) proposed a curvelet transformation in order to overcome the limitation. Monika et al. (2013) has suggested that the curvelet transformation gives better performance than the wavelet transformation method and also overcomes the wavelet transformation issues.

Recently, SVM algorithm (Simon Tong et al 2001) has been proposed for the task, but SVM is a linear classifier, thus this algorithm can only be used to classify linearly separable data. In order to classify nonlinearly separable data, this algorithm should be modified with kernel learning. Kernel learning's main idea is to transform the input space where the dataset cannot be linearly separated into a higher dimensional space where data becomes linearly separable. Many researchers are trying to develop more flexible kernel learning called Multiple Kernel Learning (MKL) (Mehmet G¨onen, 2011). Several recent studies have shown that the use of many kernels (Multi-Kernel Learning) in the SVM method has better performance than the use of a single kernel.

3. Proposed Method

Our method has four main components, namely, face and eye detection, Calculate Eye aspect ratio, facial expression classification, and drowsiness detection.

3.1 Face detection

The system starts with the process of face detection using Histogram of Oriented Gradient (HOG). HOG is feature descriptor that is used for object detection, based on intensity gradient distribution or edge detection feature directions. A detection window of specified pixels is passed over the image for each pixel within the cell in order to calculate gradients using Equation 1. Gradients include magnitude and orientation of pixels

gradient magnitude: $\text{gm} = \sqrt{gmx^2 + gmy^2}$ (1)

gradient direction $\theta = \arctan\frac{gmx}{gmy}$ (2)

3.2. Eye Closure Detection

In this section, we present a step-by-step process of the proposed eye closure detection technique. The various steps are: converting the grayscale, detecting the corner, calculating the midpoint, calculating the distance and calculating the eye status.

3.2.1. Grayscale Conversion

The first step in the calculation of eye closure is to convert a colored image into grayscale. The luminosity method is the most sophisticated method to convert a color image into grayscale. To human perception, it forms a weighted average. People are more green-sensitive than other colors, so green is weighted more than two other colors. Equation 2 gives the formula for luminosity

0.21 R + 0.72 G + 0.07 B (2)

a) Colored Image b) Grayscale Image

Figure 1: shows the conversion of a colored image into a grayscale image

3.2.2 Corner Detection

Corners are defined as an intersection of two edges. We use an eye closure detection algorithm that uses the two eye corner points and the reduced eyelid at one point. These points are identified with eye landmarks. This detector in the corner uses the fact that a corner is simply the intersection point between two edges. Figure 2 shows the corner points (landmarks) of an eye

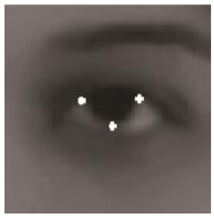


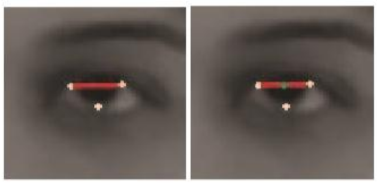
Figure 2: Corner points of eye

3.2.3 Midpoint Calculation

A midpoint is described as the middle or center point of a line segment. The next stage would be to locate a midpoint between the two upper corner points once all the required points have been identified. Let (x1,y1) be the coordinates of the upper-left corner and (x2,y2) the coordinates of the upper-right corner. A line section is taken between those two points. The midpoint of this line segment can be calculated using the formula in Equation 3.

$$\left(\frac{x1+x2}{2}, \frac{y1+y2}{2}\right) \quad (3)$$

Fig. 3 (a) and (b) show the segment line between the two endpoints and the midpoint line segment respectively.



(a) (b)

Figure 3: a) Line segment, b) Midpoint

3.2 4. Distance Calculation

The distance is a mathematical description of the distance between objects. We discover the distance from the lower eyelid point of the midpoint as a next move. In analytical geometry, using the distance formula provided by the Pythagorean theorem, the distance between two or more points is calculated. In Equation 4, the distance between two points (x1, y1) and (x2, y2) is indicated.

$$d = \sqrt{\Delta x^2 + \Delta y^2} = \sqrt{(x2 + x1)^2 + (y2 + y1)^2} \quad (4)$$

Fig. 4 shows the line segment which connects the midpoint to the lower eyelid point.

Figure 4: Line segment joining the midpoint to the lower eyelid point

3.2.5 Eye Aspect Ratio Determination

Finally, the eye state decision is made based on the distance 'd ' calculated in the previous step. If the distance is zero or close to zero, the eye state is classified as "closed" otherwise the eye state is identified as "open."

3.3 Fatigue Facial Expressions Detection

Facial expression recognition (FER) method includes three primary steps: pre-processing, extraction and classification of features.

3.3.1 Pre-Processing

Image pre-processing is important and necessary because the image captured in the real-time environment can have low contrast, noise, and background. The region of interest for facial expression recognition is face. In the proposed system, we used the HOG algorithm with a pre-trained model that gives 68 points (landmarks) of the face. Image contrast improvement is another preprocessing phase aimed at improving the amount of contrast of degraded images. Histogram equalization (HE) is the common contrast improvement that is commonly used to increase an image's dynamic range. While HE is suitable for enhancing the contrast image, it suffers from excessive contrast. In the proposed work, a technique of contrast limited adaptive histogram equalization (CLAHE) is used to solve this problem (Thamizharasi Ayyavoo et al, 2018). CLAHE is a classic technique initially suggested by Pizer et al (1987) for image improvement. The input image is first divided into separate sub-blocks in this method and then each block is enhanced.

3.3.2 Feature Extraction

Extraction/selection of features is an important step towards facial expression recognition. This step requires extracting from the face image a set of key parameters, which can be further used to classify and recognize different expressions. For recognition of expressions, all blocks may not contribute. Only the informative blocks need to be chosen. We used the entropy analysis to serve this objective. Entropy analysis is aimed at producing or selecting the informative blocks in an image. Some image sections or blocks are more informative than others. For example, the position of eyes or lips is more important for detecting the current feature or mood of a person, while chin, forehead, nose or cheeks are not that important. It is easier to recognize the more informative blocks by dividing the image into several blocks. For entropy values, we defined two threshold values, the upper (t2) and lower (t1). Blocks with an entropy value between these boundaries are considered informative, otherwise, they are not informative. If the entropy E of some blocks is less than the lower bound t1 or more than the upper bound t2, as described in Equation 5, it is considered informative.

$$\begin{cases} (informative), & t1 \leq E \leq t2 \\ (non-informative), & otherwise \end{cases} \quad (5)$$

The CVT is a multi-scale geometric wavelet transformation (WT) (Kun Su Yoo et al 2019) that can represent edge characteristics and curve singularities much more effectively than conventional WT. A trouser algorithm is implemented for sub-band decomposition. The algorithm is described in Equation 6.

$$f(x, y) = cJ(x, y) + \sum_{j=1}^{J} w_j(x, y), \quad (6)$$

Where f(x, y), cJ (x, y), and wj(x, y) are respectively an original image, low-pass filtered image, and high-pass filtered image. The cJ and wj represent the original image f detail information. The B3-spline filter decomposes the image into sub-band. In this progress of decomposition, the image information details shrink, and the image line becomes thicker as given in Equation 7 by function ' f. '

$$f \rightarrow (P0f, \Delta 1\, f, \Delta 2f, \ldots \ldots \ldots \ldots .. \Delta nf) \quad (7)$$

Where P0 is a low-pass filter and Δ1f, Δ2f, …, Δnf is a band pass filter. The image is then decomposed into four subbands, each containing details of the other frequencies. The partitioned squares Q(s, k1,k2) are given in Equation 8 to perform smooth partitioning.

$$Q(s, k_1, k_2) = \left[\frac{k_1}{2^s}, \frac{k_1+1}{2^s}\right] \times \left[\frac{k_2}{2^s}, \frac{k_2+1}{2^s}\right] \in Q_s \quad (8)$$

Where ‘s’ is scale.

May-June 2020
ISSN: 0193-4120 Page No. 5346-5356

Renormalization (TQ f)(x1,x2) is intended to normalize each square unit to [0, 1] [0, 1]. (Tq f), as described in Equation 9.

$$(TQ\ f)(x_1, x_2) = 2^8 f(2^8 x_1 - k_1, 2^8 x_2 - k_2) \qquad (9)$$

Where Tqf indicates the transportation operator acting on features ' f ' via gq provided in Equation 10.

$$g_Q = T^{-1}{}_Q h_Q \qquad (10)$$

where h_Q, T_Q^{-1} are partition and inverse deliver operators, respectively

After renormalization, radon transform (RDT) is performed and consists of three steps. First, the calculation is made of the two-dimensional quick Fourier transform (FFT2D), and this step produces a rectopolar, which is a radial line. Finally, RDT may be articulated as applying rectopolar 1D reverse FFTs (FFT1D−1) (Jean-Luc et al 2002). The limitation of RDT is that only the straight line can be used for the perfect reconstruction. Generally, an image has more curves than straight lines, and to reduce errors, an additional linearization work is needed. Ridgelet transforms (RLT) are a method used to apply 1D WT to radon regions. As in Equation 11, each standard square is analyzed:

$$\alpha Q, \lambda = < gQ, \rho\lambda > \qquad (11)$$

where α_Q, λ are the RLT coefficients.

A 50x50 pixel size image is acquired after CVT is deployed on the input facial picture. The fundamental features engaged in the method are the differentiation and average of coefficients of input information produced at distinct levels.

In the next step, we adopted DCT to the LL sub-band and chosen the elevated variance coefficients. DCT is applied to all informative facial areas (Gu-Min Jeong et al). The 2-D DCT for M×N image can be described as in Equation 12:

$$C(u,v) = \alpha(u)\,\alpha(v) \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y) \cos\left[\frac{\pi(2x+1)u}{2N}\right] cos\left[\frac{\pi(2x+1)v}{2N}\right] \qquad (12)$$

Where $0 \le u \le N$ & $0 \le v \le N$ and

$$a(u) = \begin{cases} \sqrt{\frac{1}{N}} & for\ u = 0 \\ \sqrt{\frac{2}{N}} & for\ u \neq 0 \end{cases} \qquad (13)$$

The image block row and column indices are represented here by x and y. Similarly, the DCT coefficient block two and column indices can be represented by u and v. In our study, the DCT is applied to informative regions and the coefficients are selected zigzag as shown in Fig. 5.

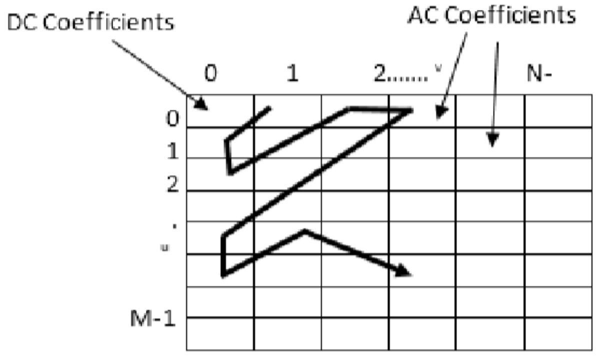


Figure 5: Applying DCT on Informative region

The first coefficient represents the proportional average of the M import N block and is known as the DC coefficient. The remaining coefficients are known as AC coefficients, indicating changes in intensity between image blocks. Equation 14 used to write all the DCT coefficients on the entire image after using DCT.

$$\sum_{w=0}^{MN-1} C(w) = \sum_{u=0}^{M-1}\sum_{v=0}^{N-1} C(u,v) \qquad (14)$$

CVT is applied to all facial inputs in our work. LL, i.e. the approximate band contains the most important and important information about the facial image. A 50x50 pixel size image is obtained after the deployment of DWT on the input facial image. The fundamental functions involved in the process are the differentiation and average of the input data coefficients produced at various levels. We implemented DCT to the LL sub-band in the next step and chosen elevated variance coefficients. The final feature vector is developed by combining chosen features based on DCT from informative face parts and sub-band approximation. The feature vector obtained provides a more comprehensive expression description and is robust as well. The last and important step after extraction/selection of the feature is to apply the appropriate classifier.

3.3.3 Expression Classification

Determining the appropriate kernel during the learning process is difficult. Therefore, many researchers are trying to develop more flexible kernel learning called Multiple Kernel Learning (MKL). Several recent studies have shown that the use of many kernels (Multi-Kernel Learning) in the SVM method has better performance than the use of a single kernel.

The main idea of multiple kernel learning is to combine several kernel functions into a single function, suppose given some kernel $K1\ (x, x')$, $K2\ (x, x')$, … , $KM(x, x')$ then the linear combination of the kernel can be written as in Equation 15.

$$K(x, x') = \sum_{m=1}^{M} d_m K_m(x, x') \qquad (15)$$

Subject to

May-June 2020
ISSN: 0193-4120 Page No. 5346-5356

$$d_m \geq 0, \sum_{m=1}^{M} d_m = 1$$

M is a number of the kernel function. If the optimal hyperplane of a single kernel is obtained by solving the following optimization problem (Equation 16).

$$min \frac{1}{2}\|w\|^2 + c\sum_{i=0}^{n} \varepsilon_i \qquad (16)$$

Subject to

$$y_i(x_i.w + b) \geq 1 - \xi_i \quad \forall i$$
$$\xi_i \geq 0 \quad \forall i$$

Where C and ξi have added a variable for penalty value to reduce the number of misclassification data. In MKL, where decision function of the form $(x)+ b = \sum f_{m.}(x) + b$ and f_m are associated with a kernel K_m.

3.4 Functioning of method

The functioning of the proposed method is shown in Fig. 6. The overall functioning of the method comprises of following components: Face Detection, Calculate Eye Aspect Ratio, Fatigue Facial Feature Classification, and Drowsiness Detection

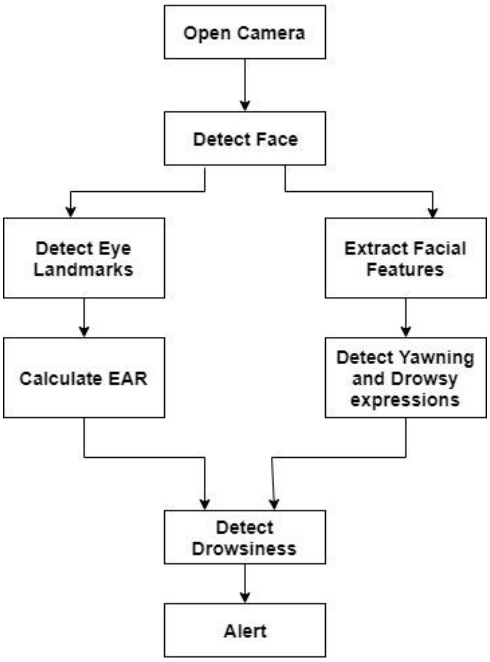


Figure 6: Methodology for the proposed system

The first step is to open the camera and detect a face in the frame using the face detection method. The next step is to obtain the landmarks on the face. The fundamental concept of this method is to find 68 specific points on the face, such as mouth corners, eyebrows, eyes, etc. In the C++ dlib library (based on HOG), it is a pre-trained detector that can locate 68 coordinates on any face. On the 300-W iBUG dataset, the predictor was trained.

Fig. 7 shows the sample facial landmark.

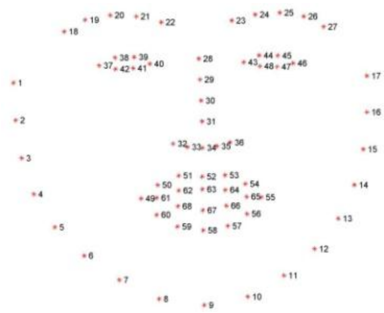


Figure 7: Facial landmarks

Table 1 shows the range of landmarks for individual facial features.

Table 1: Set of landmarks for important facial features

Features	Face landmarks
Full face	1-68
Jawline	1-17
Left Eyebrow	18-22
Right Eyebrow	23-27
Nose	28-36
Left eye	37-40
Right eye	43-46
Mouth outer points	49-60
Mouth inner points	61-68

To calculate eye aspect ratio, following steps are followed: a grayscale image of an eye is used to detect the corner of the eye, the midpoint of eye corner is calculated, calculate the distance of midpoint from the lower eyelid, Finally, the eye status decision is based on the distance from the lower eyelid.

For fatigue facial expression classification, The image is pre-processed to remove noise and improve image contrast, then the extraction technique of the function is used to acquire the vector. The extracted feature vector is compared with the trained model for the classification of expressions.

Detecting drowsiness is based on both eye aspect ratio and fatigue facial expressions. If more than 5 successive frames eyes are closed, the system will detect this state as drowsiness. And parallelly system monitors facial expressions of the driver. If the driver is feeling sleepy and closed partially eyes, then also system detects as drowsiness. Once drowsiness is detected, the proposed method triggers the alert system.

TEST Engineering & Management

May-June 2020
ISSN: 0193-4120 Page No. 5346-5356

4. Experimental Setup

In this section, we present tools for the implementation of our proposed framework, facial fatigue expression database, and experimental performance parameters.

4.1 Tools

For developing our proposed system we have used Anaconda environment and Python programming language. We performed all experiment on the computer with Intel i3 processor that has 4 GB RAM and 500 GB memory.

4.2 Database

We created our own database for facial expressions that contains 500 images of yawning, neutral and drowsy faces. These images are labeled in yawning, drowsiness and neutral classes. Fig. 8 illustrates about facial expression database.

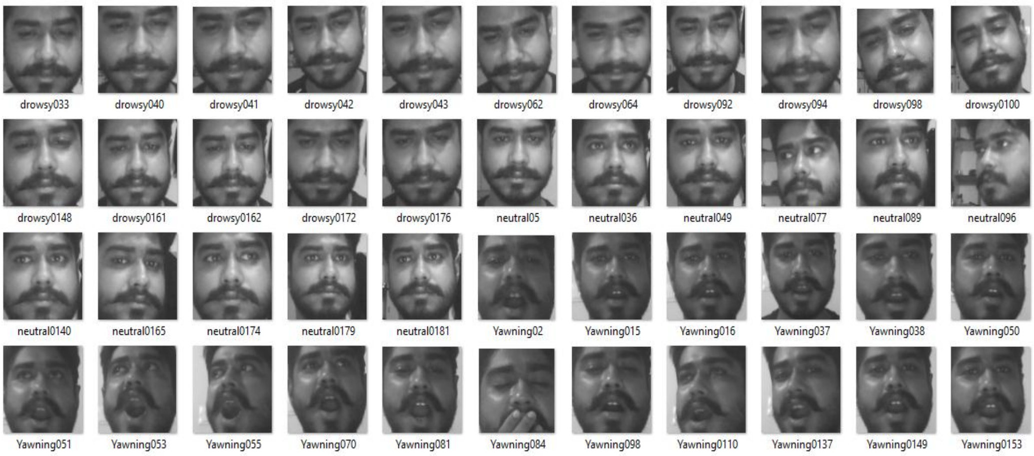


Figure 8: Face database

Throughout the image acquisition process, face images are cropped and converted into gray images. Later, these images are saved in the expression folder to make face databases for feature extraction task.

4.3 Performance Parameters

Parameters evaluated in the proposed work are a precision score, recall score, F1 score, and accuracy.

4.3.1 Precision Score

t is the ratio of correct positive predictions (TP) to the total predicted positives (TP + FP) as given in Equation 17, where TP is True Positives (correctly predicted event values) and FP is False Positives (incorrectly predicted event values. It is also called Positive predictive value.

$$\text{Precision} = TP/(TP + FP) \qquad (17)$$

4.3.2 Recall Score

It is defined as the ratio of correct positive predictions (TP) to the total positives examples (TP+FN) as given in Equation 18, where FN is False Negatives (incorrectly predicted no-event values).

$$\text{Recall} = TP/(TP+FN) \qquad (18)$$

4.3.3 F1 Score

F1 Score is a useful metric to compare two classifiers. F1 Score takes precision and recall into account. It is created by finding the harmonic mean of precision and recall (Equation 19).

$$F1 = 2 \times (\text{precision} \times \text{recall})/(\text{precision} + \text{recall}) \qquad (19)$$

4.3.4 Accuracy

Accuracy is the ratio of correct results to total results as provided in Equation 20, where TN is True Negatives (properly predicted values of no-event)

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \qquad (20)$$

5. Results and Discussion

In this section, we address the outcomes of our proposed system and comparison between the proposed method and existing work (Sajid Alik Khan et al, 2018) which is based on SVM and WT.

5.1 Face detection

The camera will initially open and identify faces in the frame using a dlib library which uses a pre-trained face detector based on a modification of Histogram Oriented Gradient + Linear SVM. If any person detected in the frame, the system detects the facial landmarks (68 facial landmarks on face) otherwise it will continue monitoring the face. Fig. 9 shows the facial landmarks on the face.

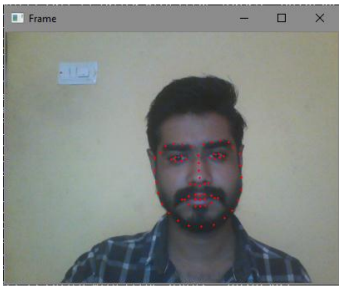


Figure 9: Facial landmarks detection

5.2 Calculate Eye Aspect Ratio (EAR)

After detecting the facial landmarks on face, the system detects eyes in the frame and calculate eye aspect ratio using midpoint value and distance from the lower eyelid. Fig. 10 shows the eye aspect ratio.

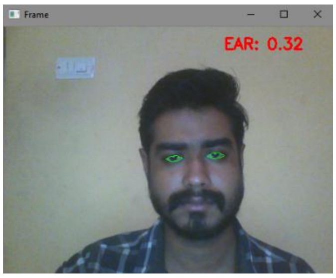
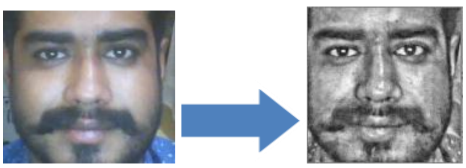


Figure 10: Eye aspect Ratio

5.3 Facial Feature Classification

5.3.1 Pre-Processing

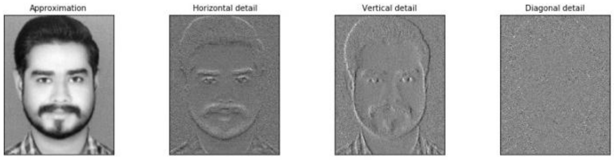
First, we applied pre-processing techniques to improve the image contrast and get the person's face information, like position, the mouth is opened or closed, whether the eyes are opened, closed, looking up and etc. Fig. 11 shows the images improved by the technique of CLAHE.



a) b)

Figure 11: Applying CLAHE a)original image, b)Enhanced image

After pre-processing, image is carried to curvelet transform to generate four sub-band pictures such as Approximation, Horizontal Details, Vertical Details, and Diagonal Details. Figure 12 demonstrates the outcome after implementation of curvelet transform on the image.



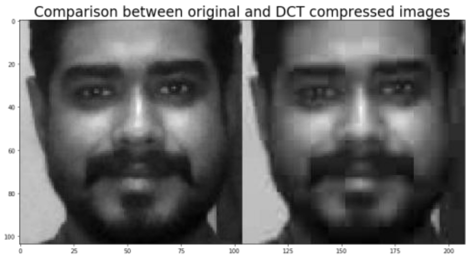
(a) (b) (c) (d)

Figure 12: Applying Curvelet Transform on image – a) Approximation b) Horizontal detail c) Vertical detail d) Diagonal detail

5.3.2 Feature extraction

Step 1: After CVT is applied, the image is divided into 8 x 8 blocks and the multi-scaled images are simultaneously passed through the entropy analysis phase to identify informative blocks.

Step 2: The DCT is used in a zigzag way after recognizing informative blocks to extract features from each block then a 64 size feature vector is being generated. Fig. 13 Shows the comparison between the initial image and the compressed DCT image.



a) b)

Figure 13: Comparison between the original image and DCT compressed image – a)original image b) DCT compressed image

5.3.3 Classification

In this phase, multiple kernel learning classifier is trained and tested on the extracted features for classification.

5.4 Detection of fatigue expressions

After applying feature extraction, we get a feature vector that is compared with the trained model and classify expressions in fatigue facial expression class such as yawning and drowsiness. The result of facial emotions is as follows.

5.4.1. Neutral State

If the eye aspect ratio is more than 0.30 and facial expressions are normal the system will show the neutral state of the driver. Fig. 14 shows the neutral state of the driver.

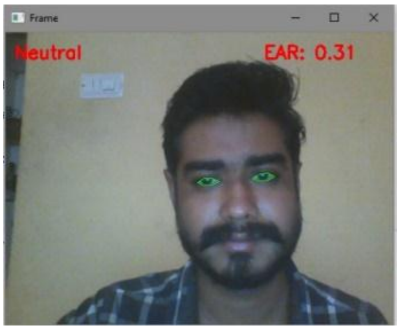


Figure 14: Neutral State

5.4.2. Yawning

If the driver is yawing, then system extracts facial expression and classify it as the yawning state. Fig. 15 show the yawning expression of the driver.

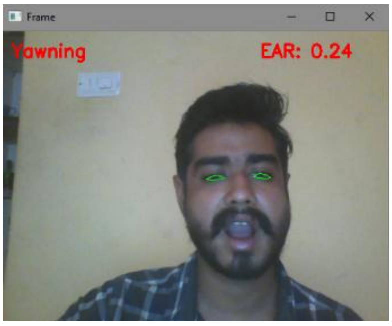


Figure 15: Yawning

5.4.3. Drowsiness detection

If the driver is trying to sleep and closing his/ her eyes partially below 0.25 eye aspect ratio, the system compares facial expressions with drowsiness class and show the result as the driver is drowsy. Fig. 16 presents the drowsiness state of the driver.

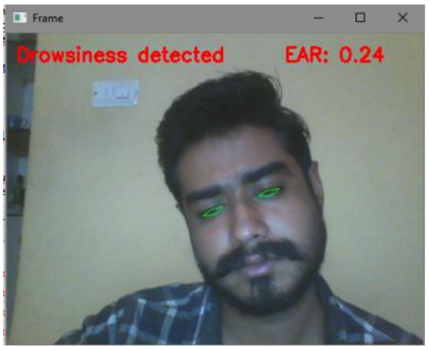


Figure 16: Drowsiness detection

5.5 Drowsiness Alert

If the driver closes his/her eye more than 5 consecutive frames, the system alert will be on. Fig. 17 shows the drowsiness alert system.

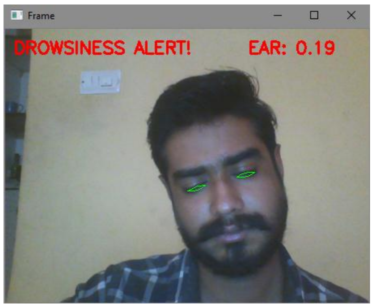
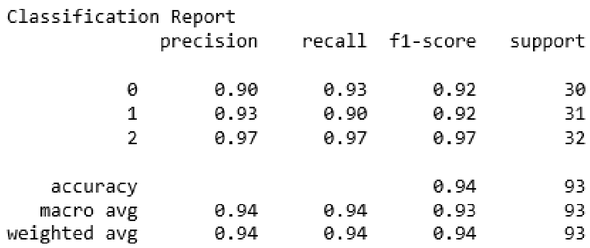


Figure 17: Drowsiness Alert

5.6 Performance Parameter Analysis

We compared SVM-WT with MKL-CVT (Proposed work) in terms of the classification report. The classification report of SVM-WT is shown in figure 18.



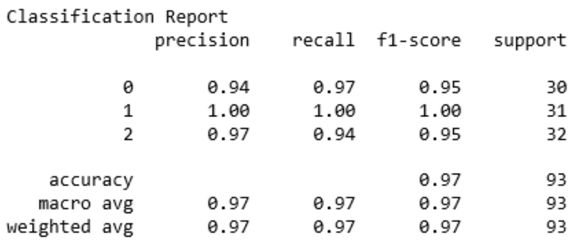
```
Classification Report
              precision    recall  f1-score   support

           0       0.90      0.93      0.92        30
           1       0.93      0.90      0.92        31
           2       0.97      0.97      0.97        32

    accuracy                           0.94        93
   macro avg       0.94      0.94      0.93        93
weighted avg       0.94      0.94      0.94        93
```

Figure 18: Classification report of SVM

Classification report of MKL-CVT is shown in figure 19.



```
Classification Report
              precision    recall  f1-score   support

           0       0.94      0.97      0.95        30
           1       1.00      1.00      1.00        31
           2       0.97      0.94      0.95        32

    accuracy                           0.97        93
   macro avg       0.97      0.97      0.97        93
weighted avg       0.97      0.97      0.97        93
```

Figure 19: Classification report of MKL

Table 2 shows the comparison between SVM+WT and MKL+CVT in terms of precision, recall, f1- score and accuracy.

Table 2: Comparison between SVM+WT and MKL+CVT

	Precision	Recall	F1-score	Support	Accuracy
SVM+WT	0.94	0.94	0.94	93	0.94
MKL+CVT	0.97	0.97	0.97	93	0.97

MKL+ CVT has better accuracy than SVM+WT.

May-June 2020
ISSN: 0193-4120 Page No. 5346-5356

6. Conclusion and Future Work

In this paper, we have suggested an effective detection technique for driver's drowsiness using eye aspect ratio and fatigue facial expressions (yawning and drowsiness) detection based on facial features. The method presented more accurate results for drowsiness detection as compared to the existing method. We used curvelet for denoising the image which is better than other denoising technique. Using Discrete Cosine Transform, The system is capable of extracting the most important discriminatory features from a face pic to represent facial features. This compression feature helps to reduce computational time. MKL offers high accuracy, 97% on real-time detection.

The possible limitations of our proposed system are that it may not work properly if the surrounding lighting condition is very low and occlusion occurs while detecting facial landmarks. We consider it as future work where we would like to improve the proposed technique in low light conditions and occlusions.

References

[1] Bappaditya Mandal, Liyuan Li, Gang Sam Wang, and Jie Lin "Towards Detection of Bus Driver Fatigue Based on Robust Visual Analysis of Eye State", IEEE Transactions On Intelligent transportation systems, Vol. 18, No. 3, March 2017

[2] Dipalee Gupta, Siddhartha Choubey, "Discrete Wavelet Transform for Image Processing", International Journal of Emerging Technology and Advanced Engineering, Vol. 4, No. 3, March 2015.

[3] Gu-Min Jeong, Hyun-Sik Ahn, Sang-Il Choi, "Pattern Recognition using Feature Feedback: Application to Face Recognition," International Journal of Control, Automation and Systems, Vol. 8, pp. 141-148, 2010.

[4] Jean-Luc Starck, Emmanuel J. Candès, David L. Donoho, "The Curvelet Transform for Image Denoising", IEEE Transactions On Image Processing, Vol. 11, No. 6, June 2002

[5] Kun Su Yoon, Wan-Jin Kim, "Efficient Edge-Preserved Sonar Image Enhancement Method based on CVT for Object Recognition", IET Image Process, Vol. 13, No. 1, pp. 15-2, 2019

[6] Kusuma Kumari B.M, Prof. Ramakanth Kumar, "A Survey on Drowsy Driver Detection System", International Conference On Big Data Analytics and Computational Intelligence (ICBDACI), 2017

[7] Lei Zhao, Zengcai Wang, "Driver Drowsiness Detection using Facial Dynamic Fusion Information and a DBN", IET Intell. Transp. Syst, Vol. 12, No. 2, pp. 127-133 2018

[8] Liu Cao, Longxu Jin, Hongjiang Tao, Guoning Li, "Multi-Focus Image Fusion Based on Spatial Frequency in Discrete Cosine Transform Domain", IEEE Signal Processing Letters, Vol. 22, No. 2, February 2015

[9] M. Sugur, B. Hemadri, and P.Kulkarni. "Drowsiness Detection Based On Eye Shape Measurement," International Journal of Computer and Electronics Research, Vol 2, No. 2, 2013.

[10] M. Omidyeganeh, A. Javadtalab, S. Shirmohammadi, "Intelligent Driver Drowsiness detection through Fusion of Yawning and Eye Closure", IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems Proceedings, 2011

[11] Maryam Keyvanara, Nasrin Salehi, Amirhassan Monadjemi, "Robust Real-Time Driver Drowsiness Detection based on Image Processing and Feature Extraction Method", Int. J. Vehicle Safety, Vol. 10, No. 1, 2018.

[12] Mehmet Gonen, Ethem Alpaydın, "Multiple Kernel Learning Algorithms", Journal of Machine Learning Research Vol. 12, pp: 2211-2268, 2011.

[13] Mehran Kafai, Kave Eshghi, Bir Bhanu, "Discrete Cosine Transform Locality-Sensitive Hashes for Face Retrieval", IEEE Transactions on Multimedia, Vol. 16, No. 4, June 2014.

[14] Miss Monika, Soni changlani, "A Comparative Study of Wavelet and Curvelet Transform for Image Denoising", IOSR Journal of Electronics and Communication Engineering (IOSR-JECE) e-ISSN: 2278-2834,p- ISSN: 2278-8735. Vol. 7, No. 4, pp. 63-68, 2013.

[15] Mohammad A. Haque, Ramin Irani, Kamal Nasrollahi, Thomas B. Moeslund, "Facial Video based Detection of Physical Fatigue for Maximal Muscle Activity", IET Comput. Vis., Vol. 10, No 4, pp. 323-329, 2016.

[16] Muhammad Athoillah, M. Isa Irwan, "Support Vector Machine with Multiple Kernel Learning for Image Retrieval", International Conference on Information, Communication Technology and System, 2015.

[17] Muhammad Ramzan, Hikmat Ullah Khan, Shahid Mahmood Awan," A Survey on State-of-the-Art Drowsiness Detection Techniques", IEEE Access, Vol. 7, 2019.

[18] National Highway Traffic Safety Administration Research on Drowsy Driving. https://www.nhtsa.gov/, Accessed on October 20, 2015

[19] N. Dalal, B. Triggs, "Histograms of Oriented Gradients for Human Detection in Computer Vision and Pattern Recognition", 2005. CVPR 2005. IEEE Computer Society Conference, Vol. 1, pp. 886– 893, 2005

[20] Ralph Oyini Mbouna, Seong G. Kong "Visual Analysis of Eye State and Head Pose for Driver Alertness Monitoring", IEEE Transactions On

Intelligent Transportation Systems, Vol. 14, No. 3, September 2013.

[21] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, "Adaptive histogram equalization and its variations," Computer vision, graphics, and image processing, Vol. 39, pp. 355368, 1987

[22] Sajid Ali Khan, Shariq Hussain, Sun Xiaoming, Shunkun Yang, "An Effective Framework for Driver Fatigue Recognition Based on Intelligent Facial Expressions Analysis", IEEE Access, Vol, 6, pp: 67459 – 67468, 2018.

[23] Samuel Lawoyin, Ding-Yu Fei, Ou Bai, “Evaluating the efficacy of an Accelerometer based Method for Drowsy Driving Detection ”, Int. J. Vehicle Safety, Vol. 8, No. 2, 2015

[24] Sang-Joong Jung, Heung-Sub Shin, Wan-Young Chung, “Driver Fatigue and Drowsiness Monitoring System with Embedded Electrocardiogram Sensor on Steering Wheel”, IET Intell. Transp. Syst., Vol. 8, No.1, pp. 43–50, 2014

[25] Thamizharasi Ayyavoo, Jayasudha John Suseela “Illumination Pre-Processing Method for Face Recognition using 2D DWT and CLAHE”, IET Biometrics, Vol. 7, No. 4, pp. 380-390, 2018

[26] Venkatesh Balasubramanian, K. Adalarasu, A. Gupta, “EEG based Analysis of Cognitive Fatigue during Simulated Driving”, Int. J. Industrial and Systems Engineering, Vol. 7, No. 2, 2011.

[27] Zhuoni Jie, Marwa Mahmoud, Quentin Stafford-Fraser, Peter Robinson, “Analysis of Yawning Behavior in Spontaneous Expressions of Drowsy Drivers” 13th IEEE International conference on Automatic Face & Gesture Recognition, 2018