International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Intelligent Driver Warning System using Deep Learning-based Facial Expression Recognition

S.Suchitra, S.Sathya Priya, R.J. Poovaraghan, B.Pavithra, J.Mercy Faustina

Abstract: ***Driver's inattention is one of the major factors and reasons in occurrence of many road accidents and unforeseen crashes. Hence it is crucial to develop an automatic driver warning system that can send timely warning signals to the drivers. This issue involves determining the driver's mental state that is ultimately based on the driver's facial expressions. Automated facial emotion recognition is a recent development in the image processing domain and is the need of the hour in applications like driver warning systems. The existing methods are capable of recognizing facial emotions even when provided with a noisy signal or imperfect data, but ultimately it lacks accuracy. It is also ineffective in dealing with spontaneous emotions, and recognition. The proposed approach develops a driver warning system that extracts the facial expressions based on a novel efficient Local Octal Pattern (LOP) and effectively recognizes the facial expressions based on Deep Neural Networks, Convolutional Neural Networks (CNN). The LOP feature map serves as an input to CNN and guides in the selection of CNN learning data thereby improving and further enhancing the understanding and learning of CNN. It also has an ability to recognize both natural and spontaneous emotions, as well as image and video can be considered as an input.The experimental results consideringYawDD dataset indicates that the proposed system has been efficiently evaluated by considering the with metrics such as Precision, Recall and F-Score and thereby it is observed and inferred that the proposed system obtained a high recall rate of 96.09% in comparison with the other state-of-the-art methods.***

Keywords : Convolution Neural Networks, Deep Learning, Driver Warning System, Facial Emotion Recognition, Local Octal Pattern.

I. INTRODUCTION

The high degree of mental distraction of a person driving the vehicle along with the consumption of alcohol and over-speeding is the most important cause of occurrence of majority of traffic accidents and unforeseen crashes. According to statistics from related departments, it is observed that driving fatigue condition accounts for the largest proportion of road traffic accidents. Many drivers who are in the fatigue state may experience sleepiness, short loss of consciousness and that they usually don't perceive that they are within the tired and drowsy state. The driver driving for duration greater than eight hours a day, excessive physical activity, lack of sleep, etc may be the cause of fatigue in drivers. There arises weakness in driver's mental stability and he will not be able to handle traffic problems in the road accurately.

Driving fatigue is authenticated and confirmed by one of two aspects of consideration: First, before the fatigue behavior is found, the driver driving the vehicle cumulatively within 24 hours is being calculated. If it is more than eight hours, it is automatically considered as fatigue though the driver has no tired feeling. Second, through investigation, albeit the additive driving time, driving fatigue is authenticated even if it is not equivalent to the standard eight hours of driving daily and if the driver with an observation that is confirmed as excess of physical hard-work due to the following reasons like lack of sufficient sleep, driving fatigue weakness and sleepiness, he cannot accurately process traffic-related issues. Though there are completely different technical testing strategies available, relying solely on appropriate testing methodology like by-law social control isn't comfortable.This is due to limited human material resources, and driver unconsciousness on fatigue driving. Road traffic accidents causes many road safety problems around the world thereby greatly threatening and influencing the life and property of social public [1-3]; the probability of occurrence of traffic accidents due to driving fatigue is 4 to 6 times more in comparison to normal driving [4]; henceforth it has gained a lot of global attention. Hence investigation and elaborated analysis on the driver fatigue warning has important sensible significance and enhanced application value.

An automated driver warning system is required to alert the driver during crucial conditions when his mental stability is poor. This issue involves determination of driver's mental state, which can be done based on the facial expressions of the driver. There are many different driving fatigue detection approaches particularly the following: physiological-based approaches [5-7], behavior-based approaches, operation state of vehicle-based approaches [8-10], and facial expressions-based approaches [11,12].

Manuscript published on November 30, 2019.
* Correspondence Author

S.Suchitra*, Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, India. Email: ssuchitra@hindustanuniv.ac.in

S.Sathya Priya, Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, India. Email: sathyap@hindustanuniv.ac.in

R.J. Poovaraghan, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, India. Email: poovaraghan1976@gmail.com

B.Pavithra, Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, India. Email: bpavithra@hindustanuniv.ac.in

J.Mercy Faustina, Department of Computer Science and Engineering, Hindustan Institute of Technology and Science, Chennai, India. Email: mercy93faustina@gmail.com

The physiological-based technique necessitates the sensor to be in contact with the driver's body and therefore the detailed information based on the points of contact are required, which in-turn is unsuitable for applying to the normal driving scenario. It depends on head movements of the driver, different techniques obtainable that have plenty of probability owing and randomness correlation between the detailed head movements and fatigue like swing driving and lane change mirrored within the vehicle driving behavior. Even if these behaviours are partly mirrored to find out the driver is in a state of fatigue or not and also it is difficult to find out more precisely and accurately if the driver is in a state of extreme tiredness. The use of the camera to capture driver's image in real-time and analyze the images to find the driver's fatigue state is done by various different machine learning based methods. The method has the foremost essential characteristic of non-contact, however needs additional improvement in terms of accuracy and sensible quality.

There is a lot of research interest towards Artificial Intelligence (AI) due its recognition ability in human emotion. Artificial intelligence plays an indispensable role in human computer interaction with the strong ability to recognize different variations of facial expressions. To precisely identify a person's emotions, there are certain psychological based characteristics such as human body movements, speech, heartbeat, hand gestures, blood pressure, facial expressions, and so on are observed. Out of these, the most effective and essential characteristic for identifying the emotion of a person is facial expressions. Mehrabian [13] research enumerates that the facial expressions effectively transmit 55% of the messages in terms of communication between face to face. An emotion exhibited in terms of facial expression, it is basically the variation in the corresponding position of the muscles underneath of the skin on face. These observed movements on human face precisely indicate the emotional state of an individual.

A facial expression is a highly powerful non-verbal communication that is used for direct one-to-one communication. The basic essential facial expressions as identified by Ekman et al. are anger, disgust, fear, happiness, sadness, and surprise. By detecting and analyzing these facial expression based features and facial muscular movements, one can easily identify the emotion as expressed by an individual and eventually detect the different mood swings of the individual such as pain, frustrated, happy, elited and so on. Nowadays, many real-time applications in various fields such as medicine, data-driven animation, E-Learning based domains, interactive gaming, entertainment, and so on are based on the automated facial expression recognition system framework.

Traditional methods of recognizing emotions in a human face are carried out only at laboratory controlled environments. Recognizing facial emotions at real time scenario is complex as well as a challenging task. Evolution in image processing and pattern recognition makes facial emotion recognition as an automated process with improved accuracy and recognition percentage. But often still images which are static are considered for recognition; Videos which are dynamic are still challenging. Facial expression recognition system faces critical real time challenges such as pose variation, illumination and occlusion which impacts greatly on the recognition accuracy.

In this paper, the proposed system develops an automated driver warning system which recognize the facial emotions under real time road driving scenario and captures the emotion to provide an alert for the driver at extreme or crucial condition where it tracks the frames of images which is a dynamic video image and matches the difference between the frames of images by low rank detection and detects the odd frame out and recognize the facial emotion by LOP-CNN.

Convolutional Neural Network (CNN), which is one of the most vital deep learning methodologies, is inferred to be successful for its exceptional and outstanding categorization in the related domain of image recognition and classification. The CNN has the stupendous capability of automatic extraction of features and transformation of variance, which in-turn projects it a significant neural network methodology for image classification. Each of the layers in CNN tries to extract the most unique characteristic features as taken from the given input image which in turn makes it a more robust neural network. The main aim of the CNN is to get the proper destined output by converting a given set of inputs into more accurate and meaningful outputs as destined. The LOP-CNN is meticulously employed and implemented in this research work, which optimally achieves maximum efficiency and effectiveness.

The key contributions of this paper include:

- The design of a Local Octal Pattern-Convolutional Neural Network (LOP-CNN) method for an efficient Driver warning System using Deep Learning based Facial Expression Recognition.
- CNN based feature extraction minimizes the semantic gap and improves the overall efficiency by using Facial Expression Recognition.

The rest of the paper is structured and organized as follows. Section II discusses the existing work related to facial recognition methods. Section III introduces the proposed method including Driver Warning System using CNN-based Face Expression Recognition. Section IV enumerates the experimental implementation and summarizes the results. Section V concludes the paper along with the future scope for enhancement.

II. RELATED WORKS

There are some efforts made in the development of a system that detects the drowsiness characteristic of the driver based on certain factors. The factors like head movements, variability in heart rate, grip quality and steering wheel movement against the path markings on the road were meticulously recorded to detect the drowsiness. The eyes of driver were concentrated to investigate the drowsiness using the developed drowsy driver detection system. Based on the parameters used for detection, the drowsiness detection techniques are classified into two types namely the intrusive method and non-intrusive methods of detection. The distinction is based on the condition whether an instrument is connected and coupled with the driver or not.

Retrieval Number: C4028098319/19©BEIESP
DOI:10.35940/ijrte.C4028.098319
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

In the intrusive method, an instrument is well-connected to the driver and the value of that instrument is checked and recorded. This method has high levels of accuracy, where the accuracy is proportional to the driver's discomfort; so it is rarely used.

The driver's eyes as opened or closed are checked by the system [14]. A warning sign is disseminated to the driver when the eyes are closed for a longer duration. The driver drowsiness monitoring and early warning system makes use of machine learning techniques based on telemetry data of the vehicle [15]. This system offers real time monitoring of the driving pattern, thereby ensuring safe driving. A major solution for efficient driver monitoring and event detection is to employ the 3-D information gathered from a range camera. Head pose estimation and region-of interest identification are provided by the system by effectively combining 2-D and 3-D techniques.

The clouds of 3-D points from the sensor are captured and 2-D projects are analyzed based on head movements and also extracted for further analysis [16]. In order to detect drowsiness, an ingenuity method which combines the methods comprising of both computer vision as well as physiological bio-signals can be systematically employed. PCA model is employed initially to detect the face region; the eye region is determined by employing GA based on face segment. The changes in the signal waveform from active awake state to drowsy state can be systematically analyzed using Photo Plethysmography (PPG) [17].

The main idea and contribution of the proposed work is the central focus on facial expressions based driving fatigue detection system in which a deep learning based feature extraction using LOP is used for facial expression recognition. There are three conventional methods available for facial recognition namely appearance-based methods, Geometric-based methods, and neural network based methods. The foremost classic method for observing face recognition is face geometry. The human face is represented and characterized by an optimal set of facial landmark points in Geometrical feature based approaches. The location and shape of facial components are determined by the angle and distances between those facial points. To classify the input face, the feature vectors that suitably represent the face are fed as input to the classifier. Inorder to accurately locate the landmark points on face, facial point detectors have to be found which is difficult task. The features are unanimously extracted from the relative pixel gray values as given in the face image in case of appearance-based methods. The pixel intensity is determined by the intensity of light emitted from the image. Some of the conventional appearance-based methods which does not include facial points information are Principal Component Analysis (PCA) including Eigen face, Local Binary Pattern (LBP), Linear Discriminate Analysis (LDA), Local Gabor Binary Patterns (LGBP) and multi-orientation multi-resolution Gabor wavelets. The features which are extracted from the above methods are given as input to the neural network, which then tries to identify the input image. This is evaluated based on the analysis of the descriptive features, which are obtained from geometric and appearance based methods, Classifiers such as Artificial Neural Networks (ANNs), Support Vector Machines, (SVMs), K-Nearest Neighbors (KNNs), Hidden Markov Models (HMMs), which classifies the input image into one of the expressions. In recent years, the methods like Deep Learning have gained prominence and are considered as a promising technique to perform facial expression recognition.

Archana Shirsat et al. [18] has enumerated a novel FER using an efficient LBP for feature extraction and employed with Artificial Neural Network (ANN) for classification. The illumination variant based local binary pattern tries to detect the features using very simple mathematical computation. Some of the extracted features from the given LBP methodology were comprehensively fed into ANN for further classification. Three feature extraction methods were used by Reza Azmi [19]: Gabor filters, LBP and local Gabor binary pattern (LGBP). The KNN classifier, coupled with calculation for the vector distance measure is utilized as a classifier. The LGBP due to its more powerful and effectiveness tends to show high accuracy based on different occlusion conditions on FER. Banu, Simona et al. [20] designed a model for detection of overall face, with detailed features of eyes and mouth, which makes use of the Haar feature functions and is applied with Bezier curves, which are used to extract features for geometrically different facial parts. Two layer Neural Networks method with K-means algorithm is used for pre-classification. Facial expression based recognition method using Hidden Markov Model was proposed by Jun Wang et al. [21]. Jun Wang et al. [22] has proposed a deep CNN for identifying and recognizing a deeper feature representation of facial expression as part of the process of facial expression recognition system, which is mainly used in order to gain automatic recognition. These methods achieve high recognition rate, but the drawback is that needs additional improvement in terms of accuracy and sensible quality.

Though these works achieve salient performance on facial expression recognition, an effective automatic driver warning system is proposed by extracting the facial features using LOP-CNN for facial expression recognition.

III. PROPOSED METHOD

The main contribution of this paper encompasses the optimal extraction of facial expression based on an efficient texture pattern LOP and further enhanced recognition of facial expression based on deep learning that is using CNN. In this proposed work, it utilizes LOP feature vector map, given as an input to CNN, which will in turn aid in assistance for the selection of CNN learning data, thereby improving the understanding and learning of CNN.

A. System Design

Fig.1 shows the overall framework of the Driver Warning System (DWS) using CNN-based Facial Expression Recognition. The DWS consist of three phases like face detection, face feature extraction and face recognition or retrieval. Face detection is done to locate the faces in the image or video using Viola-Jones Face Detector method [23]. Then the Active Shape Model is employed to find totally different facial landmarks on the face image. The utilization of those totally different facial landmarks to align each face images with face mean shape method [24] is done by barycentric coordinate based mapping process. The feature extraction using LOP algorithm in the face elements after the face alignment is obtained and the facial expression is recognized using CNN.

The recognition and retrieval is accomplished using Support Vector Machine (SVM). The system will send an alarm to the driver based on the classified facial expression.

i) Face Detection

For the images or videos that are captured by the camera, Viola-Jones Face Detector method is employed to detect and locate the faces in the image or video. By incorporating this method, although the training becomes slow, but the detection is made fast. This algorithm makes use of Haar basis feature filters and it does not use multiplications. Detection tends to happen inside a detection window. A minimum and maximum frame size is optimally chosen, and for each of the recognized sizes, a sliding step size is eventually chosen; after which the detection frame is moved accurately across the corresponding face image. Then to locate the different available facial landmarks on the facial image, the Active Shape Model is employed. In general, the shape of an object is represented with a set of points. An optimal barycentric coordinate based mapping process is duly applied on the facial landmarks to categorically align every face with the corresponding face mean shape model. The use of these different facial landmarks to align every face image in the database associated with the face mean shape method [24] is done by barycentric coordinate based mapping process.

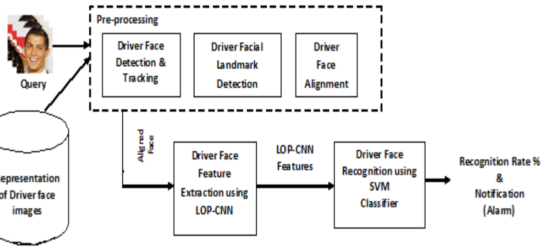

Fig. 1. Overview of the proposed method

ii) Face Feature Extraction

In the feature extraction phase, for each of the detected facial components, 7X5 grids are inherently extracted; in which each grid is represented as a square patch [25]. Using this technique the same as projected in [25], there will be a total of 175 grids represented from five significant elements of face including two mouth corners, nose tip and two eyes on the aligned face image. A face image patch is rooted out and extracted from each grid and LOP feature descriptor [26] is accordingly computed and evaluated as a local features. The CNN is then trained and map with the LOP feature vector, which serves as an input to CNN. Every LOP feature vector mapped is then subsequently merged to fully connected networks, and also the facial expression is vitally recognized to a selected category supported based on the output of CNN. Similarly, in the test classification face images, initially LOP feature vector is accurately extracted, after which the LOP features is successfully fed into the CNN classifier for further classification. Based on the facial expression recognized; our proposed system will alert the driver with an alarm.

a) LOP-based Face Feature Extraction

After face alignment, the feature extraction in the face is obtained with LOP. The LOP algorithm successfully encodes the information between the center pixel and the reference pixel, purely based on the direction of the center pixel. In the LOP, encoding information is calculated for all directions (horizontal, vertical and diagonal) of pixel values using first-order derivatives for each center pixel in the given face image. From the pixels, 8-bit octal pattern is obtained for each centre pixel in the given Image I. The octal patterns are divided into eight parts from the calculated directions. Finally, three binary patterns are generated from each direction of centre pixel as a total of 24 (8*3) binary patterns.

The g_c denotes the center pixel in I, and let g_h, g_v and g_d denote the horizontal, vertical and diagonal neighbourhoods of g_c, respectively. Then, the first-order derivatives at the center pixel g_c can be written as

$$I^1_{0^o}(g_c) = I(g_h) - I(g_c) \quad (1)$$

$$I^1_{45^o}(g_c) = I(g_d) - I(g_c) \quad (2)$$

$$I^1_{90^o}(g_c) = I(g_v) - I(g_c) \quad (3)$$

and the direction of the center pixel in the given image can be calculated as

$$I^1_{\mathrm{Directions}}(g_c) = \begin{cases} 1, I^1_{0^o}(g_c) \geq 0 \text{ and} I^1_{90^o}(g_c) \geq 0 \\ 2, I^1_{0^o}(g_c) < 0 \text{ and} I^1_{90^o}(g_c) \geq 0 \\ 3, I^1_{0^o}(g_c) < 0 \text{ and} I^1_{90^o}(g_c) < 0 \\ 4, I^1_{0^o}(g_c) \geq 0 \text{ and} I^1_{90^o}(g_c) < 0 \\ 5, I^1_{0^o}(g_c) \geq 0 \text{ and} I^1_{45^o}(g_c) \geq 0 \\ 6, I^1_{0^o}(g_c) < 0 \text{ and} I^1_{45^o}(g_c) \geq 0 \\ 7, I^1_{45^o}(g_c) < 0 \text{ and} I^1_{90^o}(g_c) < 0 \\ 8, I^1_{45^o}(g_c) \geq 0 \text{ and} I^1_{90^o}(g_c) < 0 \end{cases} \quad (4)$$

The LOP algorithm effectively constructs the pixel of matrix within the face elements with success encodes the knowledge of the given image between the center pixel and its reference pixel, strictly supported the direction of the center pixel. The center pixels of the diagonal and adjacent pixels considers as center pixels and computing encoding information based on the threshold value for all direction and interprets the binary number as the final result. The process continues for the entire adjacent pixels in the given face image. From the process, extraction of feature values of the face elements is effective and efficient since each pixel is considered as the center pixel as shown in Fig..2.

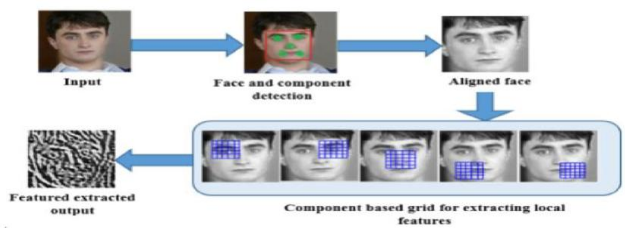

Fig. 2. Face Feature Extraction based on Local Octal Pattern (LOP) feature descriptor [26]

b) CNN-based Facial Expression Recognition

One of the best known optimal representative means of network structures in the field of deep learning is the Convolution neural network [27], [28].

It has unanimously gained prominence in the area of facial image recognition and classification. The resulting data in the bundled form of multiple set of arrays are efficiently processed by ConVnets. The CNN is capable of taking raw information of an image as an input there by passing the feature extraction and reconstruction procedures as part of the standard learning algorithms. It is similar to the bio-logical neural network in terms of its weight sharing network structure. This in-turn lessens the complexity of the overall network model and also the number of weights associated with it.

The CNN remain unaltered to any kind of orientations such as translation, scale, tilt, and other deformation.

Deep learning based Facial Emotional Recognition approaches are highly dependent on physical elements of the face and learn efficiently from the input face images using other different pre-processing techniques [28]. Among the all available deep-learning models, the CNN is that the preferred network model in use. In CNN-based method, to produce an extracted feature vector map using the input image is convolved through a filter group in the convolution layers. Each feature vector map is added and combined with fully connected network, and the facial expression is inherently recognized.

The CNN comprises of several layers. These layers can be of three different types namely Convolutional, Max-Pooling and Fully-Connected layers. First Convolutional layers are applied i.e. the rectangular grid of neurons on the primary components of the face such as left eye, right eye, mouth two corners and nose tip. Each neuron takes and utilizes input from a rectangular section of the previous layer of the face components, with the characteristic that the weights for this rectangular section are the same for each neuron in the convolutional layer. Thus, it is an image convolution of the previous layer, where the weights are used to exclusively specify the convolution filter. Further added to this, there may be a number of grids in each convolutional layer, where each of this grid takes inputs from the previous layer by exploitation probably totally different filters.

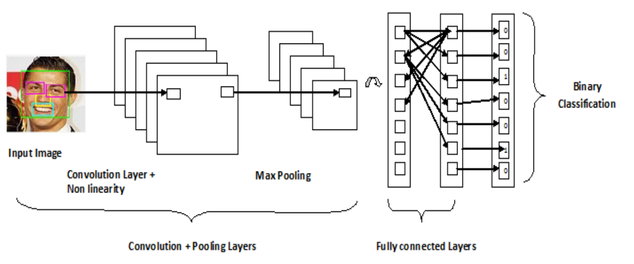

Fig. 3. Procedure of LOP-CNN Method

After the completion of convolutional layer, the second layer which is the pooling layer is applied. The pooling layer utilizes tiny rectangular blocks from the previous convolutional layer of the face components and subsamples it to yield a single conglomerated output from that block. There are different ways available to do this pooling like considering the maximum or the average, or a learned linear combination of the neurons in the block. The pooling layers considered can be max-pooling layers i.e. it take the maximum number of the blocks that are pooling in the face elements.

Finally, when applied to many convolutional and max-pooling layers, the high-level modes of interpretation within the neural network are done via multi-layer i.e. fully connected layer. This multi-layer takes inputs from all the neurons as part of the previous layer of the face elements and attaches it to every single individual neuron. It is inferred that fully connected layers are not much spatially located anymore, so there are no convolutional layers after a fully connected layer. Fig. 3 shows the procedure of LOP-CNN method.

Suppose, consider $N \times N$ neuron layer which is followed by convolutional layer. If use an $m \times m$ filter, convolutional layer output will be of size $(N - m + 1) \times (N - m + 1)$. In order to compute the pre-nonlinearity input to some unit x^l_{ij} in our layer, we need to produce output by summing up the weighted filter components from the previous layer cells:

$$x^l_{ij} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \omega_{ab} y^{l-1}_{(i+a)(j+b)} \qquad (5)$$

Then, the convolutional layer applies its nonlinearity as follows:

$$y^l_{ij} = \sigma\left(x^l_{ij}\right) \qquad (6)$$

Algorithm of CNN-LOP based FER

Inputs (Total): $x^0_{ij}, \dots x^l_{ij}$

Output: y^l_{ij}

Step 1: Begin

Step 2: Load the sample frame and convert into gray scale

Step 3: Apply LOP for all facial components

Step 4: Compute outputs for layers with Known inputs $y^l_i = \sigma(x^l_i) + I^l_i$

Step 5: Compute inputs for next layer from these outputs

$$x^l_i = \sum_j w^{l-1}_{ji} y^{l-1}_j$$

Step 6: Repeat steps (3) and (4) until reach the output layer and know

values of y^L

Step 7: Similarly repeat steps (3), (4) and (5) for all facial components

Step 8: End

B. Quantization Process and Indexing

In the quantization process and indexing phase, every descriptor is quantized into sparse codewords, after obtaining local feature descriptors, and the same procedure is applied to all components in a single face image to find different semantic codewords. Ultimately all these semantic codewords are combined together to represent the final resulting image. By using sparse codewords [29] for face image, the following optimization problem is solved:

$$\min_{D,V} \sum_{i=1}^{n} \left\|x^{(i)} - Dv^{(i)}\right\|_2^2 + \lambda \left\|v^{(i)}\right\|_1 \text{subjectto} \left\|D_{*j}\right\|_2^2 = 1, \ \forall j \qquad (7)$$

Where $x^{(i)}$ is the original feature values that are extracted from the face image patches i, $D \in R^{dXK}$ represents to be learned sparse dictionary,

Retrieval Number: C4028098319/19©BEIESP
DOI:10.35940/ijrte.C4028.098319
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

which contains K centroids with dimensions. $V = [v^{(1)}, v^{(2)}, \ldots, v^{(n)}]$ is the semantic codewords representation of the face image patches. The limit on each column to make and keep D from becoming randomly large size. Using semantic code, a feature value is a linear combination of the column vectors of the dictionary. The equation contains two parts like sparse dictionary learning (D) and semantic encoding information (V). Using random sampled face image patches as a dictionary, it can achieve similar performance as the one obtained by using a learned sparse dictionary and if the sampled patches provide a set of over complete basis, in turn representing input data. Next, consider for each face image, where after computing the semantic sparse representation using the method described in the above process, representation of semantic sparse codewords set c(i) can be used to represent it by taking non-zero entries in the semantic sparse representation as codewords. Later, the matching between two face images are then computed as follows,

$$S(i,j) = ||C^{(i)} \cap C^{(j)}|| \tag{8}$$

It is inferred that the face image matching according to this similarity score can be efficiently and successfully found in the database.

C. SVM based Recognition

One of the supervised machine learning algorithms available is the Support Vector Machine (SVM) [30] that efficiently constructs the hyperplane in a high-dimensional space to perform the classification or recognition task. SVM learning intends at finding the best hyperplane that has the ability to maximize the margin among any two classes. The basic principle of SVM is a binary classifier, which segregates or classifies data into two classes.

SVM classification is applied to the trained images to detect the difference of face image and improve the reliability of the categorized face images. SVM method proposes that training data of two clustered face image collections $\{(x_I, y_I), \ldots (x_i, y_i), \ldots (x_n, y_n), x_i \in R_m, y_i \in +1, -1$, which can be separated by a hyper plane without error. If there are m-dimensional vector w considered with constant as b, then:

$$\{(wx_i) + b > 0, y_i = 1 \ (wx_i) + b < 0, y_i = -1 \tag{9}$$

The classification problems are primarily used to seek the suitable (w, b) set and divide different forms of face images to investigate and to find an optimal classification resultant hyperplane. This classification proportionally depends on a smaller number of sample points, proximate to it and ultimately has no bonding relation to other samples. When this classification hyperplane is in a normalized state , then:

$$y_i\big((wx_i) + b\big) - 1 \geq 0, i = 1, \ldots n \tag{10}$$

The optimal classification that meets condition (10) and also minimizes the classification interval (i.e.$\frac{1}{2}\|w\|^2$) is known as the optimal classification hyperplane, which deals with the better prevention of difference face images. By introducing and utilizing Lagrange function, followed by directing those hyperplanes to meet the Karush Kuhn Tucker theorem, the optimal classification function can be obtained as follows:

$$f(x) = \mathrm{sgn}\{\textstyle\sum_{SV} \ \alpha_i^* y_i (x_i . x_j) + b^*\} \tag{11}$$

IV. EXPERIMENTAL RESULTS

For experimental analysis and corresponding results, YawDD [31] used two video datasets of drivers with various facial expressions. The videos are taken and shot in real time, under varying and capricious illumination conditions. In YawDD datasets, different mouth conditions such as normal, yawning, talking or singing are part of each video.

- In the first video dataset, a camera is installed under the front mirror of the car. This provides 322 videos comprising of each of the male driver and female driver, with and without glasses or sunglasses, from completely different races, and in three varied situations such as normal driving (no talking), talking or singing while driving, and yawning while driving.
- In the second video dataset, the camera is directly placed on the driver's dash. It provides twenty nine videos consisting of each male driver and female driver, with and without glasses or sunglasses, from completely different race sets, under varying situations.

All videos in the YawDD dataset contain 640x480 24-bit true color (RGB) 30 frames per second and AVI format without audio. The considered total dataset size is about 5 Gigabytes. The parameters which are considered for experimentation is listed in Table 1.

- Set 1 contains action of different person such as normal, talking, and yawning
- Set 2 contains different types of illumination such as sunny, rainy, and cloudy
- Set 3 contains different types of eye condition such as closed, open, etc.

Table- 1: Experimental setup for LOP-CNN method

Total no of videos tested	150 Set 1: 50 (action) Set 2: 50 (Illumination) Set 3: 50 (eye condition)
Type of videos taken	640x480 -bit true color (RGB) 30 frames per second AVI format without audio

The metrics which are considered to efficiently evaluate the effective performance of the proposed system are Precision, Recall, and F-Score. To evaluate the optimal efficiency and robustness of the algorithm, metrics such as Precision and Recall rates are evaluated and computed based on the recognition rate. For the given input, our system recognizes the facial expression; it classifies and provides the warning notification (alarm) to the driver if the recognition result falls under defined categories such as closed eyes, fatigue face, sleepy face, etc. Fig.4 shows that the proposed system produces the highest recall rate for all types of parameters like action, illumination and eye conditions. The average of all measures for the proposed system has been listed in Table2.

Retrieval Number: C4028098319/19©BEIESP
DOI:10.35940/ijrte.C4028.098319
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

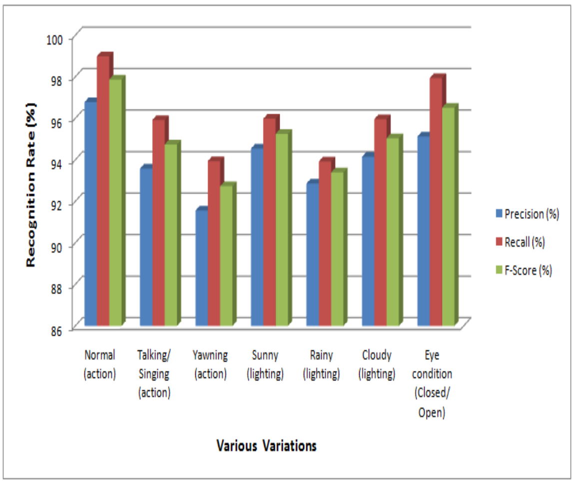

Fig. 4. Average Performance measures of Precision, Recall and F-Sore for LOP-CNN Method

To give an average estimate of the performance of the LOP-CNN system, the results have also been verified against two other existing algorithms: that is Template Matching [32] and Vision-based system and PERCLOS Eye openeness [33]. To show the ability of the proposed system to extract local features in a better way for the variation in Driver Warning System, the proposed LOP-CNN method is compared with Template Matching and Vision-based system and PERCLOS Eye openeness. The results indicate that LOP-CNN method shows clear improvement over the above existing methods due to the deep learning based feature descriptor. The performance measures for comparison are enumerated in Table 3. The Figure 5 depicts the Comparison of average performance measures of Precision, Recall and F-score for LOP-CNN and other methods.

Table-2 Average Performance measures for the LOP-CNN Method

Type of parameter	Precision (%)	Recall (%)	F-Score (%)
Normal (action)	96.78	98.99	97.87
Talking/ Singing (action)	93.58	95.92	94.74
Yawning (action)	91.56	93.94	92.73
Sunny (lighting)	94.54	95.98	95.25
Rainy (lighting)	92.86	93.92	93.39
Cloudy (lighting)	94.14	95.96	95.04
Eye condition (Closed/ Open)	95.12	97.94	96.51

In this paper, the proposed LOP-CNN has attempted over 150 videos with 3 sets of various groups (action, illumination, and eye closed/ open). These data set videos have been gathered from YawDD data set. From the experiment on the above data, it is noticed that the overall performance of the proposed system in providing the matching results for the face images of different action, illumination, and eye closed/ open has a Precision, Recall and F-score of94.08%, 96.09%, and 95.08% respectively, which is an encouraging result when compared to the Template Matching and Vision-based system and PERCLOS Eye openeness. The proposed method is compared and contrasted with the Template Matching method, whose overall Precision, Recall and F-score are 93%, 90.16%, and 91.56% respectively and the Vision-based system and PERCLOS Eye openeness method whose overall Precision, Recall and F-score are 89.3%, 86.54%, and 87.90% respectively.

Table-3: Comparison of average performance measures for LOP-CNN and other Methods

Method	Precision (%)	Recall (%)	F-Score (%)
Template Matching	93.00	90.16	91.56
Vision-based system and PERCLOS Eye openeness	89.30	86.54	87.90
LOP-CNN Descriptor **(Proposed Method)**	94.08	96.09	95.08

Fig. 5 shows the overall comparison of the average performance measures of Precision, Recall and F-score for LOP-CNN and other methods.So, it is concluded from the results that, in all cases, the proposed method is found to be better than the Template Matching and the Vision-based system and PERCLOS Eye openeness for all kinds of variations by providing better precision, recall, and F-score rates.

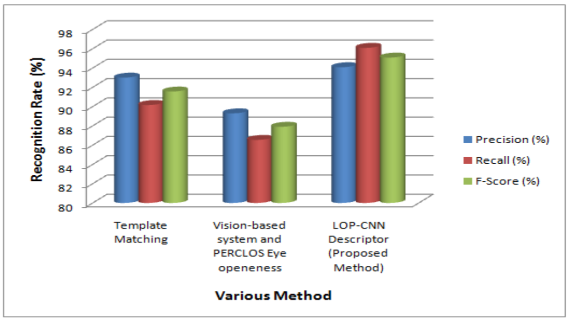

Fig. 5. Comparison of average performance measures of Precision, Recall and F-score for LOP-CNN and other methods

V. CONCLUSION AND FUTURE WORK

An automatic driver warning system is proposed and developed to send timely warning signals to the drivers based on the recognized facial expressions to avoid road accidents and unforeseen crashes. The driver warning system proposed extracts the facial expressions based on an efficient texture Local Octal Pattern (LOP) and recognizes the facial expressions based on Deep Neural Networks, i.e., CNN. It has an ability to recognize both natural and spontaneous emotions from both image and video inputs. The proposed system is efficiently evaluated with metrics such as Precision, Recall and F-Score and it is observed and inferred that the proposed system obtained a high recall rate of 96.09%, in comparison to other state-of-the-art methods. The proposed driver warning system outperforms the existing methods with respect to all the metrics considered.The proposed system can be further extended to include different maneuvers to ultimately make the driving system capable of dealing with all types of vehicles, under varying driving environments and scenarios.

Further enhancements are to be considered as part of future work to be able to withstand failure and retain robustness for extreme poses and heavy illumination as well as occlusions at certain point.

REFERENCES

1. Wang F, Chen X, Wang, D and Yang B 2017 An improved image-based iris-tracking for driver fatigue detection system, pp. 11521–11526.
2. Abtahi S, Omidyeganeh M, Shirmohammadi S and Hariri 2014 YawDD: A Yawning Detection Dataset", Proc. ACM Multimedia Syst. Singapore, March 19 -21, pp. 24-28, 2014. DOI: 10.1145/2557642.2563678.
3. Roshan Chitrakar and Huang Chuanhe 2012 Anomaly based Intrusion Detection usingHybrid Learning Approach of combining k-Medoids Clustering and Naïve Bayes Classification, the 8th Intl. Conf. on Wireless Comm., Networking and Mobile Computing, Shanghai, China.
4. Mandal B, Li L, Wang G S and Lin J 2016 Towards Detection of Bus Driver Fatigue Based on Robust Visual Analysis of Eye State, IEEE Trans. Intel. Transp. Syst., vol. P, no. 99, pp. 1–13.
5. Bor-Chun Chen, Yan-Ying Chen, Yin-HsiKuo and Winston H Hsu 2013 Scalable Face Image Retrieval Using Attribute-Enhanced Sparse Codewords, IEEE Trans. on Multimedia, Vol. 15, No. 5, pp. 1163-1173
6. Walecki R and Rudovic O 2017 Deep structured learning for facial expression intensity estimation. Image Vis. Comput., 259, 143–154.
7. Kahou S E, Michalski V and Konda 2015 Recurrent neural networks for emotion recognition in video, In Proceedings of the ACM on Intl. Conf. on Multimodal Interaction, Seattle, WA, USA, pp. 467–474.
8. Suchitra S and Chitrakala S 2017 Face Image Retrieval of Efficient Sparse Code words and Multiple Attribute in Binning Image, Intl. Journal of Brazilian Archives of Biology and Tech., Vol.60: e17160480.
9. Wu Z, Ke Q, Sun J and Shum H Y 2010 Scalable face image retrieval with identity-based quantization and multi-reference re-ranking, in Proc. IEEE Conf. Computer Vision and Pattern Recognit.
10. Park. P and Jain A k 2010 Face matching and retrieval using soft biometrics, IEEE Trans. Inf. Forensics Security, vol. 5, no. 3, pp. 406–415
11. of simple features, in Proc. IEEE Conf. Computer Vision and Pattern Recognit.,
12. Viola P and Jones M 2001 Rapid object detection using a boosted cascade
13. Ke Shan, Junqi Guo, Wenwan You, Di Lu and RongfangBie 2017 Automatic Facial Expression Recognition Based on a Deep Convolutional-Neural- Network Structure IEEE 15th Intl. Conf. on Software Engg. Research, Mgmt. and Appln. (SERA).
14. Subotica, Serbia Alptekin D and YavuzKahraman 2014 Early Facial Expression Recognition Using Hidden Markov Models, Intl. Conf. on Pattern Recog.
15. Simona Maria Banu, Gabriel Mihail Danciu, Gabriel Boboc Moga and Cristiana 2012 A Novel Approach for Face Expressions Recognition" IEEE 10th Jubilee Intl. Sym. on Intl. Systems and Informatics, 20-22 Sep.
16. Reza Azmi and Sam iraYegane 2012 Facial Expression Recognition in the Presence of Occlusion Using Local Gabor Binary Patterns, 20th Intl Conf. on Electrical Engg., May 15-17, Tehran, Iran.
17. Kiran Talele, Archana Shirsat, Tejal Uplenchwar and Kushal Tuckley 2016 Facial Expression Recognition Using General Regression Neural Network, IEEE Bombay Section Sym.
18. Lee B G, Jung S J and Chung W Y 2011 Real-time physiological and vision monitoring of vehicle driver for non-intrusive drowsiness detection, IET Commun., Vol. 5, Iss. 17, pp. 2461–2469.
19. Kashyap Vasudevan, Anjana Das P, Sandhya B and Subith P Driver Drowsiness Monitoring by Learning Vehicle Telemetry Data, Tata Elxsi Ltd., Technopark, Trivandrum, India
20. Gustavo A, Peláez C, Fernando García, Arturo de la Escalera and José María Armingol 2014 Driver Monitoring Based on Low-Cost 3-D Sensors, IEEE Trans. on Intl. Trans. Systems, Vol. 15, No. 4
21. Javed Ahmed, Jian-Ping Li, Saeed Ahmed Kran and Riaz Ahmed Shaikr Eye Behavior based Drowsiness Detection System, School of Computer Sci. & Engg. VESTC, Chengdu 611731, China.
22. Mehrabian A 1971Silent Messages". Wadsworth, Belmont, California.
23. skandarian A, Mortazavi A and Akbar Sayed R 2012 Drowsy and Fatigued Driving Problem Significance and Detection Based on Driver Control Functions, Handbook of Intel. Vehicles, Springer, pp. 941-974.
24. Garcia I, Bronte S and Bergasa L M 2010 Vision-based drowsiness detector for a realistic driving simulator, In: 13th Intl., IEEE Conf., on Intel. Transp., Systems, pp. 887- 894, Funchal.
25. Friedrichs F and Yang B 2010 Drowsiness monitoring by steering and lane data based feathers under real driving conditions, In: 18th European Signal Proc., Conf., Aalborg, pp. 23-27.
26. Forsman P M, Vila B J, Short R A, Mott C G and Dongen H P A V 2013 Efficient driver drowsiness detection at moderate levels of drowsiness, Accident Analysis and Prevention, vol. 50, pp. 341-350.
27. Kim Y, Youmin K and Hahn M 2008 Detecting driver fatigue based on the driver's response pattern and the front view environment of an automobile, In: 2nd Intl., Sym., pp. 237-240; Osaka, USA.
28. Fan X, Sun Y, Yin B and Guo X 2010 Gabor based dynamic representation for human fatigue monitoring in facial image sequences, Pattern Recog. Letters, vol. 31, no. 3, pp. 234-243.
29. Murphy E C and Trivedi M M 2009 Head pose estimation in computer vision: a survey, IEEE Trans. on Pattern Analysis and Machine Intel. vol. 31, no. 4, pp. 607-626.
30. Takahashi I and Yokoyama K 2011 Development of a Feedback Stimulation for Drowsy Driver Using Heartbeat Rhythms, Engineeringin Medicine and Biology Society, Boston.
31. Klauer S G, Dingus T A, Neale V L, Sudweeks J and Ramsey D J 2006 The Impact of Driver Inattention on Near Crash/crash Risk:An Analysis Using the 100 Car Naturalistic Driving Study Data, National Highway Traffic Safety Administration, Washington.
32. Jin L S, Niu Q N, Hou J H and Hu S 2012 A pedestrian detection method using 3D laser scanner, Information An Intl. Interdisciplinary Journal, vol. 15, no. 12A, pp. 5481-5490.
33. NHTSA Traffic Safety Facts 2011 Report DOT HS 811 402, NHTSA, US Dept. of Trans.
34. World Health Organization 2009 Global Status Report on Road Safety: Time for Action, Geneva: http://www.who.int/violence_injury_prevention/road_safety_status/2009/en/

AUTHORS PROFILE

Dr.S.Suchitra, received her B.E. degree from the University of Madras, Chennai, India in 1997. M.E. degree from Anna University, Chennai, India in 2009 and received her Ph.D from Anna University, Chennai, India in 2017. She has 21 years of teaching experience and currently she is working as an Associate Professor in the Department of Computer Science and Engineering, Hindustan Institute of Technology & Science, Chennai, India. She has published more than 30 research papers in international journals and conferences. Her research interest is image processing, image retrieval, computer vision, Big Data, Artificial Intelligence, Machine Learning and Deep Learning.

Dr.S.Sathya Priya is an Associate Professor in School of Computing Sciences, Hindustan Institute of Technology and Science, Chennai, India. She received her Ph.D in Information and Communication Engineering from Anna University, Chennai, India, M.E. in Computer Science and Engineering and B.E in Electronics and Instrumentation from Annamalai University, Chidambaram, India.She has a work experience of around 12 years and her areas of expertise include Wireless Networks, Network Security and Machine Learning.She has nearly 15 publications in reputed International Journals and International Conferences to her credit.

Mr.R.J.Poovaraghan, received his B.E. degree from the University of Madras 1997. M.E. degree from Sathyabama University 2012. He has 20 years of teaching experience and currently he is working as an Assistant Professor (SG) in the Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, India. He has published more than 25 research papers in international journals and conferences. His research interest is image processing, computer vision, Artificial Intelligence, Machine Learning and Deep Learning.

Ms.B.Pavithra, working as Assistant Professor in the Department of Computer Science and Engineering at Hindustan Institute of Technology & Science, Chennai, India. She received B.Tech degree in 2013 from Anna University and received M.E.degree in Computer Science Engineering (Specialization in Networks) in 2016 from Anna University, Chennai. She has published more than 7 papers in reputed International Journals and Conferences. Her research areas of interest include Cloud Computing, Wireless sensor Networks, MANET, Artificial Intelligence and Machine Learning.

Ms.J.Mercy Faustina received the Bachelor of Engineering in Computer Science and Engineering from Panimalar Institute of Technology, Anna University, Chennai in 2015. Completed her Masters in the Department of Computer Technology at Anna University-MIT Campus, Chennai. She is working as an Assistant Professor in Hindustan Institute of Technology, Padur. Her research interest includes Mobile Networks, Knowledge & Data Engineering.