A Novel Machine Vision-Based 3D Facial Action Unit Identification for Fatigue Detection

Gulbadan Sikander and Shahzad Anwar

***Abstract*— Fatigue has been attributed to traffic accident with higher fatality rate and causes severe damage to the surroundings compared to accidents where drivers are alert. This study presents and demonstrates an innovative driver fatigue detection method based on fatigue related facial action units' (AU) identification employing a photometric stereo (PS) testbed for 3D AU reconstruction. Initially, normal vectors were extracted for 3D AUs, subsequently, 'Quiver/Bump map' were constructed from the normal vectors. The quiver maps were further utilized for training deep neural networks. The findings exhibit that the proposed method outperforms 2D image based classification in terms of validation accuracy, for AU1, AU15 and AU41 detection. A novel method incorporating machine vision for intelligent transportation has been developed and demonstrated for driver's fatigue detection. The proposed method is also compared to other established methods and significant (95%) improvement in terms of accuracy is achieved.**

***Index Terms*— Intelligent transportation, machine vision, fatigue detection, facial action units, 3D reconstruction, photometric stereo.**

I. Introduction

DRIVING is a complex phenomenon; involving performance of various tasks including, simultaneous quick and accurate decision making. Fatigue drastically decreases human response time, which leads to human inability to drive efficiently. Fatigue symptoms also evolve as performance deteriorates.

In recent years, road accidents have drastically increased and in a US survey [1] it is stated that 4,121 fatal crashes between 2011-2015 involved driver compromised by fatigue. The same study revealed that in 2017 fatigue related crashes caused 795 deaths and over 90,000 vehicle crashes. As per The Royal Society for the Prevention of Accidents [2], 20% road accident in the UK were associated to driver fatigue and fatigue related accidents were reported to be 50% more likely to result in death or serious injury.

Fatigue is categorized into active, passive and sleep related fatigue [3]. Mental depletion due to active engagement in a task is active fatigue. Monotonous task or inattention causes passive fatigue. Sleep related fatigue is caused by the circadian rhythm [4]. Circadian rhythm, a 24-hour sleep/wake cycle, for adults, the largest dip in energy is experienced at midnight (02:00 to 04:00 hours) and midday (13:00 to 15:00 hours) and if driving is performed in between these time slots, there is a high chance of experiencing sleep-related fatigue. Even though fatigue has been categorized in three categories, the combination of any two or all, also leads to fatigue.

The alarming increase in accidents seek research community's serious attention to address the issue via developing a driver fatigue detection system in order to reduce accidents and improve transportation safety.

This study is motivated by developing a novel fatigue detection method which is capable of accurately detecting fatigue well in advance. The study was conducted on a custom designed test bed (in a laboratory environment), as real road test involves safety concerns [5]. The main contributions of this research are: (i) proposing a novel facial muscle based fatigue detection system which is developed and evaluated on a custom designed test bed; (ii) 3D representation of the deduced fatigue related facial action units; (iii) implementing four different convolutional neural networks (CNNs) for the detection of driver fatigue related facial action units;(iv) developing 3D based fatigue action unit image dataset using the testbed.

The remaining of this manuscript is organized as following. Section II presents a review of existing driver fatigue detection methods and establishes the need for this study. Section III presents the overall methodology of driver fatigue action unit detection method. Section IV elaborates the experimental setup and test bed development. Section V discusses the results. Finally, Section VI concludes this article.

II. Related Work

Driver fatigue detection methods could be categorized into subjective reporting, biological, physical, vehicular and hybrid features based driver fatigue detection systems. A more detailed state of the art review is presented in [6].

A. Subjective Reporting

Subjective reporting is one of the earliest technique in drowsiness detection. Karolinska Sleepiness Scale (KSS) [7] is incorporated as a benchmark for validating other's systems accuracy, which are under development.In a recent study [8] drowsiness is considered a by-product of sleep related fatigue, therefore; KSS was incorporated in fatigue detection [9]. One of the limitation is subjective reporting being prone to interference from frequent reporting and erroneous results due to drivers' deteriorating ability to efficiently judge their fitness

Manuscript received May 8, 2019; revised September 12, 2019 and December 5, 2019; accepted February 10, 2020. The Associate Editor for this article was L. M. Bergasa. *(Corresponding author: Gulbadan Sikander.)*

The authors are with the Department of Mechatronics Engineering, University of Engineering and Technology, Peshawar 25000, Pakistan (e-mail: gulbadan@uetpeshawar.edu.pk; shahzad.anwar@uetpeshawar.edu.pk).

Digital Object Identifier 10.1109/TITS.2020.2974263

after few hours of driving activity making its applicability limited.

B. Biological Feature Based Fatigue Detection

Biological signals, such as, electroencephalography (EEG), electrocardiogram (ECG), electro-oculography (EoG) and surface electromyogram (sEMG) have been widely utilized in driver fatigue detection systems. Sun and Yu [10] utilize non-intrusive electrodes through clothes to measure ECG, EEG with electrodes attached to a cap and EoG from an electrode hanging from the vehicle roof. Eyelid activity, EEG and HRV were recorded to deduce fatigue. It was deduced that blink duration and frequency increase with fatigue, power density of alpha and beta wave decrease with fatigue and LF/HF and SDNN decrease with fatigue while RMSSD, LF and HF increase with fatigue.

Biological features are considered as an accurate fatigue indicator, however, sensors for data acquisition are highly intrusive. Heart rate and its variability, activity in alpha EEG wave, electric muscle activation and corneo-retinal potential difference between back and front of the eye have been utilized in numerous driver fatigue detection studies [11]. As biological sensors are complex, expensive and requires much data pre-processing (to avoid noise). Moreover it is prone to driver movement, therefore it finds limited applicability in real time driver fatigue detection systems.

C. Vehicular Feature Based Fatigue Detection

Fatigue compromises driver's performance, leading to deviation in driving features, such as, lane crossing and steering wheel angle (SWA). Extensive research has been performed on SWA based fatigue detection and it has been classified as an accurate indicator of fatigue. Research indicates high correlation in SWA and fatigue. Li et al. [12] present a back propagation based neural network approach, where SW and yaw angles were monitored for driver fatigue detection, achieving average accuracy of 88.02%. Personal driving attributes affects vehicle-based fatigue features along with weather and traffic conditions.

D. Physical Feature Based Fatigue Detection

Physical features include blink frequency, eye closure duration, percentage of eyes close (PERCLOS), pose, gaze, yawning and nodding frequency. Omidyegane et al. [13] presented a yawn detection technique. Face and mouth regions were identified in the input image via Viola Jones [14], a back propagation technique was introduced for yawn detection achieving 75% accuracy. In a recent approach [15], SVM based fatigue detection system was designed specifically for bus drivers through estimating eye openness and PERCLOS. Eye openness is calculated using linear and spectral regression. Where fatigue detection works for low resolution images however, the technique is designed for the dome camera installed in buses and is not applicable to all vehicles.

E. Hybrid Feature Based Fatigue Detection

Recently, research has been geared towards the fusion of various features for fatigue detection. In [16] features such as ECG, EEG, sEMG, SWA, lateral position and PERCLOS were fused for fatigue classification. According to their study, it is desirable to only incorporate features that could effectively deduce fatigue. In a comparative analysis presented in [17], biological, physical and vehicular features were compared to deduce most effective features for fatigue detection. Artificial neural networks were trained with either biological, physical, vehicular or all features to deduce fatigue. Additionally parameters like driving time and participant information were also recorded. It is concluded that the physical features, drive time and personal information fusion produce enhanced fatigue detection accuracy. Sun *et al.* [18] discussed contextual features, driver facial features, and vehicle behaviour features employment for driver fatigue detection. Three separate multiclass support vector machine (MCSVM) classifiers were trained to detect fatigue. Best result was achieved with the fusion of all three type of features. It has been established from review that image based methods are commonly used for fatigue detection however, the fatigue detection systems are mostly been geared towards single image based methods. As 3D shape of the face will provide minute details which would be viewpoint and illumination invariant. Additionally, it will provide new information previously not accounted for driver fatigue detection. The novelty of this research lies in the fact that changes caused by fatigue in face texture has not been explored prior to this research. This study aims at the development of a driver fatigue detection system utilising 3D facial action units. Due to the unavailability of any existing 3D fatigue action unit dataset, a customized dataset was developed for this study using seven light source Photometric Stereo testbed. Additionally, performance of four transfer learning based deep networks is compared in terms of validation accuracy. The proposed method is also compared with existing methods and significant improvement is terms of accuracy is achieved.

III. Materials and Method

This research study presents a novel drive fatigue detection method. The proposed method incorporates minute facial muscle changes to deduce fatigue. To explore minute facial changes, facial action coding system is employed. Through experimentation suitable fatigue related action units are short-listed. The information is not accurately depicted in 2D images as the changes are very subtle. Three dimensional (3D) representation of facial muscles is incorporated as 3D shape of the face provide minute detail previously not observed in 2D images. Plentiful of 3D reconstruction techniques are available in literature, in this research Photometric Stereo is incorporated as it has the ability of providing accurate detailed results. 3D information is presented in the form of pixel-wise normal vectors (bump map), these normal vectors represent the surface orientation. The minute changes in facial muscles is accurately translated into the direction of the normal vectors. This bump map data was fed to the deep networks for fatigue detection. Once images for Photometric Stereo are captured, three regions are extracted from the images as they contain the desired fatigue related action units. After region of

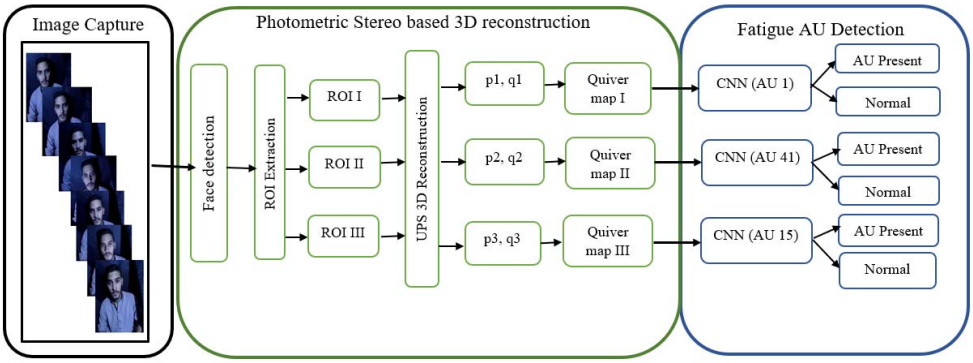


Fig. 1. Proposed framework.

interest extraction, photometric stereo based 3D reconstruction is performed to obtain the normal vectors. The normal vectors are represented in the form of quiver/bump maps and this data is used to train the deep networks. Independent networks are trained to detect each action unit and results are classified into action unit present and normal where the presence of action unit represents fatigue state. The input layers of the networks is configured to accept quiver maps as input data, as in its original state the networks are trained on image data. Four different deep networks are tested in this study, to show that the proposed method performs well under different circumstances.

The details of the framework of the proposed 3D machine vision based facial action unit identification method for fatigue detection (as shown in fig. 1) are presented in the following subsections.

A. *Photometric Stereo and 3D Reconstruction*

Three dimensional shape of an object provides detailed properties that are invariant to changes caused by the imaging process, such as, illumination and viewpoint. 3D reconstruction is the estimation of shape and appearance of an object. Many 3D reconstruction techniques are available in literature including shape from shading [19], shape from focus [20], point cloud [21] and Photometric Stereo (PS) [22].

PS was initially introduced by Woodham [22] in 1980. PS applies reflection models to recover surface shape and albedo. PS requires at least three images of an object taken from the same viewpoint but under different illumination conditions. PS is capable of providing highly accurate detailed results, hence PS is incorporated in the proposed method for driver facial 3D reconstruction. Dynamic PS is presented by Smith and Smith [23] to recover 3D shape of a moving object. PS has been incorporated for recognition and verification [24], [25] however, limited/no research has been performed to incorporate PS in driver monitoring and fatigue detection methods.

Three source PS recovers both local surface orientation and albedo given the assumption that shadow and specular factors are absent. However, in real world application, surfaces may not be Lambertian and could contain a specular component. There could be some points that will be in shadow for one or more of the images. Shadows arise in real images in two possible scenarios. Firstly, by attached shadows where some

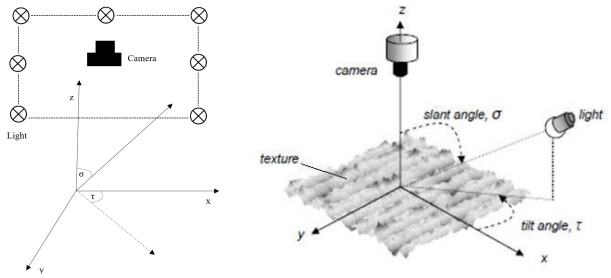


Fig. 2. (a) Experimental setup with 7-lights, and (b) Photometric Stereo schematic diagram [26].

pixels are not visible in image as the object may be facing away from the light source and such regions will appear dark. Secondly, shadows could occur once shape of an object is not convex and parts of surface is occluded by other parts, these shadows are referred as cast shadows. Irrespectively, all shadows occur in an image as dark pixels with almost zero intensity values. More than three light sources could overcome the effect of shadows. Hence, in this study seven light sources were employed for image capture; the additional lights were included to reduce shadow artefacts, as depicted in Fig.2(a). Images with shadow in the ROI are omitted from the reconstruction process as shadows cause erroneous results. The image selection process is currently manual and an intelligent automatic technique would be presented in the future. Light direction is defined by slant and tilt angles. The slant angle is between illumination vector and z-axis and tilt angle is between x-axis and projection of illumination vector onto x-y axis as shown in Fig. 2(b).

Surface normal vectors were calculated by solving the irradiance equations, and form the 3D surface description. Pixel orientation is governed by surface normal ‘n’ which is perpendicular to the surface at that point. Surface normal is represented by surface gradients ‘p’ and ‘q’. Texture could be described by a height function z(x, y) and the surface gradients are as given by eq. 1.

$$p = \frac{\partial z(x,y)}{\partial x}$$
$$q = \frac{\partial z(x,y)}{\partial y} \quad (1)$$

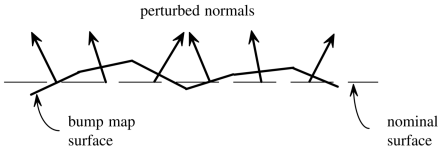


Fig. 3. Surface bump map [23].

Quiver/Bump maps are directed graphs that represent photometrically captured 3D surface. Quiver maps are composed of surface normal vector at each pixel. Bump map is the description of the surface topographic texture ignoring the surface colour. A representation of a bump map is shown in fig. 3.

According to Lambertian model the relationship in between Light intensity (I), surface albedo (ρ), the normal vector (N) and light direction (S) can be written as in eq. 2.

$$I = \rho N S \tag{2}$$

Writing $M = \rho N$, the model can be converted to linear system where I=M.S. If S is known one can recover both ρ and N with eq. 3.

$$\begin{aligned} M &= S^{-1}.I \\ N &= \frac{M}{||M||} \quad \rho = ||M|| \end{aligned} \tag{3}$$

For complete 3D reconstruction accurate light directions are highly desirable. For traditional PS position of the object with respect to light sources and camera should not change. However, traditional PS has limited applicability for this study as the human subject may move causing erroneous light directions and subsequently, incorrect 3D reconstruction. A possible solution is to incorporate uncalibrated PS as the light directions are estimated during the reconstruction process.

Uncalibrated Photometric Stereo (UPS) is the estimation of surface reflectance and surface normals without priori knowledge of the light source direction and intensity. As light directions are unknown, therefore, estimation has to be performed on both M and S [27]. Least square sense using singular-value decomposition [28] could be employed to calculate M and S that satisfy $I = MS$. However, it doesn't provide a unique solution. For the estimation of N the ambiguity has to be reduced to generalized bas-relief transformations [29] by imposing integrability constraint as described in eq. 4.

$$\overline{curl}N = 0 \quad where\ \overline{curl}[a,b,c] = \frac{\partial}{\partial x}\frac{a}{c} - \frac{\partial}{\partial x}\frac{b}{c} \tag{4}$$

As both the Lambertian assumption and the integrability constraint should hold so, $M' = MG$ and $S' = G^{-1}S$ must hold true where, G is the generalized bas-relief transformation matrix. G and G^{-1} can be represented as shown in eq. 5.

$$\begin{aligned} G(\mu,\nu,\lambda) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & \nu & \lambda \end{pmatrix} \\ G^{-1}(\mu,\nu,\lambda) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{\mu}{\lambda} & -\frac{\nu}{\lambda} & \frac{1}{\lambda} \end{pmatrix} \end{aligned} \tag{5}$$

Total Variation (TV) was incorporated to estimate the values of μ, λ and ν. TV of a function is a technique used for regularization. TV is defined in eq. 6, $||J(f)_F||$ presents Frobenius norm of Jacobian matrix of function f.

$$TV(f) = \int_{R_2} ||J(f)||_F dx \tag{6}$$

The values of μ, λ and ν minimizes the total variation of function TV(MG(μ, ν, λ)). Standard convex optimization method is employed to perform minimization.

B. Face Region Detection and 3D Reconstruction

Face detection and recognition have always attracted researches attention [30]. In this study face region extraction is performed via cascade object detector [14]. Skin based segmentation is performed on the face image region to remove as much of the background pixels as possible. Advanced convex optimization techniques [31] are incorporated to remove the effect of sparse, gross errors like corrupted and missing pixels with low-rank matrix completion. UPS performs 3D reconstruction after the correction process.

C. Facial Action Coding System

Facial Action Coding System (FACS) is one of the most widely used technique for facial expression recognition [32]. FACS encodes facial expression into 46 facial muscle movements also known as Action Units (AUs). Various AU combination result in a unique expression. For instance a smile expression is considered as a combination of "pulling lip corners (AU 12+13), mouth opening (AU 25+27) with upper lip raiser (AU 10) and furrow deepening (AU 11)" [33]. In this study investigation was performed on discovering potential fatigue related AUs that could lead to in time fatigue detection.

D. Fatigue Related Facial Action Units

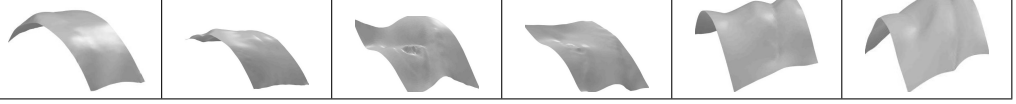
FACS has become a benchmark for human expression recognition however, literature on human fatigue recognition is limited. Vural et al. [34] proposed FACS based driver fatigue detection. Support Vector Machine (SVM) was incorporated to detect the presence of AUs. One SVM was trained for each AU, subsequently, leading to the conclusion that Blink (AU45), Inner Brow Raise (AU1),Tongue show (AU19) and Jaw Drop (AU26) are good fatigue indicators. Their study focus on AUs present 60 seconds prior to any crash.

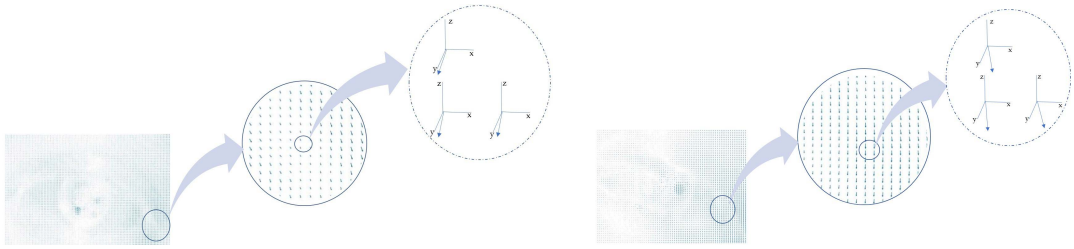
Facial action units associated with yawning (AU25,26,27) and blink rate (AU45) are commonly incorporated in driver fatigue detection systems, while other potential AUs are relatively less explored.

All AUs were taken into consideration in this study. Subsequent rigorous experimentation revealed that AU1, AU15 and AU41 are good indicators of early fatigue. AU1 presents changes on the forehead caused by eye brow raise, the AU is caused by changes in Frontalis (pars lateralis) muscle. AU15 represents corner mouth depressor, when the driver's mouth corners are pulled downwards, it is caused by the Depressor anguli oris muscle. AU41 presents eyelid droop, when the drivers eyes are relaxed due the action of Levator palpebrae superioris muscle.

TABLE I

IMAGE REPRESENTATION OF ROIS WITH EITHER AU PRESENT OR ABSENT

	ROI 1		ROI 2		ROI 3	
	Normal	AU1	Normal	AU41	Normal	AU15
Images						
Quiver						
3D						



(a) (b)

Fig. 4. (a) Normal eye and (b) AU 41 eye bump maps for pixel positions (761, 350), (762, 350) and (762, 351).

E. Region of Interest Extraction

Fatigue symptoms are prominent on the human face in areas, such as, forehead, eyes and lip corners, therefore these region of interest (ROI) were selected for experimentation. ROI 1 contains AU1, ROI 2 has AU41 while ROI 3 includes AU15. The method for ROI extraction is based on our previous work presented in [35], while [36] presents a fuzzy based facial feature tracking method. Example of training images is represented in table I. The previously proposed method [35] incorporates the Viola-Jones [14] method for the identification of possible ROIs and the subsequent calculations are based on human face symmetric properties for ROI detection. In this research right eye, mouth and forehead detection were performed. Eye and mouth detection have been performed with cascade object detectors trained on Haar features. Forehead is defined as the region included in the face area above the eye region. These ROIs data were fed to train and test deep learning method for fatigue AU detection.

F. 3D Facial Action Units

2D facial action unit detection draws the attention of researchers due to its potential application in fatigue detection [37], [38]. The existing research has been performed in 2D only and 2D data has limitations caused by head pose and illumination. 3D has the ability to capture true facial surface data and hence include more information.

3D reconstruction of the facial regions (mentioned in section III-E) was performed and vector p and q were extracted. From the 'p' vector and 'q' vector 'bump maps' were formed which were utilized to train the deep networks.

TABLE II

SLANT AND TILT ANGLES FOR THREE POSITIONS ON XY AXIS; VALUES ASSOCIATED WITH FIG. 4

Pixel position	Normal slant & tilt angles	Fatigue slant & tilt angles
(761,350)	95° & 87°	175° & 27°
(762,350)	97° & 88°	172° & 32°
(762,351)	94° & 85°	176° & 29°

Quiver maps are directed graphs, it displays vectors 'p' and 'q' as arrows. Quiver map/Bump maps represents the surface normal at each pixel. The surface normals can be analysed with separate analysis of the slant and tilt angles. The tilt angle τ and slant angle is α as shown in fig 2(a). The normal vector can be separated in the slant and tilt gradients (based on τ and α), also known as the surface gradients.

It was found that increased disruptions in the surface gradients (tilt and slant direction of surface normals), was associated with the presence of an action unit, as shown in fig. 4. Table II illustrates slant and tilt angles for normal and AU41 in ROI II. Under normal state the n vector points towards positive y-axis, when fatigued the n vector changes direction, subsequently, substantial change can be observed in tilt and slant angles. On the other hand, in 2D image the slight changes occurring in the ROIs (in between normal and early fatigue) are not prominent. It shows that 3D representation in the form of

TABLE III
CNN ARCHITECTURE FOR EXPERIMENTATION

	[39]	[40]	[41]	[42]
Input	Image	Image	Image	Image
Network Layers	Conv3-48 Conv3-48* Max pooling layer Conv3-128 Conv3-128* Max pooling layer Conv3-192 Conv3-192* Conv3-192 Conv3-192* Conv3-128 Conv3-128* Max-pooling layer FC layer-2048 FC layer-2048* FC layer-1024 FC layer-1024* FC layer-2	Conv7-64 Max pooling layer Conv3-192 Max pooling layer Inception3-256 Inception3-480 Max-pooling layer Inception4-512 Inception4-512 Inception4-512 Inception4-528 Inception4-832 Max-pooling layer Inception5-832 Inception5-1024 Average-pooling layer dropout(40%) FC layer-2048 FC layer-1024 FC layer-2	Conv7-64 Max pooling layer 3 [conv1-64 conv3-64 conv1-256] [Conv1-128 conv3-128 conv1-512] 7 [conv1-128 conv3-128 conv1-512] [conv1-256 conv3-256 conv1-1024] 35 [conv1-256 conv3-256 conv1-1024] [conv1-512 conv3-512 conv1-2048] 2 [conv1-512 conv3-512 conv1-2048] Average-pooling layer FC layer-2048 FC layer-1024 FC layer-2	Conv7-96 Max pooling layer Fire55-128 Fire55-128 Fire55-256 Max pooling layer Fire27-256 Fire27-384 Fire27-384 Fire27-512 Max pooling fire13-512 conv1-2 Average pool 13-2

* Parallel stream layers

surface gradients of the face surface shows minute changes that are otherwise not visible in 2D images hence, provides sensitive information for early fatigue detection.

G. Deep Learning Based Driver Fatigue Detection

To develop a real-time fatigue detection system, Convolutional Neural Network (CNN) is utilized in this study, four CNN architectures were introduced and compared in terms of accuracy to identify the algorithm that is best suited for implementation. Table III shows the network architecture of the deep learning networks under consideration in this study. The subsequent subsections explain the CNN models, more information can be found in [5], [39]–[42].

1) AlexNet: AlexNet network [39] includes two streams of five convolutional layers. Some of the convolutional layers are followed by max-pooling layers and fully connected layers. Dropout layers are included to prevent overfitting in the fully connected layers. The network has the ability to classify into 1000 categories. AlexNet has 60 million parameters and 650,000 neurons and takes $224 \times 224 \times 3$ images as input.

2) GoogleNet: GoogleNet [40] designed by Google is a 22-layer CNN with local network topology. GoogleNet introduced the concept of local Inception module in CNN, which finds optimal local construction and spatially repeats it. This allows the increase in number of units at each stage without considerably increasing computational complexity. GoogleNet takes $224 \times 224 \times 3$ images as input and has the ability to classify into 1000 categories. An Inception module is formed by the combination of 1×1, 3×3, and 5×5 convolutional layers with 3×3 max pooling layers.

3) ResNet: ResNet [41] designed by Microsoft Research is based on VGG network. The layers are reformulated as learning residual functions and replace unreferenced learning functions. ResNet takes $224 \times 224 \times 3$ images as input and has the ability to classify into 1000 categories. If the output feature map is of the same size then the convolutional layers have same number of filters, however, if the feature map is reduced to half the size, filters are doubled in number so, the time complexity remains the same. Output layers include a global average pooling layer and a fully connected softmax function.

4) SqueezeNet: Squeezenet [42] has the same level of accuracy as AlexNet with 50 times less parameters. SqueezeNet requires less communication across servers, lesser exportation bandwidth and are more feasible to be deployed on hardware with limited memory due to the smaller architecture size. Further reduction could be achieved by replacing the 3×3 filters by 1×1 filters, for efficient calculation.

5) Model Training: Extensively trained deep neural networks could be customized to fit a specific application through transfer learning [43]. Transfer learning grants the ability to train and test on different tasks, domains and distribution. Transfer learning can be extensively observed in the real world. For example, learning to recognize peaches would aid in identification of tomatoes. The motivation behind transfer learning is the human ability to apply knowledge learned previously and solve new problems. The fully connected layer and output classification layer were replaced by new layers for GoogleNet, AlexNet, ResNet and SqueezeNet. The input layers for all the networks are configured to accept pixel-wise

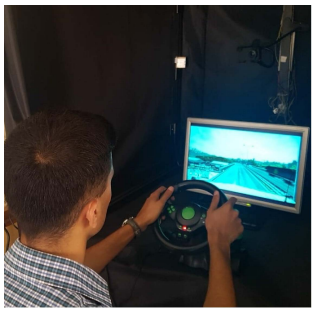


Fig. 5. Photometric stereo hardware set-up.

normal vector data (in the form of matrices) in the proposed method. The output of the new fully connected layer equals to two classes (i.e. normal and AU present). Classification layer defines output classes of the network hence, it is also changed to output two classes. For SqueezeNet the final convolution layer and classification output layer were replaced. The number of filters in the final convolutional layer are kept the same as the number of classes.

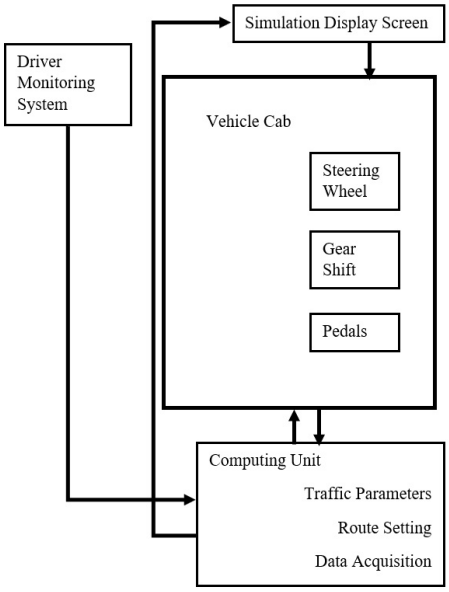


Fig. 6. Schematic diagram of the simulator.

IV. Experimentation

A. Hardware Setup

Fatigue AU detection was performed employing PS setup in a laboratory environment. Fig. 5 shows 3D facial feature reconstruction setup, which consists of seven light sources and a single camera for driver monitoring. The device is designed for practical driver face 3D geometry. The frame rate is high enough so, effect of inter frame motion can be minimized. Interfacing code is performed using Matlab. Light source switching was performed in coordination with Arduino UNO microcontroller board with ATmega328P microcontroller. Imaging device (4K IP) is incorporated (in the hardware) for image capture. Seven images are taken sequentially by the imaging device in coordination with the light sources.

The simulator platform offers an interactive and high resolution environment with fixed base driving. The simulator has been specially designed for the research study. The vehicle cab includes a steering wheel, gear shift and throttle,brake and clutch pedals. A projection screen displays the drive simulation in coordination with the computing unit and the vehicle cab. The computing unit controls the traffic parameters, route setting and data acquisition. The computing unit also communicates with the driver monitoring systems. Fig. 6 illustrates the various parts of the driving simulator.

B. Test Protocol and Database

In this study 80 participants (65 male, 15 female) between the ages of 24 to 40 were invited for experimentation. Test scenario was set at 60 miles with both straight and curve road sections. The experiment requires a long and monotonous drive segment to induce boredom and fatigue. The participant drove the simulator for four days, each day there were three test segments; 0900 to 1000, 1400 to 1600 and 2300 to 0100. The subjects were advised to sleep at night but remain wake from 0700 till the last test phase. Throughout the day the participants (drivers) were advised to follow their usual daily activities. During the experiments the drivers were advised to keep a safe distance from other vehicles and to follow the speed limits. Practice session were held for all participants prior to the commencement of the experiments to familiarize them with the simulator enviroment. In the morning session, drivers only drove for one hour to avoid fatigue induced due to road monotony. In the afternoon and night sessions, the participants arrived a couple of hours prior to start of the experiment. The experiments were repeated for four days for each participants to generalize the data. Images were captured in various poses; the following triggers examined: (i) sleep deprivation, (ii) extended duration of wakefulness and (iii) time of day (circadian rhythm effect). This research study is approved by the ethical body and all participants were briefed with the objective of the study. Images were captured at various intervals such as (i) when participant is alert and (ii) when early fatigue sign exhibits. Visual cues, questionnaire based self classification and circadian rhythm are deployed as ground truth. The participants own estimation of fatigue level is estimated by the KSS questionnaire [7]. Images were taken of the participants on different days to generalize the data. Images are recorded for each participants during active and fatigued stages for AU1, AU15 and AU41. Hence, about 25,000 different data matrix are available, as each data matrix is recorded as 'images' and 'quiver maps' making the dataset of about 50,000.

To test the capability of the proposed method, four CNN models were trained on a machine with the following configuration: Intel i7 7700HQ quad processor, 16 GB RAM and NVIDIA GeForce GTX 1050 GPU. Often overfitting occurs, which was reduced by incorporating Stochastic Gradient Descent with Momentum (sgdm) having a learning rate of 3×10 e-4 and momentum of 0.9. The minimum batch

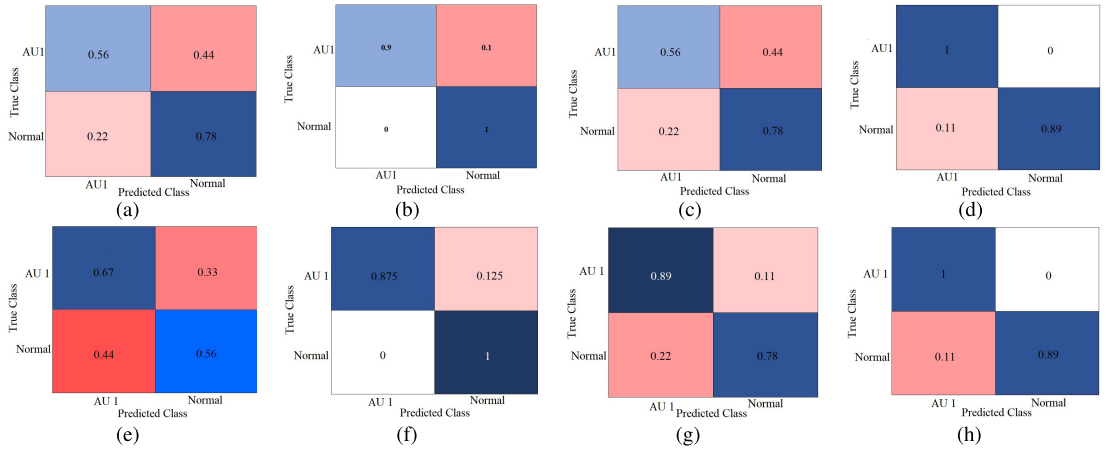


Fig. 7. (a) Confusion matrix for AU1 detection AlexNet trained on 2D images (b) Confusion matrix for AU1 detection AlexNet trained on bump maps (c) Confusion matrix for AU1 detection GoogleNet trained on 2D images (d) Confusion matrix for AU1 detection GoogleNet trained on bump maps (e) Confusion matrix for AU1 detection ResNet trained on 2D images (f) Confusion matrix for AU1 detection ResNet trained on bump maps (g) Confusion matrix for AU1 detection SqueezeNet trained on 2D images (h) Confusion matrix for AU1 detection SqueezeNet trained on bump maps.

size was kept at ten and the maximum epochs were forty with data shuffle at each epoch. Data was divided into training and validation sets to evaluate the performance of the models, 70% of the data was reserve for training and 30% of the data was kept hidden in training stage for validation. The driver's data at validation stage were different from driver's data at training stage to avoid a falsely high test accuracy. The following section shows the performance of each of the four CNN models.

V. Result and Discussion

The developed technique introduces 3D data as input to the CNN models. The effectiveness of 3D representation for fatigue AU detection is accessed and comparison is performed with other techniques.

A. Action Unit 1 Detection

2D images and quiver maps obtained by the proposed method were utilized to train Alexnet, GoogleNet, ResNet and SqueezeNet respectively. The results show that the developed method outperforms 2D images for all networks. AlexNet trained on 2D images had an accuracy of 67% where, 56% AU1 were correctly identified, 2D image trained GoogleNet had a validation accuracy of 67% where, 44% AU1 were misidentified. ResNet trained on 2D images had 61.5% validation accuracy, where 67% AU1 were correctly identified, however, 44% normals were misidentified. SqueezeNet should similar performance with validation accuracy of 83.5% accurate where, 89% AU1 were correctly identified, however, 22% normals were misidentified.

SqueezeNet trained on quiver map had a validation accuracy of 94.5%, all the AU1 were correctly distinguished. Quiver map trained ResNet had validation accuracy of 93.75% and all the normal were correctly identified. GoogleNet displayed similar performance 94.5% and all the AU1 were correctly identified. The developed method outperformed with a validation accuracy of 95%, all the normals were correctly identified by AlexNet trained on quiver map. The confusion matrices for all the networks are presented in fig. 7.

B. Action Unit 15 Detection

AlexNet, GoogleNet, ResNet and SqueezeNet were respectively trained on 2D images and quiver maps obtained by the developed method. The confusion matrices are shown in fig. 8.

The developed method outperformed 2D image trained CNNs. The developed method trained on AlexNet was able to identify all AU15 correctly, 89% of AU15 were correctly detected with GoogleNet, 78% correct identifications with ResNet and 82% with SqueezeNet. On the other hand, with 2D images as input 70% AU15 were correctly identified with AlexNet, 80% identification rate with GoogleNet and only 58.3% with SqueezeNet. AlexNet trained on quiver maps obtained by the developed method outperform all the others where all the AU15 were correctly identified and 89% of normals of also identified correctly.

C. Action Unit 41 Detection

For AU41 detection AlexNet was trained on 2D images and quiver maps obtained by the developed method. Developed method was able to correctly recognize all the AU41 and 80% of normals. 2D image trained AlexNet was also able to correctly identify all the AU41s however, 50% of normals were misclassified. The validation accuracy of GoogleNet for 2D images and developed method was 60% and 95% respectively, here again the developed method was able to identify all AU41 correctly. For AU41 identification ResNet had the worst performance, The accuracy of 2D images was 60% and of quiver maps was 74.65%, still the developed method performed better than 2D images. In image trained CNN 90% AU41 were located correctly however, 70% of the normals were misidentified. Developed method performed better as 80% AU41 were correctly recognized and 70% normals were correctly identified.

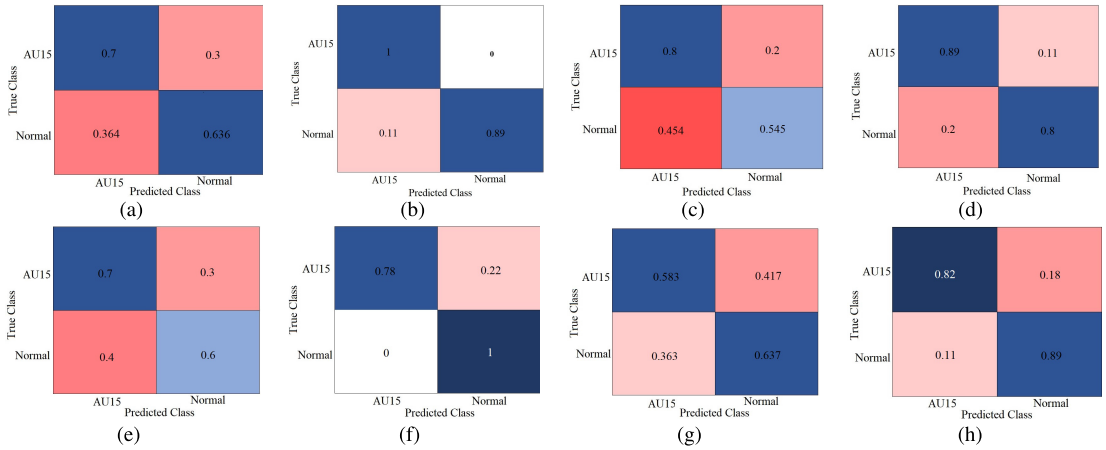


Fig. 8. (a) Confusion matrix for AU15 detection AlexNet trained on 2D images (b) Confusion matrix for AU15 detection AlexNet trained on bump maps (c) Confusion matrix for AU15 detection GoogleNet trained on 2D images (d) Confusion matrix for AU15 detection GoogleNet trained on bump maps (e) Confusion matrix for AU15 detection ResNet trained on 2D images (f) Confusion matrix for AU15 detection ResNet trained on bump maps (g) Confusion matrix for AU15 detection SqueezeNet trained on 2D images (h) Confusion matrix for AU15 detection SqueezeNet trained on bump maps.

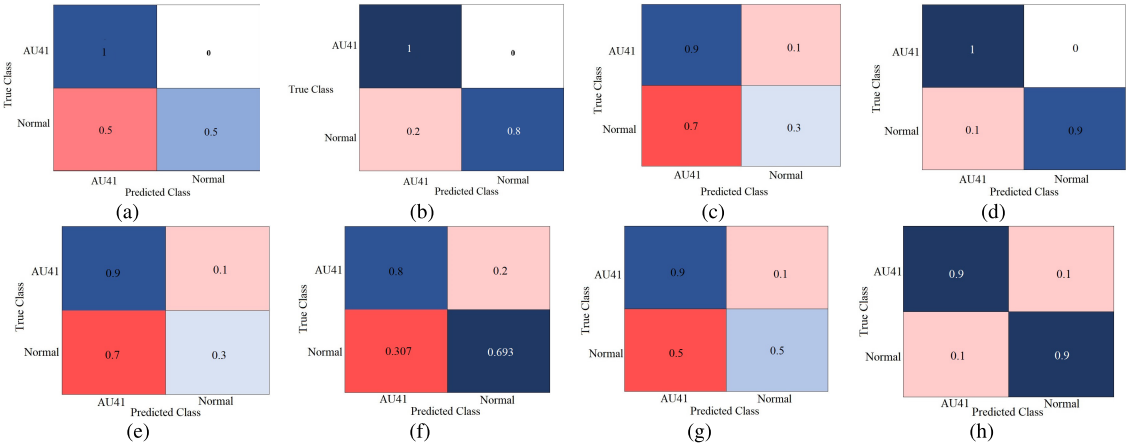


Fig. 9. (a) Confusion matrix for AU41 detection AlexNet trained on 2D images (b) Confusion matrix for AU41 detection AlexNet trained on bump maps (c) Confusion matrix for AU41 detection GoogleNet trained on 2D images (d) Confusion matrix for AU41 detection GoogleNet trained on bump maps (e) Confusion matrix for AU41 detection ResNet trained on 2D images (f) Confusion matrix for AU41 detection ResNet trained on bump maps (g) Confusion matrix for AU41 detection SqueezeNet trained on 2D images (h) Confusion matrix for AU41 detection SqueezeNet trained on bump maps.

For AU41 detection 2D images show comparative results to the developed method however, the performance deteriorates considerably for normal detection. For 2D images 90% AU41 were correctly located, however, half of the normals were misclassified. Developed method outperformed with a validation accuracy of 90%, where, 90% AU41 and normals were correctly recognized. The confusion matrices are shown in fig. 9.

A summary of the results is presented in table IV. Deep networks trained with developed method have considerably better validation accuracy compared to deep networks trained on 2D images. The developed method outperformed with all CNN architectures. This proves that the developed method would perform better than 2D images under all deep networks platforms. It is shown that AlexNet performs better for AU1 and AU15 detection where as, GoogleNet performs best for AU41 detection. 3D information in the form of quiver maps gives pixel level information related to surface orientation. Previously, minute detailed information was not explored for fatigue detection. Pixel level detail provides information to detect initial stages of fatigue, details which are not visible in 2D images. Therefore, 2D images fail to accurately show fatigue related AUs.

TABLE IV
VALIDATION ACCURACY (IN %) OF TRANSFER LEARNING BASED CNN

Feature	Data Type	Evaluation Method [39]	[40]	[41]	[42]
AU1	2D Images	67	67	61.5	83.5
	Developed Method	95	94.5	93.75	94.5
AU15	2D Images	66.8	67.25	65	61
	Developed Method	94.5	84.5	89	85.5
AU41	2D Images	75	60	60	70
	Developed Method	90	95	74.65	90

D. Performance Evaluation: A Comparison

To test the capability of the developed technique, comparison was made with existing methods such as, [13], [15]

TABLE V
PERFORMANCE COMPARISON WITH EXISTING METHODS

Method	Feature	Technique	Accuracy
[13]	Yawn	Back propagation	75%
[15]	Eye Openness	Linear & spectral regression	90.1%
[18]	CF FF VBF	MCSVM	93.5%
Developed Method	3D based Quiver maps	deep network	95%

and [18]. The developed method outperforms the existing approaches (comparison is shown in table V). The developed method incorporates unique features for driver fatigue detection that have not been explored prior to this research study. Much research has already been performed on eye activity and yawn detection, therefore, the possibility of advancement with the said features is limited and new avenues are being explored in this research.

Quiver maps (as defined in section III) also known as bump maps are direct representation of the direction vectors and consist of only the direction related data. This proves that albedo carries no valuable information regarding fatigue action unit detection.

The proposed novel method offers minute detail previously absent in fatigue monitoring. The proposed non-invasive scheme is based on machine vision techniques and explored the analysis of surface normal data for AU detection. It also demonstrated new valuable and potentially complementary 3D indicators, which could be employed for early fatigue detection. There have been increasing demands for a non-invasive computer vision based system which would offer the benefit of detailed descriptions of fatigue AUs. Such a system would be a combination of multi-disciplinary engineering, such as, computer vision, machine learning, convolutional neural networks and facial muscle movement. The proposed system could find application in many areas, for instance, driver monitoring, emotion recognition and duty fitness monitoring.

VI. CONCLUSION AND FUTURE WORK

This study presents 3D facial action unit based driver fatigue detection. Uncalibrated Photometric Stereo testbed was developed for 3D reconstruction of regions of interest. Dataset was developed on the testbed and the data was further fed to train deep networks for action unit detection. The results show that developed method outperform networks trained on 2D images. The developed method works well for AU1, AU15 and AU41 detection with validation accuracy of 95%, 94.5% and 95% respectively.

As non-invasive lighting have become a possibility, in the future a real time 3D acquisition and fatigue detection device with sufficient accuracy will be developed for in vehicle driver monitoring. To detect fatigue it is necessary to fuse the decision taken by the AU1, AU15 and AU41 classifier, in the future classifier decision fusion algorithm will also be proposed. As reconstruction with Photometric stereo is sensitive to the presence of shadows and error in 3D reconstruction could be caused by the presence of shadow in the region. Currently, images are manually selected to omit the effect of shadows. In future an intelligent technique would be developed for light source configuration for shadow omission.

ACKNOWLEDGMENT

The authors would like to thank the associate editor and anonymous reviewers for their insightful comments, which improved the quality and presentation of the article.

REFERENCES

[1] US Department. of Transportation National Highway Traffic Safety Administration. *Traffic Safety Facts*. Accessed: Jul. 10, 2019. [Online]. Available: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812318?-ga=1.78055380.1104132544.1489526594

[2] *Driver Fatigue and Road Accidents Factsheet*. Accessed: Jul. 10, 2019. [Online]. Available: https://www.rospa.com/rospaweb/docs/advice-services/road-safety/drivers/driver-fatigue-factsheet.pdf

[3] J. F. May and C. L. Baldwin, "Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies," *Transp. Res. F, Traffic Psychol. Behav.*, vol. 12, no. 3, pp. 218–224, May 2009.

[4] W. Harris, "Fatigue, circadian rhythm, and truck accidents," in *Vigilance*. Boston, MA, USA: Springer, 1977, pp. 133–146.

[5] D. Tran, H. Manh Do, W. Sheng, H. Bai, and G. Chowdhary, "Real-time detection of distracted driving based on deep learning," *IET Intell. Transp. Syst.*, vol. 12, no. 10, pp. 1210–1219, Dec. 2018.

[6] G. Sikander and S. Anwar, "Driver fatigue detection systems: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2339–2352, Jun. 2019.

[7] K. Kaida *et al.*, "Validation of the karolinska sleepiness scale against performance and EEG variables," *Clin. Neurophysiol.*, vol. 117, no. 7, pp. 1574–1581, Jul. 2006.

[8] A. Anund, C. Fors, and C. Ahlstrom, "The severity of driver fatigue in terms of line crossing: A pilot study comparing day-and night time driving in simulator," *Eur. Transp. Res. Rev.*, vol. 9, no. 2, p. 31, 2017.

[9] A. Craig, Y. Tran, N. Wijesuriya, and P. Boord, "A controlled investigation into the psychological determinants of fatigue," *Biol. Psychol.*, vol. 72, no. 1, pp. 78–87, Apr. 2006.

[10] Y. Sun and X. Yu, "An innovative nonintrusive driver assistance system for vital signal monitoring," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 6, pp. 1932–1939, Nov. 2014.

[11] Y. Dong, Z. Hu, K. Uchimura, and N. Murayama, "Driver inattention monitoring system for intelligent vehicles: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 596–614, Jun. 2011.

[12] Z. Li, L. Chen, J. Peng, and Y. Wu, "Automatic detection of driver fatigue using driving operation information for transportation safety," *Sensors*, vol. 17, no. 6, p. 1212, May 2017.

[13] M. Omidyeganeh *et al.*, "Yawning detection using embedded smart cameras," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 3, pp. 570–582, Mar. 2016.

[14] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.

[15] B. Mandal, L. Li, G. S. Wang, and J. Lin, "Towards detection of bus driver fatigue based on robust visual analysis of eye state," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 3, pp. 545–557, Mar. 2017.

[16] S. Samiee, S. Azadi, R. Kazemi, A. Nahvi, and A. Eichberger, "Data fusion to develop a driver drowsiness detection system with robustness to signal loss," *Sensors*, vol. 14, no. 9, pp. 17832–17847, Sep. 2014.

[17] C. Jacobé de Naurois, C. Bourdin, A. Stratulat, E. Diaz, and J.-L. Vercher, "Detection and prediction of driver drowsiness using artificial neural network models," *Accident Anal. Prevention*, vol. 126, pp. 95–104, May 2019.

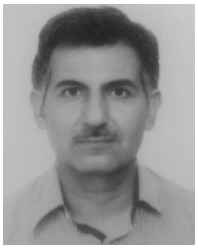
[18] W. Sun, X. Zhang, S. Peeta, X. He, and Y. Li, "A real-time fatigue driving recognition method incorporating contextual features and two fusion levels," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 12, pp. 3408–3420, Dec. 2017.

[19] R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah, "Shape-from-shading: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 8, pp. 690–706, Aug. 1999.

[20] S. K. Nayar and Y. Nakagawa, "Shape from focus: An effective approach for rough surfaces," in *Proc. IEEE Int. Conf. Robot. Autom.*, vol. 2, May 1990, pp. 218–225.

[21] S. Izadi *et al.*, "KinectFusion: Real-time 3D reconstruction and interaction using a moving depth camera," in *Proc. 24th Annu. ACM Symp. User Interface Softw. Technol.*, 2011, pp. 559–568.
[22] R. J. Woodham, "Photometric method for determining surface orientation from multiple images," *Opt. Eng.*, vol. 19, no. 1, Feb. 1980, Art. no. 191139.
[23] M. L. Smith and L. N. Smith, "Dynamic photometric stereo—A new technique for moving surface analysis," *Image Vis. Comput.*, vol. 23, no. 9, pp. 841–852, 2005.
[24] S. Zafeiriou *et al.*, "Face recognition and verification using photometric stereo: The photoface database and a comprehensive evaluation," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 1, pp. 121–135, Jan. 2013.
[25] K. Emrith, L. Broadbent, L. N. Smith, M. L. Smith, and J. Molleda, "Real-time recovery of moving 3D faces for emerging applications," *Comput. Ind.*, vol. 64, no. 9, pp. 1390–1398, Dec. 2013.
[26] S. Anwar, L. Smith, and M. Smith, "Innovative machine vision technique for 2D/3D complex and irregular surfaces modelling," *Int. J. Comput. Sci. Issues*, vol. 9, no. 5, pp. 113–121, 2012.
[27] Y. Quéau, F. Lauze, and J.-D. Durou, "Solving uncalibrated photometric stereo using total variation," *J. Math. Imag. Vis.*, vol. 52, no. 1, pp. 87–107, May 2014.
[28] H. Hayakawa, "Photometric stereo under a light source with arbitrary motion," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 11, no. 11, p. 3079, Nov. 1994.
[29] P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille, "The bas-relief ambiguity," *Int. J. Comput. Vis.*, vol. 35, no. 1, pp. 33–44, 1999.
[30] B. S. Manjunath, C. Shekhar, R. Chellappa, and C. von der Malsburg, "A robust method for detecting image features with application to face recognition and motion correspondence," in *Proc. 11th IAPR Int. Conf. Pattern Recognit. Conf. B, Pattern Recognit. Methodol. Syst.*, Aug. 1992, pp. 208–212.
[31] L. Wu, A. Ganesh, B. Shi, Y. Matsushita, Y. Wang, and Y. Ma, "Robust photometric stereo via low-rank matrix completion and recovery," in *Proc. Asian Conf. Comput. Vis.* Berlin, Germany: Springer, 2010, pp. 703–717.
[32] M. A. Sayette, J. F. Cohn, J. M. Wertz, M. A. Perrott, and D. J. Parrott, "A psychometric evaluation of the facial action coding system for assessing spontaneous expression," *J. Nonverbal Behav.*, vol. 25, no. 3, pp. 167–185, 2001.
[33] I. A. Essa and A. P. Pentland, "Coding, analysis, interpretation, and recognition of facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 757–763, Jul. 1997.
[34] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, and J. Movellan, "Drowsy driver detection through facial movement analysis," in *Proc. Int. Workshop Hum.-Comput. Interact.* Berlin, Germany: Springer, 2007, pp. 6–18.
[35] G. Sikander, S. Anwar, and Y. A. Djawad, "Facial feature detection: A facial symmetry approach," in *Proc. 5th Int. Symp. Comput. Bus. Intell.(ISCBI)*, Aug. 2017, pp. 26–31.
[36] G. Sikander, S. Anwar, and M. T. Khan, "Non intrusive selective facial feature tracking: A fuzzy control approach," in *Proc. 5th Int. Conf. Electr. Electron. Eng. (ICEEE)*, May 2018, pp. 394–398.
[37] E. Vural, M. Çetin, A. Erçil, G. Littlewort, M. Bartlett, and J. Movellan, "Automated drowsiness detection for improved driving safety," in *Proc. Int. Conf. Automot. Technol. (ICAT)*, 2008, pp. 1–15.
[38] E. Vural, M. Bartlett, G. Littlewort, M. Cetin, A. Ercil, and J. Movellan, "Discrimination of moderate and acute drowsiness based on spontaneous facial expressions," in *Proc. 20th Int. Conf. Pattern Recognit.*, Aug. 2010, pp. 3874–3877.
[39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
[40] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[42] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," 2016, *arXiv:1602.07360*. [Online]. Available: http://arxiv.org/abs/1602.07360
[43] S. Jialin Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

Gulbadan Sikander is currently pursuing the Ph.D. degree in machine vision-based intelligent transportation with the University of Engineering and Technology, Peshawar, Pakistan, where she is also a Lecturer. She has served as a Researcher at the Centre of Intelligent Systems and Networks Research (CISNR), Peshawar. Her research interests include intelligent transportation systems, machine vision, artificial intelligence, and machine learning.



Shahzad Anwar received the Ph.D. degree from UWE Bristol (Frenchay Campus), U.K. He is currently serving as an Associate Professor with the Department of Mechatronics Engineering, University of Engineering and Technology, Peshawar, Pakistan. His work focuses on computer vision and artificial intelligence, with particular attention to innovative intelligent system techniques.