IEEE Transactions on Circuits and Systems for Video Technology Volume: 29 , Issue: 4 , April 2019

Recognizing Distractions for Assistive Driving by Tracking Body Parts

Tashrif Billah, S. M. Mahbubur Rahman*, M. Omair Ahmad, *Fellow, IEEE*, and M. N. S. Swamy, *Fellow, IEEE*

***Abstract*—Busy life as well as prevalence of infotainment is increasingly making people more occupied even during tasks that require serious attention. One such task is driving and at the same time getting involved in activities that may distract them cognitively from watching the road and cause fatal accidents. This paper presents a method that is capable of monitoring different types of distractions such as talking and texting on cell phone, casual eating and operating cabin equipment while driving, so that a driver can be assisted to remain cautious on road. The proposed method automatically detects and tracks fiducial body parts of a driver from video captured by a camera mounted on the front windshield inside a vehicle. Relative distances between the tracking trajectories are used as features that represent actions of the driver. Then, the well-known kernel support vector machine is applied for recognizing a particular distraction from the features extracted from body parts. The proposed feature is also compared with previously employed features for tracking-based human action recognition schemes to substantiate its better result in terms of mean accuracy and robustness for distraction recognition. The effectiveness of the proposed method of distraction recognition is also analyzed with respect to tracking errors.**

***Index Terms*—Assistive driving, distraction recognition, tracking of fiducial body parts, trajectory classification.**

Fig. 1. Typical distractions during driving. Examples include talking on cell phone shown in (a), texting in (b), operating equipment in (c) and casual eating in (d).

I. Introduction

Road accident is one of the ten leading causes of death all over the world [1]. Distraction, recklessness, drunkenness, wrong maneuver, and inclement weather are major causes of road accidents. In the recent years, distracted driving has been intensely becoming an apprehension for global road safety. For example, about ten individuals were killed and one thousand and two hundred of people were injured in crashes that involved distracted driving in 2014 in USA [2]. Moreover, there were at least three thousand fatal crashes involving distraction which comprises 10% of all crashes the same year. The use of cell phone alone was reported as a distraction for an approximately 15% of those crashes involving distractions. The gravity of this issue compels the Federal Railroad Administration to ban the use of cell phone and electronic devices for train engineers and conductors on track [3]. These facts point to the need for recognition of distractions to assist driving. In other words, identifying distraction from the activities of drivers has emerged as a significantly important area of research in artificial intelligence.

Distraction during driving can occur in many forms. Typical distractions are depicted in Fig. 1 where the driver is either talking or texting on cell phone, operating cabin equipment or eating food. If these common forms of distractions are identified autonomously, then necessary warnings can be rendered during their occurrences to bring back the focus of the driver on the road. Thus, the possibility of accidents due to distracted driving can be greatly reduced, which is the ultimate goal of the work presented in this paper. Because of the availability of low cost cameras to monitor a subject and the ability to obtain a large amount of spatio-temporal information from such cameras, several studies have investigated video-based algorithms for assistive driving. The videos are captured either from inside or outside the vehicles. In the following subsections, first a literature review is conducted, then the scope of analysis is described, and finally contribution of the paper is delineated.

A. Related Works

There are three approaches that aim at predicting the driver's intent so that assistance can be provided in the event of any

Manuscript received ....The electronic version of the paper is available at IEEE Xplore.

Tashrif Billah is with the Department of Electrical Engineering, Columbia University, New York, NY 10027, USA.

S. M. Mahbubur Rahman is with the Department of Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka 1000, and University of Liberal Arts Bangladesh, Dhaka 1209, Bangladesh.

M. Omair Ahmad is with the Department of Electrical and Computer Engineering, Concordia University, Montréal, QC, H3G 1M8, Canada

M. N. S. Swamy is with the Department of Electrical and Computer Engineering, Concordia University, Montréal, QC, H3G 1M8, Canada

Initial results of this paper were presented in part at the 13th IEEE International Conference on Advanced Video and Signal-based Surveillance (AVSS), Colorado Springs, CO, USA, August 2016.

Tel: +88-02-5516-7100 Ext. 7919 Fax: +88-02-861-1594

E-mail: tashrif.billah@columbia.edu, mahbubur@eee.buet.ac.bd, omair@ece.concordia.ca, swamy@ece.concordia.ca

uncalled maneuver. The first approach relies on monitoring the surroundings of a vehicle. The second approach, on the other hand, monitors the driver inside the cabin. The last approach combines information from both the driver and the vehicle surroundings. For a better organization, the related works of the approaches are described individually.

1) Monitoring Vehicular Surroundings: In this approach, information obtained from multiple cameras and sensors mounted on a vehicle assists in lane changing, collision avoidance, and autonomous cruising. For example, Sultan *et al.* [4] proposed a context-aware architecture that collects information such as the speed and distance of neighboring vehicles from camera as well as the data from traffic management center to infer the maneuver of the driver. Shia *et al.* [5] developed a semi-autonomous controller for correcting the driving inputs by monitoring the vehicle surroundings along with some signals received from the steering wheel, speedometer, brake and accelerator pedals. Vehicle trajectories were also used by traffic monitoring system to predict the possibility of lane changing or turn taking in [6]–[8]. Data of vehicular dynamics have also been exploited for recognizing distractive activities of a driver. Such an attempt can be found in [9], wherein the distractions in a simulated driving environment are identified as a failure to keeping the lane, maintaining speed, and following the vehicle ahead at a safe distance. Notably, the distractions in this work are simulated by performing secondary visual tasks rather than driving in a real-life environment.

2) Monitoring Driver Inside a Vehicle: Types of inattention such as the drowsiness and fatigue of a driver can be identified by monitoring the driver inside the cabin by means of a front camera. Chutorian and Trivedi [10] continuously tracked the position and orientation of the head of the driver in six degrees of freedom (clockwise and counter clockwise parameters of pitch, yaw, and roll) to infer the attention of the person. Liang et al. [11] studied the relation between cognitive distraction and types of eye movements in the context of driving in a simulated environment. The movements of head and eyes as well as the personal characteristics of drivers were used to identify distraction during driving [12]. In [13], it was shown that the visual and cognitive distractions estimated from the facial action units and head pose variations are strongly correlated. Videos monitoring the hands and head of a driver are employed to identify three activity regions, namely, the wheel, gear, and instrument cluster inside a vehicle [14]. In-vehicle tasks such as watching video, reading news, and writing email are recognized by extracting features from the tracking coordinates of head and eyes [15]. Moreover, in a recent unpublished work [16], the gaze of eyes, orientation of head, action units of faces, and position of arms estimated from Kinect camera have been employed for distraction recognition in a simulated environment. In a similar environment, Google-glass has been used to count the frequency of the eye blinks to distinguish between alertness and drowsiness of a driver [17].

3) Monitoring Both Driver and Surroundings: Studies have been conducted by considering information from both the inside and surroundings of a vehicle for a better prediction of the intent of a driver. Monitored eye gaze combined with detected road sign has been reported to be an effective cue for assistive driving. For example, in [18], the safety concern of identification of collocated road signs has been investigated in a simulated driving environment by tracking eye gaze of the driver. In particular, when multiple signs or directions of traffic are placed together on a road, the driving performance has been evaluated in terms of identifying the desired signs from the collocated ones. Rezaei and Klette [19] have correlated the head pose of a driver with road signs to identify the level of fatigue of the motorist. Fletcher *et al.* [20] have prescribed an assistive system that reads the speed-signs on the road by employing a radial symmetry algorithm, analyzes the gaze of the driver by face detection, and finally provides a feedback when a sign is missed by the driver. In a similar work by Jain *et al.* [21] the facial feature points of the driver are tracked and the information acquired from global positioning system are incorporated for predicting the maneuverability of a vehicle. A detailed review on monitoring systems that identify inattention of drivers for intelligent vehicles can be found in [22].

B. Scope of Analysis

Existing literatures mainly focus on either predicting vehicular maneuver or detecting the inattention or fatigue of a driver. Most of the methods use information collected from inside or outside the vehicle to determine whether the driver is in correct lane or making correct turn. A few methods have focused on recognizing the inattention or distraction of drivers in simulated environments of driving. We argue that the study of actions of a driver inside a vehicle on road in real-life can provide a deeper understanding about identifying the common distractions during driving. Besides, observing only the head, face or gaze of a driver is not sufficient to detect what the person is doing apart from driving. Hence, there is a scope for work on identifying and categorizing common distractions by exploiting the simultaneous movements of different body parts of the driver. In this context, this work adopts a tracking-based approach for identifying the activity of the driver with particular attention given when the person is distracted.

Recognition of human activities from spatio-temporal data is a prolific area of research in computer vision. There are different approaches of action recognition from video sequences such as the algorithms that employ the histogram of oriented optical flow [23], temporal order invariance [24], silhouette analysis [25], two dimensional principal component analysis (PCA) of motion energy image [26], Fourier transform of tracking trajectories [27], and deep convolutional neural network [28]. Feature descriptors including the histogram of oriented gradient (HOG), histogram of optical flow (HOF), and motion boundary histogram (MBH) of neighboring pixels along the pipeline of tracking trajectories have also been employed for recognizing human actions [29], [30]. Among the existing methods of action recognition, we prefer the tracking-based approach for identifying distractions of the driver. There are certain issues that motivate us to use the tracking trajectories to determine the activities of a driver. In this particular problem, the body trunk of the driver mostly remains still and the eyes observe the surroundings. Hence, the face shows noticeable movement when monitored by a camera mounted in front of the vehicle. Again, the hand of a driver steers the wheel or shifts the lever. Apart from driving activities, the hand may engage in others tasks such as holding the cell phone or food, and operating cabin equipment. Since hand and face perform most of the tasks during driving, tracking of hand and facial movements may provide useful information about the actions being performed. Moreover, the tracking-based approach of recognizing actions during driving has specific advantage over other methods in view of the requirement of small data volume for the storage of trajectories.

There exists a few instances where the tracking-based approach has been implemented to identify different actions such as the picking-up and putting-down an object [31], entering and leaving a building [32], and walking and running [33]. To the best of our knowledge, tracking trajectories of body parts of a driver have not been attempted yet for predicting complex actions representing distractions such as eating food, texting a friend, talking on cell phone or operating cabin equipment during driving. Besides, there exist successful algorithms that can automatically detect face [34], [35] and hand [36], [37] even in unconstrained settings. Based on these facts, the work presented in this paper analyzes the actions of a driver on an experimental database through tracking of body parts with an aim to first detecting the occurrence of distraction and then classifying the type of distractions.

C. Specific Contributions

In this paper, we propose a video tracking-based algorithm to recognize distractions of drivers for potential application in assistive driving systems. The main contributions of this work are as follows:

- Ensuring automatic detection of fiducial body parts and continuity of tracking trajectories over the frames in the presence of tracking errors
- Modeling activities of drivers by a suitable feature set represented by tracking trajectories obtained from fiducial body parts
- Detecting whether a driver is cautious or distracted from the proposed features as well as recognizing the type of distractions including talking on cell phone, texting, casual eating and operating cabin equipment
- Experimentations on the methods of video-based recognition of distractions while driving in real-life using a publicly released database

This paper is organized in different sections as follows. In Section II, the problem is specified first and then the proposed method is elaborated. Section III describes the experiments and the results to evaluate the performance of the proposed method. Finally, Section IV presents the conclusive remarks on this work.

II. Proposed Method

The proposed method first identifies whether the driver is cautious or distracted during driving. Once the driving action is identified as distracted, the algorithm classifies the action into one of the common types of distractions denoted by l ($l \in 1, 2, \cdots, L$). The types of distractions considered in the setting of this work are talking on cell phone, texting, casual eating, and operating cabin equipment. For identifying the distractions, the proposed method tracks the gesture of the driver from frame-to-frame in order to establish a relation between tracking trajectories and actions being performed. In the tracking-based approach of distraction recognition, the following questions arise:

- What are the active body parts during driving?
- How can they be tracked?
- What are the effective feature vectors?
- How should the feature vectors be classified?

Answers to these questions are sequentially described in the following subsections.

A. Fiducial Body Parts

In this section, we identify the fiducial body parts that are required to be tracked for autonomous detection of driving activities. The proposed system has video footage captured from front view of a driver as the input data. Let the driver be involved in any of the distraction activities l ($l \in 1, 2, \cdots, L$) during driving. If the person talks during driving, his hand holds the cell phone and it goes to the ear. In case the person eats food, his hand moves back and forth between the lips and steering wheel. In a similar fashion, if the person texts during driving, he repetitively looks down to his hand-held cell phone screen and then to the road. Thus, there exists significant movements of fiducial body parts denoted by b ($b \in${hand, lips, forehead}) for the commonly-observed actions that divert the attention of the driver from road. Intuitively, if the these body parts are tracked simultaneously, irrespective of hand-held object such as cell phone or food, their tracking trajectories are expected to provide significant information about the actions being performed.

B. Tracking of Body Parts

The Kanade-Lucas-Tomasi (KLT) point tracker is used for tracking the fiducial body parts b ($b \in${hand, lips, forehead}) of the driver. The details of KLT tracker algorithm can be found in [38] and [39]. In brief, the proposed method takes a set of neighboring pixels around the centroid (x_0, y_0) of a body part b bounded by a window $\mathcal{N}(x_0, y_0)$ in the initial video frame as inputs and tracks them in terms of some features estimated from the set of pixels. Reported good features for tracking include the standard deviations of the spatial intensity profile [40], the zero crossings of the Laplacian of the spatial intensities [41], and the first and second derivatives of the intensity function [42]. Let N be the number of frames of a video captured from a driver i ($i \in 1, 2, \cdots, M$), where M is the total number of drivers considered in the training stage to estimate tracking-based features for action recognition. Let $I_b^i(x_j, y_j)$ and $I_b^i(x_{j+1}, y_{j+1})$ be the pixel intensities of the centroids of a body part b in two consecutive frames j and $j+1$ ($j \in 0, 1, 2, \cdots, N-1$) of a video sequence. Then tracking is carried out by finding an association between the two centroids of a body part b given by

$$I_b^i(x_j, y_j) \Leftrightarrow I_b^i(x_{j+1}, y_{j+1}) \tag{1}$$

$$x_{j+1} = x_j - \chi_j \tag{2}$$

$$y_{j+1} = y_j - \gamma_j \tag{3}$$

where (χ_j, γ_j) is the displacement between the coordinates of two centroids in consecutive frames in terms of their association. The displacement is obtained by minimizing a residual error given by

$$\epsilon = \sum_{\mathcal{N}(x_j, y_j)} \Big[g_b^i(x_j, y_j) - g_b^i(x_{j+1}, y_{j+1}) \Big]^2 w^i(x_j, y_j) \tag{4}$$

where $w^i(x_j, y_j)$ is a weighting function, $g_b^i(x_j, y_j)$ and $g_b^i(x_{j+1}, y_{j+1})$ are the features for KLT tracker in two consecutive frames extracted from local neighboring region $\mathcal{N}(x_j, y_j)$ around the centroid (x_j, y_j). The common choice of $w^i(\cdot)$ is a Gaussian function to emphasize location of centroid in the window being tracked.

Algorithm 1 describes the step-by-step process of automatic detection and tracking of the forehead and lips. The algorithm first detects the face using Viola-Jones algorithm [43]. The four key points, namely, RE (right eye), LE (left eye), RL (right lips), and LL (left lips) are detected on the face region by applying the algorithm described in [34]. The two key points RF (right forehead) and LF (left forehead) are found to be p ($p \in [10, 30]$) pixels above RE and LE. Then, we construct the neighboring region $\mathcal{N}(x_j, y_j)$ by centering the mid-point between RF and LF. A similar neighboring region is constructed for the lips using the key points RL and LL. The KLT tracker keeps track of the centroids of the forehead and lips using the features prescribed in [40] and [41]. As the tracking progresses over frames, the association of tracking coordinates may be poor or even lost due to a change in illumination, out of plane rotation or sudden movement of the body parts. The association of the tracking coordinates between two successive frames can be defined in terms of the 'tracking error' as

$$e = \left(1 - \frac{g_b^i(x_{j+1}, y_{j+1})\ w^i(x_{j+1}, y_{j+1})}{g_b^i(x_j, y_j)\ w^i(x_j, y_j)}\right) \times 100 \tag{5}$$

which represents the percentage of the weighted difference of the features of neighboring regions of the frames. If the tracking error is below a certain threshold η_1 ($\eta_1 > 0$), it is interpreted as a loss of only a few coordinates in $\mathcal{N}(x_j, y_j)$. In this case, the tracker can continue tracking with negligible compromise in the performance of the distraction recognition. On the contrary, if the error is in the interval $[\eta_1, \eta_2])$ ($\eta_2 > \eta_1 > 0$), then the coordinates in the neighboring region are re-sampled by considering that a significant number of coordinates in $\mathcal{N}(x_j, y_j)$ miss the association. Finally, if the error is greater than the threshold η_2, it is interpreted as irrelevant coordinates of the body part to be tracked. Then, initial detection of the body part is reapplied to retrieve the tracking coordinates. It is to be noted that if the threshold values for re-sampling and re-initialization are small, then the performance of distraction recognition is embellished only at the expense of increased computational complexity and vice versa. A similar procedure is followed for detection and tracking of the hand as described in Algorithm 2. In this case, however, the re-sampling and re-initialization thresholds η_3 and η_4 ($\eta_4 > \eta_3 > 0$) are lower than that of the forehead and lips in order to account for a higher error due to frequent movements of the hand. An important aspect of hand detection is the selection of a localized search region instead of the whole frame. This is mainly due to the fact that color and texture of skin play a major role in the hand detection (see, [36] and [37]), and hence, the localized search prohibits detection of the face region erroneously as a hand-like object. In the proposed system, the localized region is chosen as the lower-right quarter of the frame for right-hand driving and the lower-left quarter for left-hand driving in view of fact that hand is mostly found to be placed on a steering wheel in this region. In certain scenarios, when a hand appears to be outside the chosen search region or very close to face region for a while, the history of tracking coordinates (~ 20 frames) is considered for re-sampling in $\mathcal{N}(x_j, y_j)$.

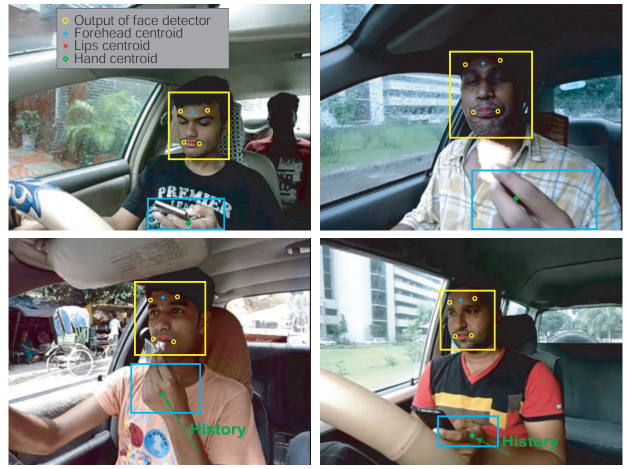


Fig. 2. Examples of initial detection and re-detection of centroids of forehead, lips and hand due to tracking errors. The images in the upper row illustrate the initial detection of the body parts when drivers are distracted due to texting and eating. The forehead and lips are detected following steps 5-9 of Algorithm 1, whereas the hand is detected following the steps 2-6 of Algorithm 2. The images in the lower row show the re-initialization of the body parts once a tracking error occurs. The forehead and lips are re-detected using a similar approach of initialization. In contrast, the centroid of the hand is re-detected from the past history of the tracking trajectory.

Fig. 2 illustrates typical examples of initial detection of the centroid of forehead, lips, and hand, and their re-detection once tracking error happens. The coordinate positions of initialization are presented in the frames in the upper row depicting the distractive actions texting and eating and that of re-initialization are shown in the frames of lower row. The bounding boxes around the face and hand are included for a better visualization. The centroids of the forehead, namely, RF and LF, and that of the lips, namely, RL and LL are obtained using steps 5-6 of Algorithm 1 for both the initialization and re-initialization. On the other hand, the centroid of the hand is obtained initially using steps 2-5 of Algorithm 2. Given the fact that the face and hand exhibit similar skin complexion, which is one of the main cues for hand detection, and very

Algorithm 1 Detection and Tracking of Forehead and Lips

1: Acquire video from interior camera;
// To force the tracker initialize directly
2: $Flag = 0$;
3: **while** Video frame exists **do**
4:   **if** $Flag == 1$ **then**
5:     Detect face using Viola-Jones algorithm [43];
6:     Apply algorithm of [34] on the face region to detect RF(right forehead), LF(left forehead), RL(right lips), and LL(left lips);
7:     Construct $\mathcal{N}(x_j, y_j)$ around centroids of the forehead and lips as illustrated in Fig. 2;
8:   **else**
9:     Re-sample $\mathcal{N}(x_j, y_j)$ in the neighborhood of existing points;
  end if
10:   **for** Each of forehead and lips **do**
11:     Apply KLT tracker on (x_j, y_j) of the body part using features [40], [41];
12:     Find tracking error e using (5);
13:     **if** $\eta_1 < e < \eta_2$ **then**
    // Association of a few coordinates is lost
14:       $Flag = 0$;
15:     **else if** $e > \eta_2$ **then**
    // Tracking is unacceptable
16:       $Flag = 1$;
    end if
  end for
end while

Algorithm 2 Detection and Tracking of Hand

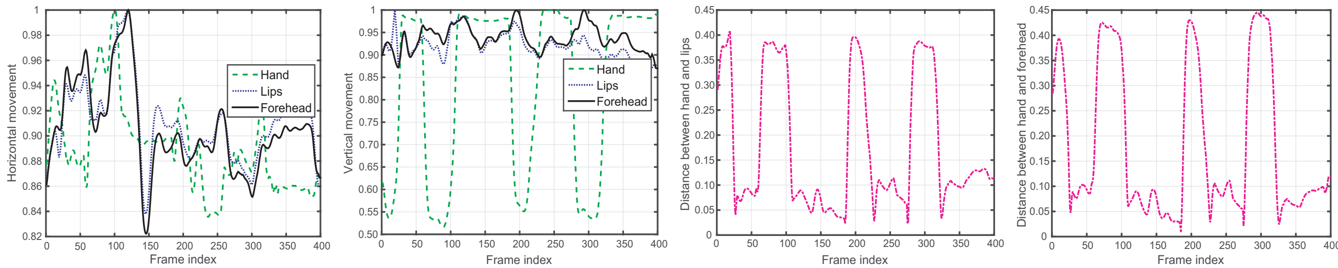
1: Acquire video from interior camera;
2: **Region**= Lower-right (or lower-left) quarter of the frame;
//Hand detection to initialize the tracker
3: Search for hand in **Region**;
4: Detect hand in **Region** applying [36];
5: Find centroid (x_j, y_j) of the segmented hand;
6: Construct $\mathcal{N}(x_j, y_j)$ around centroid of the hand (see Fig. 2);
// To force the tracker initialize directly
7: $Flag = 0$;
8: **while** Video frame exists **do**
9:   **if** $Flag == 1$ **then**
   // Hand detection for re-initialization
   // Hand may move outside ***Region***
   // Hand may even appear outside the frame
   // Re-initialize from the past ***History*** *of* (x_j, y_j)
10:     Re-sample $\mathcal{N}(x_j, y_j)$ around (x_j, y_j) obtained from **History**;
11:   **else**
12:     Re-sample $\mathcal{N}(x_j, y_j)$ in the neighborhood of existing points;
  end if
13:   Apply KLT tracker on (x_j, y_j) of hand using features [36], [37];
   // To facilitate re-initialization
14:   **History** = Last few positions of (x_j, y_j);
15:   Steps 11-16 of Algorithm 1 using η_3 *and* η_4
end while

often the hand and face are in close proximity, the centroid-based features of hand are re-initialized from its past history to reduce the possibility of erroneous detection (see steps 9-15 of Algorithm 2).

In the context of driving, commonly observed actions are constrained by the sitting arrangement inside the vehicle. As argued earlier, the movement of the driver in the seat is insignificant in most cases. It is the collateral tasks during driving that cause distractions. This natural yet constrained setting allows us to exploit relatively simple and yet effective detectors and point tracker for fiducial body parts, i.e., forehead, lips and hand to recognize distractions. Although the human body is 3D deformable in an open environment, such deformations are marginal in the driving seat and thus measurable through detection and tracking of the proposed fiducial body parts. In particular, the minor deformations of the body parts appearing in the frontal pose are easily traceable by Algorithms 1 and 2 (see Fig. 2). Even when a tracking error occurs, the re-initialization takes place almost immediately as the deformations in the body parts are marginal. In the case, when tracking of a body part is missing for a large number of consecutive frames, the system identifies that the driver is out of regular position representing complete distraction.

C. Smoothing of Tracking Trajectories

Let $P_b^{li}(x_j, y_j)$ be a point on the trajectory of a body part b ($b \in$\{hand, lips, forehead\}) on the jth frame ($j \in 0, 1, 2, \cdots, N-1$) of a video captured from the ith driver ($i \in 1, 2, \cdots, M$) performing an action l ($l \in 1, 2, \cdots, L$). The coordinate range of the tracking trajectories obtained by using the method described in Section II-B varies considerably due to the variations of body structures of the drivers (e.g., tall or short and fat or slim) and to that of the placements of video cameras in front of the driver. To account for the effect of the variations of body shapes of drivers as well as that of the mounting of the cameras on the estimated trajectories, the tracking coordinates are scaled first. Let $\tilde{P}_b^{li}(x_j, y_j)$ be the normalized version of tracking trajectories in the range $[0, 1]$ both in the x and y coordinates. In addition, the vehicle jerks due to a road bump, braking or accelerating during driving. These inherent noises result in noticeable fluctuations even in the normalized trajectories. To offset such random fluctuations,



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

Fig. 3. Trajectories representing the action eating : (a) Horizontal movement of hand, lips, and forehead. (b) Vertical movement of hand, lips, and forehead. (c) Euclidean distance between hand and lips. (d) Euclidean distance between hand and forehead.

a trajectory smoothing process is required. There are a number of choices such as the Gaussian filtering [44], cubic spline interpolation [45], and Savitzky-Golay finite impulse response filtering [46] that can be applied for smoothing the trajectories. These methods require parameter adjustment that introduces relatively high complexity in implementation to obtain an acceptable performance. Instead, we propose the use of the moving average filter with adaptive window length, which is computationally of low complexity, yet is much effective for this problem. Finally, the smoothed tracking trajectories are obtained as

$$\hat{P}_b^{li}(x_j, y_j) = \sum_{k=0}^{\tau} \tilde{P}_b^{li}(x_{j-k}, y_{j-k}) \tag{6}$$

where τ $(\tau > 1)$ is the span of the moving average filter. In particular, the span of the filter is chosen on the basis of the variance of the tracking coordinates of the trajectory of previous frames. If the variance is large, the span is chosen to be wider and vice versa.

D. Tracking-Based Features

In order to find suitable feature sets for representing each of the actions, the tracking trajectories $\hat{P}_b^{li}(x_j, y_j)$ $(j \in 0, 1, 2, \cdots, N-1)$, $(i \in 1, 2, \cdots, M)$ of body parts b (b $\in${hand, lips, forehead}) for each of the actions l $(l \in 1, 2, \cdots, L)$ are carefully observed. Figs. 3(a) and (b), respectively, show trajectories of the horizontal and vertical movements of hand, lips, and forehead during the action eating as an example. From this set of trajectories, little or no relationship can be drawn between the movement of body parts and the action being performed. On the other hand, if Euclidean distances between the trajectories of a pair of fiducial body parts are estimated, then certain signature can be found that characterizes the action. Here, quasi periodic signatures are observed as shown in Figs. 3(c) and (d), respectively, when two pairs of body parts, namely, the hand and lips, and the hand and forehead, are considered for the action eating. Similar types of favorable reasoning hold for the other classes of distractions such as talking on cell phone, texting, and operating cabin equipment.

To further substantiate the suitability of inter distances of a pair of fiducial body parts as feature sets, we evaluate its significance over traditional feature sets in terms of the tracking trajectories of each of the body parts. The suitability of a feature set is signified by the discrimination score, which can be defined as the ratio of between-class variance to within-class variance of the features [47]. The larger the discrimination score, the more significant is the feature. Let V_W^b and V_B^b be the within- and between-class variances, respectively, estimated from the trajectory features of body part b. The within-class variance of the features can be estimated as

$$V_W^b = \frac{1}{NLM} \sum_{j=0}^{N-1} \sum_{l=1}^{L} \sum_{i=1}^{M} \left[\hat{P}_b^{li}(x_j, y_j) - \mu_b^l(x_j, y_j)\right]^2 \tag{7}$$

where $\mu_b^l(x_j, y_j)$ is the mean of trajectories of lth action on jth frame given by

$$\mu_b^l(x_j, y_j) = \frac{1}{M} \sum_{i=1}^{M} \hat{P}_b^{li}(x_j, y_j). \tag{8}$$

In a similar fashion, the between-class variance of the features are

$$V_B^b = \frac{1}{NL} \sum_{j=0}^{N-1} \sum_{l=1}^{L} M \left[\mu_b^l(x_j, y_j) - \mu_b(x_j, y_j)\right]^2 \tag{9}$$

where $\mu_b(x_j, y_j)$ is the mean of the entire trajectory set considering all actions l $(l \in 1, 2, \cdots, L)$ given by

$$\mu_b(x_j, y_j) = \frac{1}{L} \sum_{l=1}^{L} \mu_b^l(x_j, y_j) \tag{10}$$

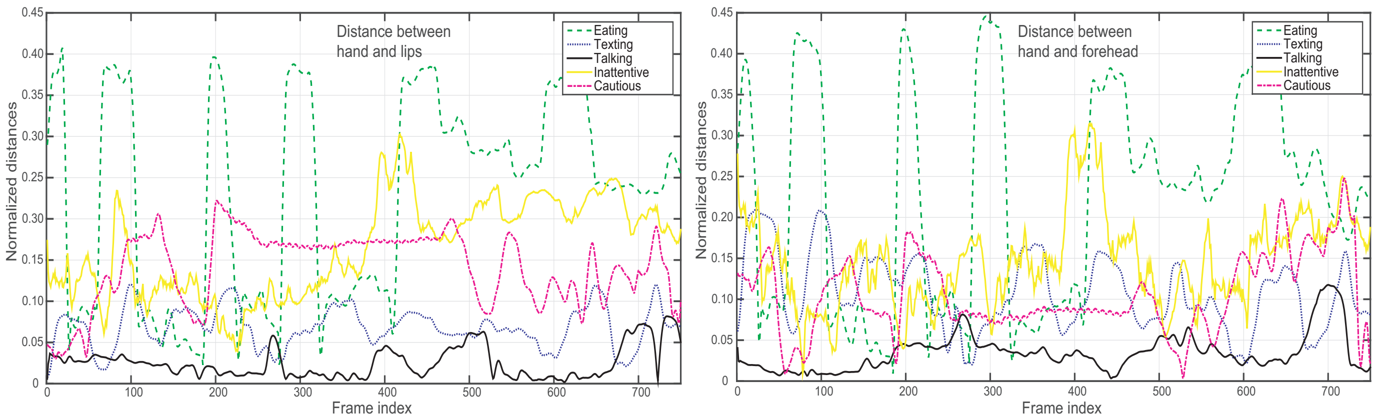
Finally, the discrimination score of the trajectory-based features of a body part b is evaluated as

$$\Theta_b = \frac{V_B^b}{V_W^b} \tag{11}$$

Table I shows the discrimination scores of features obtained from the x- and y-axis trajectories of the three body parts, viz., the hand, lips, and forehead as well as that obtained from the Euclidean distance of the trajectories of all possible pairs of body parts, when the values of L, M, and N are 4, 13, and 750, respectively. It can be observed from the table that the inter distance of trajectories of two pairs of body parts, namely, the hand and lips, and the hand and forehead show the highest scores, which substantiate their suitability as features. Hence, the proposed feature vectors for recognizing activities

TABLE I
DISCRIMINATION SCORE OF TRAJECTORY-BASED FEATURES

	Features								
	Hand-x	Hand-y	Lips-x	Lips-y	Forehead-x	Forehead-y	**Hand-Lips**	**Hand-Forehead**	Lips-Forehead
Θ_b	1.24	0.38	1.45	4.32	2.02	5.18	**7.68**	**6.06**	1.29



(a) (b)

Fig. 4. Proposed trajectory-based features $\mathbf{F}^l_{b_m b_n}$ of a driver when driving is performed with four types of distractions, namely, the eating, texting, cell phone talking, inattentive (operating cabin equipment) or cautious driving. The features are the inter distances of trajectories of pairs of fiducial body parts, namely, b_m and b_n. The two pairs of body parts are (a) hand and lips representing the features $\mathbf{F}^l_{b_1 b_2}$, and (b) hand and forehead representing the features $\mathbf{F}^l_{b_1 b_3}$.

of the driver for the problem of distraction recognition can be written as

$$\mathbf{F}^l_{b_m b_n} = \left[f^l_0, f^l_1, f^l_2, \cdots, f^l_j, \cdots, f^l_{N-1}\right] \quad (12)$$

$$f^l_j = \parallel \hat{P}^l_{bm}(x_j, y_j) - \hat{P}^l_{bn}(x_j, y_j) \parallel, \qquad 0 \leq f^l_j \leq 1 \quad (13)$$

where $b_m \in$ {hand} $(m = 1)$, $b_n \in$ {lips, forehead} $(n = 2, 3)$ and $\| \cdot \|$ represents the $\mathcal{L}_2$ norm. Fig. 4 shows typical examples of the proposed trajectory-based features $\mathbf{F}^l_{b_m b_n}$ extracted from tracking coordinates of forehead, lips and hand of a driver. The features represent different distractive actions such as eating, cell phone talking, texting, and inattentive (operating cabin equipment) as well as cautious driving. It is evident from this figure that the action class eating is exhibited by a quasi-periodic back and forth signature of variation of trajectories. A pseudo-random variation of trajectory is observed for the action inattentive (operating the cabin equipment), while this is low in magnitude for the actions texting and cell phone talking. During the action of talking on cell phone, the hand firmly holds the phone near the ear that is close to the forehead, and movements of lips occur due to conversation. It is due to the fact that the driver firmly holds the cell phone with the ear, the relative distance between the hand and lips or the hand and forehead appears to be small in magnitude. On the other hand, when the driving state of a person is cautious, the hands rest on the steering wheel except for occasional turn taking. The high magnitude of inter-distances between body parts are found for cautious driving, since the position of the hand remains on the steering wheel which is far from the ear. In cautious driving, however, flat-type features especially in the middle region are found mainly due to the resting position of the hands on steering wheel. Thus, the signatures of the proposed trajectory-based features signify their suitability for distraction recognition while driving.

A question may arise about the robustness of the proposed trajectory-based features in case of significant level of tracking errors. Fig. 5 illustrates the robustness of the proposed features obtained from hand and lips during the action of eating while driving. In the example, the movements of the body parts in the vertical direction are considered, since they are more significant than the horizontal movements. In Fig. 5, a number of abrupt transitions of hand in terms of spikes are observed. The normalized distance of lip remains close to unity because the movement of this part of the body is insignificant as compared to that of the hand during eating. The spikes in the trajectory of hand can be attributed to sudden change in illumination and quick movements of this part of body towards the mouth during the action. When such a transition occurs, the error thresholds η_3 and η_4 come into effect by forcing the system to re-initialize the tracking. Since the tracking coordinates are retrieved almost instantly, the tracking features, i.e., the distances between hand and lips are not adversely affected due to the spikes. Similar results can be observed for other body parts performing different actions, but are skipped here to avoid repetition. It is observed that the main source of tracking error that affects the proposed features is the change in illumination inside the cabin, which is overcome by introducing Algorithms 1 and 2 (see Section II-B). On the other hand, minor error of tracking due to a road bump is tackled by smoothing the trajectories (see Section II-C). Thus, the proposed trajectory-based features can be treated as robust in the event of tracking errors.

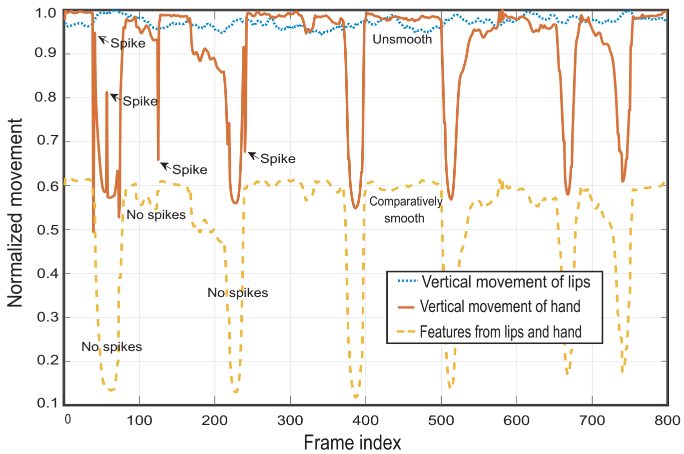


Fig. 5. Illustration of robustness of the proposed trajectory-based features against tracking errors during the action eating. The trajectory representing vertical movements of hand has spikes that mainly arise due to significant change of illumination inside the cabin. In addition, both the trajectories of hand and lips are rough due to bumps on road. The re-initialization of tracker and smoothing of trajectories result in the proposed features to be robust.

E. Classification of Features

The proposed method identifies an unknown action l ($l \in 1, 2, \cdots, L$) from the test feature vector $\mathbf{F}_{b_m b_n}$, $b_m \in \{\text{hand}\}$ ($m = 1$), $b_n \in \{\text{lips, forehead}\}$ ($n = 2, 3$) that is estimated from the video footage of a driver. We have chosen a supervised learning system called the kernel support vector machine (SVM) that determines the action through a model constructed from trained data set representing known actions in terms of $\mathbf{F}^l_{b_m b_n}$. The kernel SVM implicitly maps the feature vectors to another higher dimensional feature space, wherein the projected features of one class are separated from the rest of the classes with maximal margin by an optimal hyperplane [48], [49]. Let $S = \left\{ \left(\mathbf{F}^l_{b_m b_n}, h_{lmn} \right) \mid (l = 1, 2, \cdots, L)(m = 1)(n = 2, 3) \right\}$ be the training set where $\mathbf{F}^l_{b_m b_n} \in [0, 1], h_{lmn} \in \{-1, 1\}$ represent the classes of actions. Then, the test feature vector $\mathbf{F}_{b_m b_n}$ representing an unknown action is classified as

$$\Psi(\mathbf{F}_{b_m b_n}) = \text{sign} \left[\sum_{\substack{n=2 \\ m=1}}^{\substack{n=3 \\ m=1}} \sum_{l=1}^{L} \alpha_{lmn} h_{lmn} \Phi \left(\mathbf{F}^l_{b_m b_n}, \mathbf{F}_{b_m b_n} \right) + c \right] \tag{14}$$

where α_{lmn} are the Lagrange multipliers of a dual optimization problem that describe the separating hyperplane, c is a scalar bias, and $\Phi \left(\mathbf{F}^l_{b_m b_n}, \mathbf{F}_{b_m b_n} \right)$ is a kernel function. SVM finds the separating hyperplane with respect to the support vectors $\mathbf{F}^l_{b_m b_n}$ with $\alpha_{lmn} > 0$ that are randomly taken from the training set. The common choices of kernel functions include the linear, polynomial, multilayer perceptron, and radial basis function (RBF). Since SVM provides a binary decision on a class, the multiclass decisions are obtained from multiple two-class problems. Finally, the class of an action is obtained from a majority voting of the decisions of $\frac{L(L-1)}{2}$ number of binary classifiers. In case of a tie, preference is given to the action that provides maximum safety during driving.

III. Experiments and Results

Several experiments are performed to validate the suitability of the proposed feature sets for the problem of tracking-based distraction recognition of a driver. This section first describes the characteristic of the database on which the experiments are carried out. Then, the experimental setup is explained and the results obtained are reported. Finally, an analysis of system performance against tracking error is presented.

A. Database

The Internet was intensively searched for video clips depicting distracted driving. In general, the videos available in the web are not captured from a single direction to depict a particular type of distraction. Most of the videos are captured from the side view, while only a few are from the behind or front. Notably, the major portion of video clips are cinematized to portrait an accident in a very brief duration. These limitations annulled their use as experimental data for recognizing the distraction activities of a driver. In other words, due to the unavailability of quality data set of distracted driving, we were impelled to shoot video clips to develop a useful database. For this purpose, a good resolution camera – Sony Cyber Shot 14.1 mega pixels was affixed on the front windshield facing the driver inside the vehicles. Videos of a number of drivers were captured on city roads and university campus in Dhaka, Bangladesh. The developed data set is diverse in terms of landscape, illumination, type of vehicle and road condition (smooth or bumpy) as well as of the age or experience of the drivers. A total of 13 drivers voluntarily took part in the development of the database without being informed about the ultimate purpose of capturing the videos. The drivers were asked to drive attentively first. Then, the drivers were asked to do talking or texting on cell phone, eating food casually and operating cabin equipment during driving. Roughly five minutes of driving video was shot by each driver performing each of the actions. The captured videos have a frame rate of 30 frames/second. Each of the frames are in RGB format with bit depth 24 having a resolution of 854×480 pixels. We refer the database developed as the 'EEE BUET Distracted Driving (EBDD)' database that is released in [50]. This database is used in the experiments presented in this paper. At the same time, it is now publicly open for experimentation of algorithms related to activity recognition of the driver.

B. Setup

The experiments are conducted on Matlab environment. The processor used for computation is an Intel Core i5 with clock speed of 2.6 GHz and memory size of 8 GB. In the experiments, the captured videos are processed first for each of the actions. Then, an action is identified as either cautious or distracted by applying the method explained in Section II. Once the action is labelled as distraction, the type of distraction is predicted in the following step. In the experiment, the action recognition performance of the proposed feature vector is contrasted with existing trajectory-based feature vectors that were previously employed for recognition of human actions.

TABLE II
PERCENTAGE MEAN ACCURACIES AND STANDARD DEVIATIONS FOR THE TWO-CLASS PROBLEM OF DETECTION OF DISTRACTION

Features	Classifiers	Actions	
		Cautious	Distracted
Spatial Curvature [31], [51]	KNN	50.50±5.86	56.40±6.50
	SVM	72.50±5.76	69.20±6.40
Angle Variation [52]	KNN	61.76±6.19	67.33±5.90
	SVM	71.67±6.09	67.33±5.90
Fourier Transform [27], [53]	KNN	53.33±5.86	70.25±5.35
	SVM	83.33±5.86	80.25±5.28
Principal Component [54]	KNN	76.17±6.46	75.67±6.92
	SVM	86.17±6.40	85.67±5.92
Dense Trajectories [29], [30]	KNN	81.40±6.29	80.95±6.35
	SVM	85.55±5.37	87.62±5.25
Proposed Feature	KNN	81.50±4.45	81.67±4.35
	SVM	91.32±4.36	91.67±4.22

In order to improve the readability, the experimental setup is detailed in separate sections.

1) Processing of Videos: The videos in the database developed are broadly categorized as cautious or distracted. The videos showing cautious driving by each of the drivers are segmented into four 30-second clips having 900 frames each. In total, there are 52 video segments of cautious driving in the experimental data. Similarly, there are 52 video segments, each of which depicts one of the four commonly-observed distractions considered in this work. The frames of a 30-second clip are further grouped into six 5-second non-overlapping segments each having 150 frames. Each of the segments constitute a tracklet of the whole tracking trajectory estimated as per the method described in Section II-B. Hence, there are 312 video segments of attentive driving and equal number of video segments of distracted driving in total to carry out the experiments.

2) Settings of Tracking Threshold: The proposed features are estimated from the tracking trajectories of the forehead, lips, and hand of a driver from a video segment. In order to circumvent the tracking error, specifically to re-sample the neighboring coordinates in case of weak association of tracking over the frames and to automatically re-initialize the body parts in the case of illumination variation and occlusion, the proposed method employs four thresholds - two for the face region and the rest for hand region (see Algorithms 1 and 2). The choice of the threshold values is a trade off between efficacy and efficiency of the tracker. A large value of the thresholds results in a crude estimation of the trajectories that may yield a low accuracy in the distraction recognition. On the other hand, a small value of the thresholds implies frequent detections of the body parts and increasing rate of re-sampling of the coordinates that impose significant computational overhead on the system. That is why, two threshold values for a region are chosen in a way that result in a compromise between the accuracy and computational complexity giving a satisfactory performance. The threshold values for the face and hand regions are empirically found to be $(\eta_1 = 15\%, \eta_2 = 40\%)$ and $(\eta_3 = 10\%, \eta_4 = 35\%)$, respectively. The threshold values of the hand region is slightly lower than that of the face region, since the former is more prone to tracking error than the latter due to its significant movements.

3) Settings of Classifier: The performance of the proposed tracking-based features are evaluated first as a two-class problem, wherein the driving videos are categorized as cautious or distracted. In the experimental setup, each of the two categories have 312 video segments that are employed for performance evaluation. In the next stage, the videos with distractive driving are grouped into four categories of actions, viz., talking (on cell phone), texting, eating, and inattentive (operating cabin equipment). Thus, the recognition of the type of distraction becomes a four-class problem. As per the grouping, there are 78 video segments for each of the four classes of distractions. For both the two- and four-class problems, the data set is partitioned by randomly choosing video segments as training set and the rest as a separate testing set. There are 60% of the total number of video segments in the training set that are used to train the SVM model. The trained parameters of SVM are used to predict the label of the mutually exclusive 40% videos in the testing set. The results of classification are reported by summarizing the accuracies obtained from 15 sets of random partitions. The choice of best kernel function for SVM is another point of interest in the experiment. The linear, quadratic, polynomial, multilayer perceptron, and RBF functions are applied as suitable kernels, and the RBF shows the best results. The sigma parameter of the RBF is adjusted through auto scaling and 10-fold cross validation. It is worth mentioning that the driving distraction could be classified as a 5-class problems in one-step process instead of two-step process followed in our proposed method. Experimental results reveal that two-step classification process provides at least 8% improvement over the one-step process for recognizing driving distractions.

4) Detection of Distraction: In order to recognize a video as one of the two classes, namely, cautious or distracted, the SVM predicts the labels of each of the frames of a tracklet. If more than one-half of the total number of frames is predicted as distracted, then the whole video segment is recognized as distracted. Similar decision criteria is applied for the cautious video segments. If the total number of frames for each of the two classes are in a tie, then the video is recognized as distracted as a precautionary measure.

5) Determining Type of Distractions: Once the action is identified as distracted, the type of distraction is predicted by applying the proposed classification method. In other words, the action of the driver is further categorized into one of the four commonly-observed distractions considered in this work. Similar to the approach of determining distraction, SVM provides the labels of each of the frames of a tracklet as one of the four types of actions. The type of distraction of the video is predicted through majority voting scheme on the estimated labels of frames. In other words, if the number of a particular label is greater than ζ fraction of the total frame number of a tracklet, it is identified as that particular type of distraction with a confidence of ζ. It is observed through extensive simulation that $\zeta = 0.7$ is a good choice for this classification.

6) Execution Time: In the proposed method, the tracking trajectories of the body parts are estimated using the Algorithms 1 and 2. Typical execution times of these two algorithms for a 5-sec video of 150 frames are 0.3 sec and 0.4 sec, respectively. Overall, the time required for estimating the trajectory-based features for a 5-sec video sequence is 0.5 sec. The classifier is trained off-line, and it takes only 0.1 sec to infer whether the driver is driving attentively or engaging in a distraction action. Once the action is detected as a distraction, it takes additional 0.2 sec to classify the type of distraction in one of the four classes that we have considered.

7) Comparison with Existing Features: The proposed feature for action recognition of driver is compared with five previously reported trajectory-based features that were employed for recognizing human actions. In order to present a rational comparison on the same platform, the experimental setup including the classifier used for the proposed feature is also applied for the comparing features. In a previously published work [55], the distraction of a driver was detected from the tracking trajectories by using the K-nearest neighbor (KNN) classifier. In this paper, we include comprehensive results by employing both the SVM and KNN classifiers for comparing the tracking-based features for detection and classification of distractions. A brief overview of the methods that use the existing trajectory-based features is described as follows:

- Curvature of Trajectories [31], [51]: In general, the curvature is a measure of twisting or curving of motion trajectories. Spatio-temporal curvature of two-dimensional (2D) trajectories was used to identify natural human actions such as the picking up and putting down an object, opening and closing a door, and erasing a white board [31]. Local maxima of curvature function was also employed to study different actions including jogging, waving, and boxing [51].
- Angular Variation of Trajectories [52]: Every motion trajectory changes its angle in course of time. Such angular variations estimated from local neighboring frames form a bag-of-words for describing a particular action. In the experimental settings, five neighboring frames are found to perform the best in most cases. Event detection was demonstrated by classifying the angle-based features of the tracking trajectories. Human actions that were identified by this approach include the transition from running to walking and walking to running. In addition, the type of maneuvers of vehicles such as stopping and turn taking were also identified using this feature.
- Fourier Transform of Trajectories [27], [53]: Repetitive motions in human actions were identified through the Fourier transform-based approach. The selected Fourier coefficients were taken as features for action recognition. Cyclic human actions such as walking and running were reported to be identified well by this method.
- PCA of Trajectories [54]: The principal component analysis of motion trajectories was applied in this method for identifying actions such as jumping, running, and side stepping.
- Dense Trajectories [29], [30]: The bag-of-words defined in terms of trajectories as well as HOG, HOF, and MBH of the local neighboring pixels of trajectories were employed to recognize common forms of human actions.

C. Results

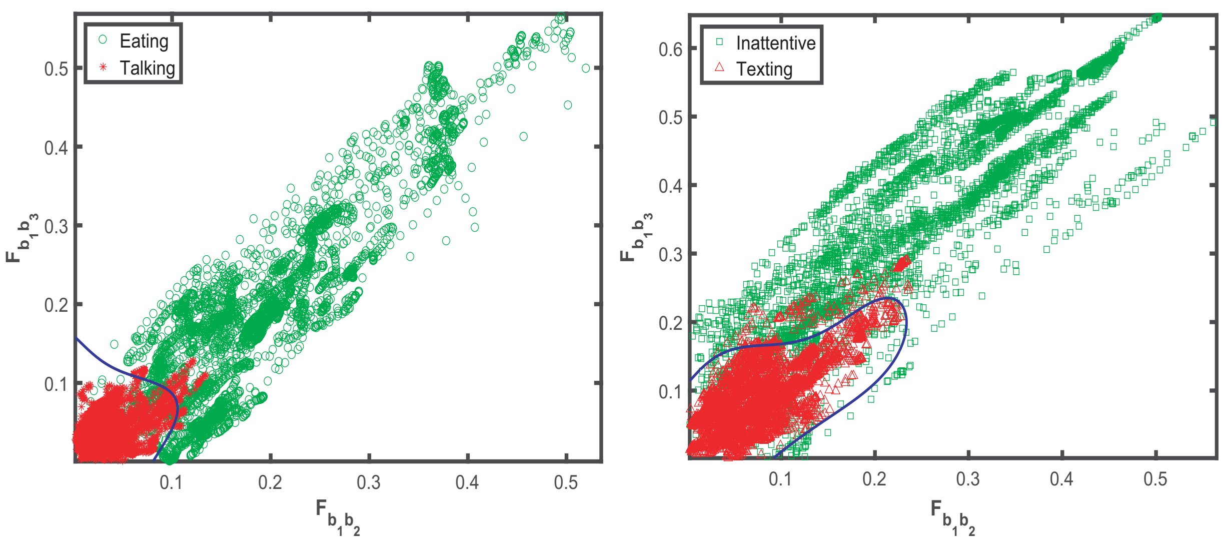
In this work, we first solve for the two-class problem of detecting the existence of distraction during driving, and then solve for four-class problem of determining the type of distraction. Table II shows the percentage of mean accuracies and standard deviations for the two-class problem of detection of distraction for the proposed method along with a comparison with existing trajectory-based action recognition methods. It can be observed from Table II that the average accuracy of distraction detection using the proposed features is more than 91% and the improvement of detection accuracy is not less than 4% in comparison to that of the other methods. Also, the robustness of distraction detection is improved by at least 1% by the proposed method as compared to others.

Table III shows the percentage mean accuracies and standard deviations for the four-class problem of classifying the type of distractions. It can be seen from this table that the overall mean accuracy of the distraction classification for the proposed features is more than 90% and the improvement of accuracy is more than 2.5% as compared to the existing features. Also, the robustness of distraction classification is improved by more than 1.0% compared with others. It can further be observed from Table III that the accuracies of recognition of action ‘eating’ by the proposed features and Fourier transform-based features are close to each other. The closeness of results of these two methods can be explained by the fact that the action ‘eating’ is characterized by back and forth movements of hand between the lips and the steering wheel, and that the Fourier transform identifies well such periodic nature of the action. Another competitive performance to the proposed feature in identifying the action ‘inattentive’ is found to be the features obtained from PCA. The closeness of results of these two methods may have occurred due to the fact that the ‘inattentive’ action is characterized by the random movement of body parts and PCA works well on random patterns.

Table IV shows the confusion matrix for the two-class problem of detecting distraction. It is seen from this table that cautious driving can be falsely identified as distracted driving and vice versa with an error margin of less than 9%. This false recognition may happen in different scenarios; for example, touching hair momentarily or adjusting air-condition system can be recognized as a state of distraction although such a state of driving can be considered as cautious. It can be also observed from Table IV that the errors due to misclassification for both the cautious and distracted driving actions are nearly same. This result can be explained by noting the fact that half the frame numbers of a tracklet is chosen as decision threshold and there are equal number of videos in each of the two classes in the experiments. Table V shows the confusion matrix for the four-class problem of classifying the type of distractions.

TABLE III
PERCENTAGE MEAN ACCURACIES AND STANDARD DEVIATIONS FOR THE FOUR-CLASS PROBLEM OF CLASSIFYING TYPE OF DISTRACTIONS

Features	Classifiers	Actions				
		Talking	Texting	Eating	Inattentive	Overall
Spatial Curvature [31], [51]	KNN	80.00±5.65	76.67±5.80	70.00±6.35	70.00±5.95	74.17±5.94
	SVM	86.67±5.60	83.33±5.33	80.00±4.90	66.67±5.45	79.17±5.32
Angle Variation [52]	KNN	76.67±6.25	73.33±7.50	70.00±6.95	73.33±5.94	73.33±6.67
	SVM	86.67±4.90	83.33±7.20	76.67±6.09	80.00±5.63	81.67±5.96
Fourier Transform [27], [53]	KNN	76.67±5.95	73.33±8.45	76.67±5.80	80.00±6.58	76.67±6.70
	SVM	83.33±5.68	80.00±4.80	83.33±4.32	90.00±6.58	84.17±5.35
Principal Component [54]	KNN	80.00±5.35	73.33±7.50	73.33±7.60	80.00±6.88	76.67±6.83
	SVM	86.67±5.25	80.00±5.96	83.33±6.60	90.00±6.88	85.00±6.17
Dense Trajectories [29], [30]	KNN	84.55±6.30	70.32±8.80	71.42±6.35	78.05±6.05	76.09±6.88
	SVM	92.50±4.50	80.72±5.05	88.32±5.95	90.40±5.50	87.99±5.25
Proposed Feature	KNN	83.33±5.60	73.33±5.20	73.33±5.45	83.33±4.95	78.33±5.30
	SVM	93.33±4.80	86.67±4.20	88.67±3.88	93.33±4.05	90.50±4.23



(a) (b)

Fig. 6. Typical scatter plots depicting class separability of the two features, namely, $\mathbf{F}_{b_1b_2}$ and $\mathbf{F}_{b_1b_3}$, for recognizing type of distractions. A significant number features remain in the respective regions marked by the decision boundary that results in correct recognition of distraction. The distractions classified are (a) eating and talking, and (b) texting and inattentive (operating cabin equipment).

It can be observed that the action cell phone 'talking' does not create any confusion with the actions 'eating' or 'texting'. In other words, the cell phone 'talking', which is the most prevalent action that causes distraction during driving in real-life, can be easily distinguished from other actions. This result may be due to the fact that the body parts of the driver remains almost static during cell phone 'talking', while the actions 'eating' and 'texting' involve considerable movements of hand and head. As per the experimental results, the actions 'eating' and 'texting' are found to be the most challenging to be distinguished. It is observed that about 6.67% 'eating' videos are falsely classified as 'inattentive' and 6.67% of 'texting' videos are falsely classified as 'eating'. These misclassifications of the actions 'eating' and 'texting' can be attributed to the fact that the hand holds cell phone or food, which introduces similar features in the trajectories for the two actions.

Fig. 6 depicts two typical scatter plots in a 2D plane showing two scenarios, where the proposed features, namely, $\mathbf{F}_{b_1b_2}$ and $\mathbf{F}_{b_1b_3}$ are characterized by noticeable separation marked by decision boundary resulting in correct recognition for determining the type of distraction. In particular, Fig. 6(a) shows a scatter plot that exhibits distinct separation for a significant number of features for recognizing two actions, viz., 'eating' and cell phone 'talking'. Because of such distinction, none of the videos representing the action 'eating' are falsely classified as the action cell phone 'talking' (see results in Table V). In a similar fashion, Fig. 6(b) shows a scatter plot for classifying two pairs of actions, viz., 'texting' and 'inattentive' that have a noticeable number of features remaining in the well-separated regions. In a few scenario, the misclassification of distractions cannot be avoided, but the classification error of any type of distractions is less than 7% as seen in Table V, which

TABLE IV
CONFUSION MATRIX FOR THE TWO-CLASS PROBLEM OF DETECTING DISTRACTION

Actions	Cautious	Distracted
Cautious	**91.32**	8.68
Distracted	8.33	**91.67**

TABLE V
CONFUSION MATRIX FOR THE FOUR-CLASS PROBLEM OF CLASSIFYING TYPE OF DISTRACTIONS

Distractions	Talking	Texting	Eating	Inattentive
Talking	**93.33**	0.00	0.00	6.67
Texting	0.00	**86.67**	6.67	6.67
Eating	2.33	3.33	**88.67**	5.67
Inattentive	0.00	3.33	3.33	**93.33**

clearly signifies the effectiveness of the proposed trajectory-based feature.

D. Recognition Performance with Tracking Errors

In order to get an insight on the viability of the proposed tracking-based method, the variation of recognition performance with respect to tracking errors is evaluated in the experiments. In particular, we present the results obtained from the video segments representing the distraction 'eating', which is one of the most confusing among all the distractions considered in the experiments (see Table V). In this experimental setting, the trained classifier for the two-class problem of distraction detection is kept functional to predict a label, which is either cautious or distracted, for each of the frames of a video. The frame-by-frame detection accuracies of the distraction 'eating' are obtained by averaging the results over 13 drivers. The frame-by-frame tracking errors of the forehead and lips are estimated from the current and previous frames by using (5) and considering that $\eta_1 = 15\%$ and $\eta_2 = 40\%$ (see Algorithm 1). Fig. 7 shows the variation in the percentage accuracies of distraction detection provided by the proposed method with respect to the tracking errors. The figure provides an insight as to how the performance of the proposed system in terms of detection accuracy is maintained at a high level by considering that the tracking thresholds described in Section II-B come into effect as and when necessary. It is seen from Fig. 7 that without any tracking error, the detection accuracy can be as high as 94%, which decreases slowly with the increasing level of errors. When the tracking error between consecutive frames is greater than 15%, the first threshold η_1 comes into effect. The immediate result of the application of the threshold, i.e., adoption of re-sampling of the feature points in the neighborhood, is an increasing level of detection accuracy. As the tracking error increases considerably ($\geq 30\%$), the detection performance deteriorates significantly due to the loss of association of the forehead and lips. At this stage, the second threshold η_2, which is set to 40%, comes into effect so that the tracking coordinates of the associated body part is re-initialized. Application of the

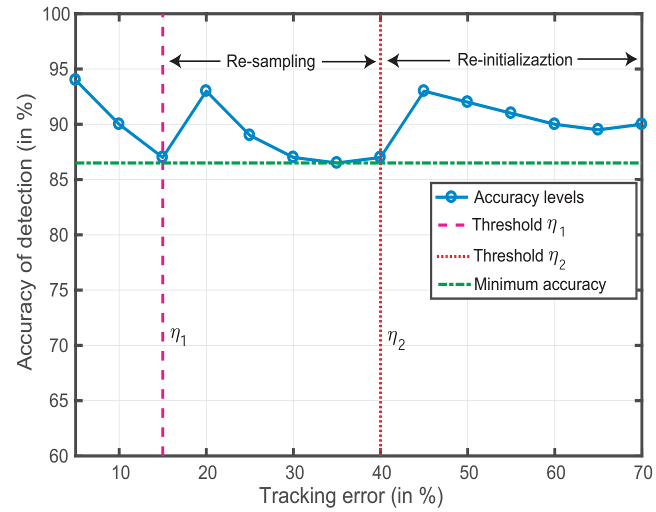


Fig. 7. Illustration of variation of accuracies for detecting distraction in the presence of tracking errors. The type of distraction is considered to be 'eating'. The tracking errors are estimated for forehead and lips.

second threshold in the tracking system results in a gradual increase of the detection accuracy, similar to that observed for the application of the first threshold. It is due to the successive applications of the two thresholds, viz., η_1 and η_2, the overall detection accuracy remains well above 89% even with certain frames having tracking errors greater than 60%. It is also found that the proposed approach of tracking can ensure a minimum level of detection accuracy for the 'eating' distraction to be 87% (see Fig. 7). Thus, the proposed tracking-based features can be viable to detect and recognize commonly occurring driving distractions.

IV. CONCLUSION

In this paper, it has been attempted to identify whether the driver is cautious or distracted during driving by capturing video from the front view of the driver. The proposed method first detects the existence of distraction from a video and then classifies the type of distraction. Fiducial body parts for activity recognition of a driver have been found to be the hand, lips, and forehead. Algorithms have been developed that not only detect the initial coordinates of the body parts automatically, but also keep track of the movements of these body parts over the video frames with acceptable tracking errors. The inter-distances among the tracking trajectories of the hand and lips, and hand and forehead have been demonstrated as suitable feature sets representing different activities of the driver. It has been shown that the proposed tracking-based features are robust, especially because the re-sampling and re-initialization of the coordinates of body parts are employed by considering the tracking errors over the frames. The proposed feature sets have been classified using the kernel SVM technique for detecting as well as recognizing the type of distraction. In order to carry out experiments for evaluating the performance of the proposed method, the EBDD dataset has been developed by considering diversity in driving environments as well as expertise of drivers. With a view to facilitating further work on distraction recognition of the driver, the dataset has been released publicly. Experimental results show that the proposed method can provide at least overall accuracies of 91% and 90%, respectively, for detecting and recognizing the type of distraction. In both scenarios of detection and recognition, the robustness of accuracy is improved by at least 1% by the proposed method as compared to the existing methods. An in-depth analysis on the accuracy of the distraction detection with respect to tracking error has been made to substantiate the suitability of the proposed tracking-based features. It is expected that the proposed distraction recognition technique be implemented in an embedded system installed in a vehicle for assistive driving so that the driver is kept more vigilant on the road. In a future research, information of vehicular dynamics such as change of speed of the vehicle and that of lane position can be incorporated into the proposed features in order to improve the performance of the distraction recognition. Moreover, to alleviate the effect of inherent noise caused by occlusion and illumination variation, an effective noise-robust classifier such as random forest algorithm can be investigated in the setting of this work.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments, which have been useful in improving the quality of the paper.

References

[1] World Health Organization, "Fact sheets," Geneva, Switzerland, Tech. Rep. FS 310, 2014.

[2] National Center for Statistics and Analysis, "Distracted driving 2013," National Highway Traffic Safety Administration, Washington, DC, Tech. Rep. DOT HS 812132, 2015.

[3] US Department of Transportation Federal Railroad Administration, "Restrictions on railroad operating employees use of cellular telephones and other electronic devices," *Federal Register*, vol. 75, no. 186, pp. 59 579–59 604, 2010.

[4] S. Sultan, A. Bayatti, and H. Zedan, "Context aware driver behavior detection system in intelligent transportation systems," *IEEE Trans. Vehicular Technology*, vol. 62, no. 9, pp. 4264–4275, 2013.

[5] V. Shia, Y. Gao, R. Vasudevan, K. Campbell, T. Lin, F. Borrelli, and R. Bajcsy, "Semiautonomous vehicular control using driver modeling," *IEEE Trans. Intelligent Transportation Systems*, vol. 15, no. 6, pp. 2696–2709, 2014.

[6] H. Berndt, J. Emmert, and K. Dietmayer, "Continuous driver intention recognition with hidden Markov models," in *Proc. 11th IEEE Int. Conf. Intelligent Transportation Systems*, Beijing, China, 2008, pp. 1189–1194.

[7] P. Kumar, M. Perrollaz, S. Lefevre, and C. Laugier, "Learning-based approach for online lane change intention prediction," in *Proceedings of IEEE Intelligent Vehicles Symposium*, Queensland, Australia, 2013, pp. 797–802.

[8] M. Liebner, M. Baumann, F. Klanner, and C. Stiller, "Driver intent inference at urban intersections using the intelligent driver model," in *Proc. IEEE Intelligent Vehicles Symposium*, Alcal de Henares, Spain, 2012, pp. 1162–1167.

[9] F. Tango and M. Botta, "Real-time detection system of driver distraction using machine learning," *IEEE Trans. Intelligent Transportation Systems*, vol. 14, no. 2, pp. 894–905, 2013.

[10] E. Chutorian and M. Trivedi, "Head pose estimation and augmented reality tracking: An integrated system and evaluation for monitoring driver awareness," *IEEE Trans. Intelligent Transportation Systems*, vol. 11, no. 2, pp. 300–311, 2010.

[11] Y. Liang, M. L. Reyes, and J. D. Lee, "Real-time detection of driver cognitive distraction using support vector machines," *IEEE Trans. Intelligent Transportation Systems*, vol. 8, no. 2, pp. 340–350, 2007.

[12] J. Zeng, Y. Sun, and L. Jiang, "Driver distraction detection and identity recognition in real-time," in *Proc. 2nd WRI Global Cong. Intelligent Systems*, vol. 3, Wuhan, China, 2010, pp. 43–46.

[13] N. Li and C. Busso, "Analysis of facial features of drivers under cognitive and visual distractions," in *Proc. IEEE Int. Conf. Multimedia and Expo*, San Jose, CA, 2013, pp. 1–6.

[14] E. O. Bar, S. Martin, A. Tawari, and M. Trivedi, "Head, eye, and hand patterns for driver activity recognition," in *Proc. 22nd Int. Conf. Pattern Recognition*, Stockholm, Sweden, 2014, pp. 660–665.

[15] C. Braunagel, W. Stolzmann, E. Kasneci, and W. Rosenstiel, "Driver-activity recognition in the context of conditionally autonomous driving," in *Proc. 18th IEEE Int. Conf. Intelligent Transportation Systems*, Canary Islands, Spain, 2015, pp. 1652–1657.

[16] C. Craye and F. Karray, "Driver distraction detection and recognition using RGB-D sensor," Cornell University, Tech. Rep. arXiv preprint:1502.00250, 2015.

[17] J. He, W. Choi, Y. Yang, J. Lu, X. Wu, and K. Peng, "Detection of driver drowsiness using wearable devices: A feasibility study of the proximity sensor," *Applied Ergonomics*, vol. 65, pp. 473–480, 2017.

[18] A. J. Filtness, G. Larue, A. Schramm, J. Fuller, A. Rakotonirainy, C. Han, and P. Cairney, "Safety implications of co-locating road signs: A driving simulator investigation," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 47, pp. 187–198, 2017.

[19] M. Rezaei and R. Klette, "Look at the driver, look at the road: No distraction! no accident!" in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Columbus, OH, 2014, pp. 129–136.

[20] L. Fletcher, G. Loy, N. Barnes, and A. Zelinsky, "Correlating driver gaze with the road scene for driver assistance systems," *Robotics and Autonomous Systems*, vol. 52, no. 1, pp. 71–84, 2005.

[21] A. Jain, H. Koppula, B. Raghavan, S. Soh, and A. Saxena, "Car that knows before you do: Anticipating maneuvers via learning temporal driving models," in *Proc. IEEE Int. Conf. Computer Vision*, Santiago, Chile, 2015, pp. 3182–3190.

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

[23] F. Fawzy, M. Abdelwahab, and W. Mikhael, "2DHOOF-2DPCA contour based optical flow algorithm for human activity recognition," in *Proc. 56th IEEE Int. Midwest Symp. Circuits and Systems*, Columbus, OH, 2013, pp. 1310–1313.

[24] A. Haq, I. Gondal, and M. Murshed, "On temporal order invariance for view-invariant action recognition," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 23, no. 2, pp. 203–211, 2013.

[25] D. Wu and L. Shao, "Silhouette analysis-based action recognition via exploiting human poses," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 23, no. 2, pp. 236–243, 2013.

[26] M. Naiel, M. Abdelwahab, and M. Saban, "Multi-view human action recognition system employing 2DPCA," in *Proc. IEEE Work. Applications of Computer Vision*, Kona, Hawaii, 2011, pp. 270–275.

[27] M. Zarzuela, F. Pernas, A. Pablos, D. Ortega, and M. Rodrguez, "Action recognition system based on human body tracking with depth images," *Advances in Computer Science: an International Journal*, vol. 3, no. 7, pp. 115–123, 2014.

[28] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Proc. Adv. Neural Information Processing Systems*, Montreal, Canada, 2014, pp. 568–576.

[29] H. Wang, A. Klaser, C. Schmid, and C.-L. Liu, "Action recognition by dense trajectories," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Colorado Springs, CO, 2011, pp. 3169–3176.

[30] H. Wang and C. Schmid, "Action recognition by dense trajectories," in *Proc. IEEE Int. Conf. Computer Vision*, Sydney, Australia, 2013, pp. 3551–3558.

[31] C. Rao, A. Yilmaz, and M. Shah, "View-invariant representation and recognition of actions," *International Journal of Computer Vision*, vol. 50, no. 2, pp. 203–226, 2002.

[32] J. Nascimento, M. Figueiredo, and J. Marques, "Trajectory classification using switched dynamical hidden Markov models," *IEEE Trans. Image Processing*, vol. 19, no. 5, pp. 1338–1348, 2010.

[33] S. Calderara, A. Prati, and R. Cucchiara, "Markerless body part tracking for action recognition," *International Journal on Multimedia Intelligence and Security*, vol. 1, no. 1, pp. 76–89, 2010.

[34] Y. Sun, X. Wang, and X. Tang, "Deep convolutional network cascade for facial point detection," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Portland, OR, 2013, pp. 3476–3483.

[35] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Providence, RI, 2012, pp. 2879–2886.

[36] A. S. Ghotkar and G. K. Kharate, "Hand segmentation techniques to hand gesture recognition for natural human computer interaction,"

International Journal of Human Computer Interaction, vol. 3, no. 1, pp. 1–15, 2012.

[37] N. Dhruva, S. R. Rupanagudi, and H. N. Kashyap, "Novel algorithm for image processing based hand gesture recognition and its application in security," in *Proc. Int. Conf. Advances in Computing, Communication, and Control*, Mumbai, India, 2013, pp. 537–547.

[38] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. 7th Int. Joint Conf. Artificial Intelligence*, Vancouver, Canada, 1981, pp. 674–679.

[39] C. Tomasi and T. Kanade, "Detection and tracking of point features," Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, Tech. Rep. CMU-CS-91-132, 1991.

[40] C. Thorpe, "FIDO: Vision and navigation for a robot rover," Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, PA, 1984.

[41] D. Marr, T. Poggio, and S. Ullman, "Bandpass channels, zero-crossings, and early visual information processing," *Journal of the Optical Society of America*, vol. 69, no. 6, pp. 914–916, 1979.

[42] L. Kitchen and A. Rosenfeld, "Gray level corner detection," *Pattern Recognition Letters*, vol. 1, no. 2, pp. 95–102, 1982.

[43] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, Kauai, Hawaii, 2001, pp. 1–8.

[44] M. P. Deisenroth, R. D. Turner, M. F. Huber, U. D. Hanebeck, and C. E. Rasmussen, "Robust filtering and smoothing with Gaussian processes," *IEEE Trans. Automatic Control*, vol. 57, no. 7, pp. 1865–1871, 2012.

[45] J. M. Chen, F. Deng, and M. Chen, "Locally adjusted cubic-spline capping for reconstructing seasonal trajectories of a satellite-derived surface parameter," *IEEE Trans. Geoscience and Remote Sensing*, vol. 44, no. 8, pp. 2230–2238, 2006.

[46] J. Chen, P. Jönsson, M. Tamura, Z. Gu, B. Matsushita, and L. Eklundh, "A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter," *Remote sensing of Environment*, vol. 91, no. 3, pp. 332–344, 2004.

[47] S. M. M. Rahman, T. Howlader, and D. Hatzinakos, "On the selection of 2D Krawtchouk moments for face recognition," *Pattern Recognition*, vol. 54, pp. 83–93, 2016.

[48] J. Friedman, R. Tibshirani, and T. Hastie, *The Elements of Statistical Learning*, 2nd ed. Springer Series in Statistics, 2008.

[49] S. M. Imran, S. M. M. Rahman, and D. Hatzinakos, "Differential components of discriminative 2D Gaussian-Hermite moments for recognition of facial expressions," *Pattern Recognition*, vol. 56, pp. 100–115, 2016.

[50] EEE BUET Distracted Driving (EBDD) Video Database. Released on April 04, 2016. [Online]. Available: https://mahbubur.buet.ac.bd/resources/ebdd_database.htm

[51] T. Nguyen and A. Manzanera, "Action recognition using bag of features extracted from a beam of trajectories," in *Proc. 20th IEEE Int. Conf. Image Processing*, Melbourne, Australia, 2013, pp. 4354–4357.

[52] H. Cheng and J. Hwang, "Integrated video object tracking with applications in trajectory-based event detection," *Journal of Visual Communication and Image Representation*, vol. 22, no. 7, pp. 673–685, 2011.

[53] P. Tsai, M. Shah, K. Keiter, and T. Kasparis, "Cyclic motion detection for motion based recognition," *Pattern Recognition*, vol. 27, no. 12, pp. 1591–1603, 1994.

[54] D. Chivers and A. Goshtasby, "Human action recognition in videos via principal component analysis of motion curves," in *Proc. Int. Conf. Image Processing, Computer Vision, and Pattern Recognition*, Las Vegas, NV, 2012, pp. 1–7.

[55] T. Billah and S. M. M. Rahman, "Tracking-based detection of driving distraction from vehicular interior video," in *Proc. 13th IEEE Int. Conf. Advanced Video and Signal-based Surveillance*, Colorado Springs, CO, 2016, pp. 23–28.

Tashrif Billah received his B.Sc. degree in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology (BUET), Dhaka, in 2016. Later on, he has received his M.S. degree in Electrical Engineering from Columbia University, New York, in 2017. He is enrolled in a Ph.D. track program at Columbia University. Since his admission at Columbia University, he was awarded full financial aid for his outstanding undergraduate performance. During his M.S. study, he completed some projects in the areas of machine learning and deep learning. His recent work is on object detection in an image using reinforcement learning techniques. During his undergraduate studies, he obtained several academic and extra curricular awards. His notable academic accolades include the Dean's List Award and the University Merit Scholarship in every semesters. Moreover, he held the title of The Best Athlete at BUET because of his outstanding performance in sports. His research interests are in the areas of computer vision, image processing, machine learning, and wireless communication network.

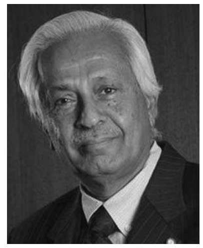
S. M. Mahbubur Rahman received the B.Sc. and M.Sc. degrees in electrical and electronic engineering (EEE) from the Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh, and the Ph.D. degree from the Department of Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada. Dr. Rahman started his career as a Lecturer in the Department of EEE, BUET in 1999. Subsequently, he served as an Assistant Professor from 2002 until 2011, as an Associate Professor from 2011 until 2014, and since then as a Professor in the Department of EEE, BUET. He was tenured in the University of Toronto, ON, Canada as an NSERC (Natural Sciences and Engineering Research Council) Postdoctoral Fellow in 2012. In 2017, he held a position of Visiting Research Professor in the University of Liberal Arts Bangladesh, Dhaka, Bangladesh. His research interests are in the areas of deep learning, biometric security system, intelligent transportation system, cognitive science, stereo vision, virtual reality, biomedical visualization, human computer interaction, video surveillance, signal processing, and communication systems.

Dr. Rahman has a strong record of research in his area that includes contributions to more than 45 publications in top-rank refereed journals, book chapters, and proceedings of international conferences. As per the Google Scholar and Research Gate, his current citation count is 670+, RG score is 20+, h-index is 13, and i10-index is 16. Among his publications, 16 papers appear in the reputed SCI-indexed journals in his area such as IEEE Transactions on Image Processing, IEEE Transactions on Circuits and Systems for Video Technology, IEEE Transactions on Intelligent Transportation Systems, IEEE Transactions on Circuits and Systems-I, and the journals of Elsevier (Pattern Recognition, Expert Systems with Applications, Signal Processing, Journal of Visual communication and Image Representation), IET, and Springer Nature. He has been associated with more than 25 international conferences in different capacities such as coordinator, convener, member of technical and organizing committees, session chair, reviewer, and presenter. He received several prestigious international awards including the Late Sydney R. Parker Best Paper Award in Signal Processing and M. N. S. Swamy Best Paper Award from Journal of Circuits, Systems and Signal Processing (CSSP), Springer Nature in 2015, Post-Doctoral Fellowship Award from NSERC of Canada in 2009, Distinguished Doctoral Dissertation Award from Concordia University in 2009. Throughout his academic career, he has received several prestigious awards at international conferences such as the Best Paper Awards in ICCIT 2016, Dhaka, ISI-RSC 2014, Malaysia and CSNDSP 2002, UK. Currently, he is an Associate Editor of CSSP, Springer Nature.

M. Omair Ahmad (S'69-M'78-SM'83-F'01) received the B.Eng. degree in electrical engineering from Sir George Williams University, Montreal, QC, Canada, and the Ph.D. degree in electrical engineering from Concordia University, Montreal, QC, Canada.

From 1978 to 1979, he was a Faculty Member with New York University College, Buffalo, NY, USA. In September 1979, he joined the Faculty of Concordia University as an Assistant Professor of Computer Science. He then joined the Department of Electrical and Computer Engineering, Concordia University, where he was the Chair with the Department from June 2002 to May 2005 and is currently a Professor. He holds the Concordia University Research Chair (Tier I) in Multimedia Signal Processing. He was a Founding Researcher at Micronet, Ottawa, ON, Canada, from its inception in 1990 as a Canadian Network of Centers of Excellence until its expiration in 2004. He has published extensively in the area of signal processing and holds four patents. His current research interests include multidimensional filter design, speech, image and video processing, non-linear signal processing, communication DSP, artificial neural networks, and VLSI circuits for signal processing.

Dr. Ahmad was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMSI: FUNDAMENTAL THEORY AND APPLICATIONS from June 1999 to December 2001. He was the Local Arrangements Chairman of the 1984 IEEE International Symposium on Circuits and Systems. In 1988, he was a Member of the Admission and Advancement Committee of the IEEE. He has served as the Program Co-Chair for the 1995 IEEE International Conference on Neural Networks and Signal Processing, the 2003 IEEE International Conference on Neural Networks and Signal Processing, and the 2004 IEEE International Midwest Symposium on Circuits and Systems. He was a General Co-Chair for the 2008 IEEE International Conference on Neural Networks and Signal Processing. He is the Chair of the Montreal Chapter IEEE Circuits and Systems Society. He was previously an Examiner of the Order of Engineers of Quebec. He was a recipient of numerous honors and awards, including the Wighton Fellowship from the Sandford Fleming Foundation, an induction to Provosts Circle of Distinction for Career Achievements, and the Award of Excellence in Doctoral Supervision from the Faculty of Engineering and Computer Science of Concordia University.



M. N. S. Swamy (S'59-M'62-SM'74-F'80) received the B.Sc. (Hons.) degree in mathematics from Mysore University, Mysore, India, in 1954, the Diploma degree in electrical communication engineering from the Indian Institute of Science, Bangalore, India, in 1957, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Saskatchewan, Saskatoon, SK, Canada, in 1960 and 1963, respectively.

He was conferred the title of Honorary Professor at National Chiao Tong University, Hsinchu, Taiwan, in 2009. He is currently a Research Professor in the Department of Electrical and Computer Engineering, Concordia University, Montreal, QC, Canada, where he served as the founding Chair of the Department of Electrical Engineering from 1970 to 1977, and Dean of Engineering and Computer Science from 1977 to 1993. Since July 2001, he has held the Concordia Chair (Tier I) in Signal Processing. He has also taught in the Electrical Engineering Department, Technical University of Nova Scotia, Halifax, NS, Canada, and the University of Calgary, Calgary, AB, Canada, as well as in the Department of Mathematics, University of Saskatchewan. He was a founding member of Micronet, Ottawa, ON, Canada, from its inception in 1990 as a Canadian Network of Centers of Excellence until its expiration in 2004, and was also its Coordinator for Concordia University. He has coauthored eight books and five book chapters. He holds five patents. His research interests include number theory, circuits, systems and signal processing.

Dr. Swamy is a Fellow of the Institute of Electrical Engineers (U.K.), the Engineering Institute of Canada, the Institution of Engineers (India), and the Institution of Electronic and Telecommunication Engineers (India). In 2008, Concordia University instituted the M. N. S. Swamy Research Chair in Electrical Engineering as a recognition of his research contributions. In 2009 he was inducted to the Provosts Circle of Distinction for career achievements. He has served the IEEE in various capacities such as President-Elect in 2003, President in 2004, Past-President in 2005, Vice President (Publications) from 20012002, Vice-President in 1976, Editor-in-Chief of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMSI: FUNDAMENTAL THEORY AND APPLICATIONS from June 1999 to December 2001, Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS from June 1985 to May 1987, Program Chair for the 1973 IEEE CAS Symposium, General Chair for the 1984 IEEE CAS Symposium, Vice-Chair for the 1999 IEEE Circuits and Systems (CAS) Symposium, and a member of the Board of Governors of the CAS Society. He was the recipient of many IEEE CAS Society awards, including the Education Award in 2000, the Golden Jubilee Medal in 2000, and the 1986 Guillemin-Cauer Best Paper Award. He has been the Editor-in-Chief of Circuits, Systems, and Signal Processing (CSSP) since 1999. Recently, CSSP instituted a Best Paper Award in his name.