Driver Identification Leveraging Single-turn Behaviors via Mobile Devices

Yan Wang*, Tianming Zhao*, Fatemeh Tahmasbi†, Jerry Cheng‡, Yingying Chen§ and Jiadi Yu¶

*Temple University, Philadelphia, PA, USA 19122

†Binghamton University, Binghamton, NY, USA 13902

‡New York Institute of Technology, New York, NY, USA 10023

§Rutgers University, New Brunswick, NJ, USA 08901

¶Shanghai Jiao Tong University, Shanghai, P.R. China 200240

Email:{y.wang,tum94362}@temple.edu, ftahmas1@binghamton.edu, jcheng18@nyit.edu

yingche@scarletmail.rutgers.edu, jiadiyu@sjtu.edu.cn

***Abstract*—Drivers' identities are essential information that can facilitate a broad range of applications. For example, by understanding who is driving the vehicle when an accident happens, insurance companies could determine the liability and payment in a car accident claim case with high confidence. Another example, pick-up service companies could track the identities of their drivers to ensure that authorized drivers are driving esteemed clients to their destinations. While there are existing studies that can utilize video cameras and dedicated sensors to identify drivers, they either have privacy issues or require additional hardware, which is not practical enough for daily uses. In this paper, we devise a low-cost driver identification system, which can determine drivers' identities by using sensors readily available in wearable devices. Our system captures the unique driving behaviors during pervasive but momentary driving events (i.e., turning at intersections) with motion sensors, which are widely integrated into commodity wearable devices (e.g., smartphones and activity trackers). Toward this end, we extensively analyze people's driving behaviors and identify the critical turning events that capture people's unique behavioral patterns for driver identification. We design a fine-grained turning segmentation method that divides sensor data into critical turning stages (i.e., before, during, and after-turn stages), which provide multiple dimensions of turning behavioral metrics facilitating driver identification. The system extracts unique turning behavior features from time and frequency domains to enable driver identification based on drivers' turning behaviors at different types of turns. Extensive experiments are conducted with 12 drivers and various types of turns in real-road conditions. The results demonstrate that our system can identify drivers with high accuracy and low false-positive rate based on one single turning event.**

***Index Terms*—Driver Identification, Smartphone**

I. Introduction

Drivers' identities are critical information that is highly desirable by various vehicle businesses, including insurance, rental, and on-demand transportation service. For instance, vehicle insurance and rental companies could leverage the information to identify the thief for a stolen car [1] or determine the liability of the drivers involved in a car accident [2], respectively. On-demand transportation services, such as Uber and Lyft, could utilize such technology to verify their drivers' identity, tracking their service and providing security precaution for female passengers riding alone at night. Moreover, by identifying a driver, vehicle manufactures can build an in-vehicle operating system that can automatically switch the vehicle's settings, such as air-condition, rear-view mirror, navigation system and engine tuning, to the driver's preferences. In addition, individuals can also benefit from real-time driver identification. For example, parents would like to know whether or not the driver that drives their kids home is authorized. Figure 1 illustrates the applications that may benefit from having a driver's identification. Existing works on user identification either use PIN, or rely on user gestures [3]–[5]. However, their approaches require active user inputs or dedicated devices (e.g., a smartwatch), which is not safe and scalable for general driving scenarios. Therefore, a low-cost, passive user identification mechanism that can obtain the driver's identification during driving without requiring the driver's input is highly desirable.

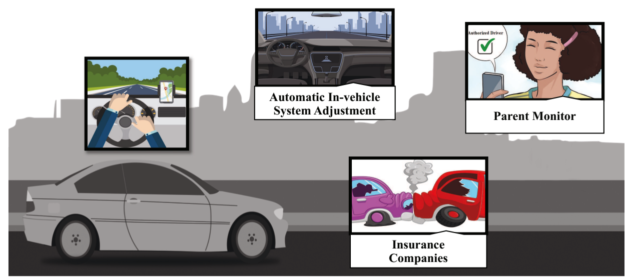


Fig. 1. Illustration of applications that may benefit from driver identification.

Recently, there are several driver identification studies try to address the issue by using drivers' driving behaviors, such as accelerating or braking patterns in different scenarios (e.g., car following, highways, and turns). For example, embedded sensors and On-Board Diagnostic (OBD-II) of vehicles are used to analyze drivers' behavior and differentiate them [1], [2]. David *et al.* [1] analyze drivers' actions when they turn at corners and use the data from the sensors embedded in vehicles to capture the activities for differentiating drivers. In a similar work, Miro *et al.* [2] study long haul driving data from various in-vehicle sensors of brakes, gas pedals, and steering wheels, and use the combination of the sensor data for driver identification. Though such studies can identify drivers with moderate accuracy, they require access to the car's sensing platform or additional sensors deployed at different positions of the vehicle, which is not convenient for normal users and various vehicles without dedicated sensors. In other cases, for example, Zhang *et al.* [6] can identify drivers based on long-time driving data, which makes the identification process less reliable in short distance driving cases. Whereas, they use a large feature set, which might be difficult to deploy in mobile devices with limited resources.

In this work, we propose to identify drivers passively based on their distinct turning maneuvers. We chose turns for driver identification mainly because 1) turning procedures contain a series of controlling events that include not only the driver's accelerating and braking behaviors but also the steering behaviors; 2) drivers' unique driving behaviors in turning procedures are less likely to be impacted by uncontrollable road conditions than those in any other times (e.g., the driver may have variant his/her braking behaviors due to unpredictable traffic or weather conditions). Toward this end, we develop a low-cost driver identification system, which exploits readily available motions sensors (i.e., accelerometers (Acc) and gyroscopes (Gyro)) in commodity mobile devices to capture drivers' unique turning-behavior characteristics for driver identification. Compared to existing works, our system is a practical solution as it only requires a single mobile device in the vehicle. Moreover, as far as we know, it is the first driver identification system that can accurately identify drivers based on single-turn data.

To use the low-cost motion sensors of a single mobile device to differentiate drivers based on the momentary data from one single turn, our system has several challenges. First, since there are different types of turns and various road conditions in practical driving scenarios. It is really hard to model a driver's distinct behavioral characteristics by using the low-cost motion sensors of a single mobile device. Secondly, the motion sensors in the mobile device capture not only the driving behaviors but also a significant amount of noise from the driving environment, including sliding, vibrations, user operations, etc. Therefore, our system needs to mitigate various sensor noises in real-driving scenarios and provide robust driver identification for practical use. Thirdly, many applications would demand timely driver identity information (e.g., parent monitoring and theft detection). Our system aims to provide the driver identification results within the time of a momentary turning event, which is also challenging for the low-cost single device system.

To better understand drivers' turning behaviors, we extensively study the motion sensor data corresponding to accelerating, braking, and turning activities of 12 drivers in various real turning scenarios. Based on our findings, we adopt a novel analysis approach by defining three critical stages for each turn, namely *before*-, *during*-, and *after*-turn stages, which capture independent driving behaviors that can facilitate driver identification. Our system is designed to detect different types of turns (e.g., left/right turns, sharp/90-degree turns, and turns with/without stop events) by using the low-cost motion sensors. For each type of turn, the system determines the three critical stages by using a behavior-based segmentation algorithm and extracts fine-grained turning features to capture drivers' distinct turning behaviors in each stage. Furthermore, our system employs a low-cost and effective classifier using the Gradient Boosting Tree (GBT) algorithm that can be easily deployed in commodity mobile devices to identify different drivers based on the turning features extracted.

The main contributions of our work are summarized as follows:

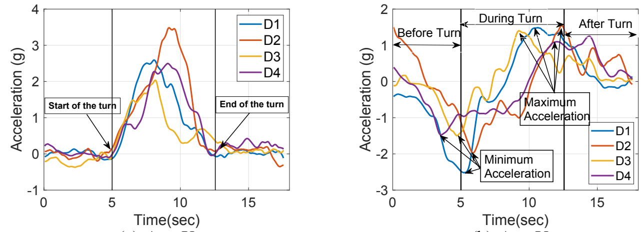
- We extensively study drivers' behaviors during turns and develop a driver identification system that can recognize a driver's identity based on the driver's driving behaviors captured by a single commodity mobile device in a single turn.
- Our system enables accurate passive driver identification by utilizing unique turning features captured by low-cost motion sensors and GBT-based classifier.
- We develop a novel behavior-based segmentation algorithm to separate motion sensor data into critical turning segments, which facilitate fine-grained analysis of drivers' turning behaviors and robust driver identification using the limited amount of data within a single turn.
- We evaluate our system in real driving experiments with 12 drivers and different models of cars. The results demonstrate that our system can achieve 98% accuracy and low false positive rate for driver identification.

II. Related Works

There are existing works trying to identify drivers using sensors in vehicles [2], [7]–[10]. Wakita *et al.* [7] use embedded sensors in a car to capture driving behaviors (e.g., accelerating and braking) and other environmental factors (e.g., turning signals, speed, and distance) for driver identification in car-following scenarios. Choi *et al.* [8] develops a system that can detect distraction and identify drivers using the collected data from the CAN-Bus in vehicles. Similarly, Van Ly *et al.* [9] analyze driving events such as acceleration, braking, and turns using vehicles' inertial sensors from the CAN-bus to identify drivers. All these works require access to multiple dedicated sensors in vehicles, which are not practical. Recently, Wallace *et al.* [10] and Miro*et al.* [2] show that drivers have consistent driving habit , and it is possible to differentiate drivers based on their long-term driving behaviors captured by the sensors in vehicles. These works require long-term monitoring on drivers' behaviors, which is not convenient and useful for real-time applications.

With the emerging use of smartphones, researchers try to differentiate drivers by using smartphones. Zhang *et al.* [6] collect data from sensors in both vehicles and smartphones for driver identification. While their approach has more promising results, the size of the feature set used to identify drivers is too large to run on smartphones. Ezzini *et al.* [11] build a system that uses the inertial sensors in smartphones and ECG sensors on drivers' fingers to differentiate drivers, which still requires long-term sensor data to improve the identification accuracy. Yan *et al.* [12] exploit the sensors in drivers' smartwatches to model their driving behaviors for driver identification. However, smartwatch-based approaches are prone to body movements. Thus, any body movement can significantly effect the results which limits the capability of the model.

Hallac *et al.*'s research [1] is the closest to this work. They study drivers' behaviors in single turns using the data from the sensors in vehicles and develop a system that can identify drivers based on drivers' behaviors in the most frequent-appear



(a) Acc-X (b) Acc-Y

Fig. 2. Comparison among four drivers' accelerations on the X- and Y-axis in a right turn without stop event.

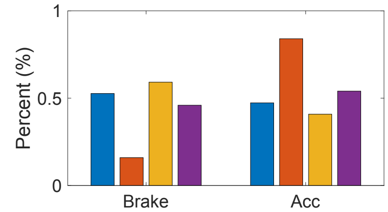


Fig. 3. Illustration of the different percentages of accelerating and braking events among four different drivers in their right turns.

turns. While this work shows it is possible to use vehicle's sensors to capture drivers' unique behaviors and identify drivers in a single turn, it does not provide a good accuracy for more than two drivers. In this work, we develop a smartphone-based driver identification system that can differentiate drivers based on their unique behaviors in single turns. Compared to the existing works, our system utilizes the features that can capture finer-grained characteristics of turning behaviors considering different driving conditions (e.g., turning types and w/o stop events), which makes our system more accurate and robust in practical driving scenarios. Moreover, our feature set is small enough to be analyzed using smart devices with limited resources while achieving high accuracy.

III. Feasibility Study

In this work, we find that the drivers' behaviors through the turning process is complicated and unique, which could facilitate robust and accurate driver identification. The insight is that drivers have different preferences in the braking and accelerating behaviors in the terms of time and the pressure applied on the brake or gas pedals [10], [13]–[15]. In addition, different drivers may have different preferred radius when making similar types of turns, which adds another layer of unique characteristics to turning behaviors on top of the accelerating and braking behaviors. We envision that all of these characteristics can be captured by sensors in smartphones (i.e., Acc and Gyro) and enable driver identification based on turning behaviors.

To study the feasibility of using turning behaviors to differentiate drivers, we ask four different drivers to drive a same car with a smartphone fixed on the dashboard. In these experiments, we log accelerations on the X and Y axes (i.e., Acc-X: centripetal acceleration and Acc-Y: tangential acceleration) of the smartphone, which are determined by the speed and the angle of the turn and accelerating/braking incidents, respectively. We collect data from 100 right turns at several locations without stop events. From Figure 2, we can observe that all drivers have distinct turning behaviors. The *start* and *end* points of a turn (i.e., the time when Acc-X is turning from zero to positive and back to zero) occur at different times in different drivers' data. In addition, Figure 2(a) shows that on average, all four drivers have maximum Acc-X at different times with different amplitudes, indicating that the drivers have different preferences of speed and turning radius. Furthermore, we find that the minimum Acc-Y indicates the moment that a driver starts to release the brake pedal, and the maximum Acc-Y indicates the moment that the driver starts to release the gas pedal. Thus, from Figure 2(b) we can infer that on average: 1) Driver D2 has a higher acceleration in the Before-Turn segment. Therefore he/she likes to brake with higher intensity compared to others. 2) Drivers D3 and D4 prefer releasing the brake pedal in the Before-Turn segment, while drivers D1 and D2 prefer releasing the brake pedal in the During-Turn segment. 3) Drivers D1, D2, and D3 like to release the gas pedal in the During-Turn segment, only Driver D4 keeps increasing the acceleration and stops increasing the acceleration in the After-Turn segment.

To better illustrate the overall differences between drivers in their turning behaviors, we define two behavioral categories, *braking event* and *acceleration event*, which can be extracted from Acc-Y. Figure 3 shows the average percentage of events in each behavioral category over all the turns for four different drivers. For example, D2 has more acceleration events (e.g., pressing the gas pedal) while D3 has more brake events, respectively. All these observations suggest that it is feasible to use sensors in smartphones to capture the unique turning behaviors and identify drivers.

IV. System Design and Challenges

A. Overview

The purpose of this study is to identify different drivers based on their driving behaviors in making a turn. We choose to use turns because they contain a variety of behavioral patterns in a short time, which could facilitate real-time driver identification. Although making a turn seems to be simple, its whole process can be divided into critical stages, such as before, during, and after-turn. Each of the critical stages contains independent and distinct driving behaviors, including wheel rotation, accelerating, and braking. We find that even in the turns that are affected by different road conditions, such as traffic lights and stop signs, the way that each driver handles such situations is still unique and distinguishable.

The basic idea of our work is to analyze drivers' behaviors in fine-grained stages of a turn. We find that each driver has his/her preferences or habits in making a turn from approaching to completing and finally leaving a turn. For example, drivers usually decelerate or brake before making turns. However, the intensity and timing of these actions vary from driver to driver, which has been illustrated in Section III. Besides, drivers also rotate the steering wheel differently during a turn. For instance, a driver may prefer to start rotating the steering wheel at an early stage of a turn and rotate it slowly. Such different turning behaviors on turns also persistent under different traffic or driving conditions, such as turns with/out stop events. The flow of our system is shown in Figure 4.

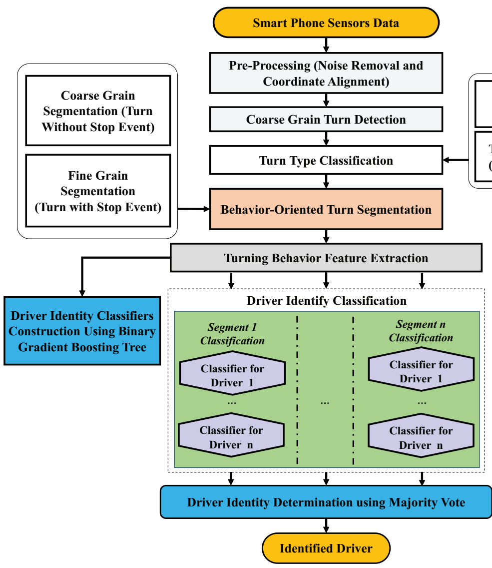


Fig. 4. Overview of the smartphone-based driver identification system.

Our system first takes the data as input from the sensors in smartphones (i.e., Acc and Gyro). Then it conducts the *Data Pre-Processing (Noise Removal and Data Rotation)* to remove noises and rotate the sensor data to the vehicle's coordinate system. Next, the system adopts the *Coarse-grained Turn Detection* to identify the sensor data contains turning behaviors based on Gyro-Z readings. After obtaining the turning data, the *Turn Type Classification* separates the turning data into four categories based on the turning direction (i.e., left or right) and traffic condition (i.e., with or without stop events (defined in the later section)). Next, the *Behavior-Oriented Turn Segmentation* applies fine-grained segmentation on the turning data using different strategies based on the turn type. Each fine-grained segment contains unique driving behaviors corresponding to a particular stage of a turn, including *before*, *during*, and *after* the turn. Note that for the turns with stop events, the during turn segment will be further divided into finer-grained segments to facilitate the driver identification. Then, the system performs the *Turning Behavior Feature Extraction* to derive the features that can capture the unique driving behaviors within each fine-grained segment of the turning data. Last, the features are processed by the *Driver Identity Classification* to identify registered drivers. In this step, a separate binary Gradient Boosting Tree (*GBT*) classifier is built for each registered driver. The final decision of the driver's identity will be determined by using a majority vote mechanism based the confidence scores from the binary classifiers in each segment.

B. Challenges

Realizing a driver identification system using the single-turn data from a smartphone is very challenging. We list several major challenges as shown below:

Practical Modeling of Unique Turning Behaviors. We use turning behaviors for driver identification in this paper. Although turning behaviors have drivers' unique characteristics, like any other driving behaviors, they could also be interfered with by various events on the road as listed in Table. I. Many of the events are completely dynamic and stochastic. For example, the waiting time in front of a stop sign could be a function of the traffic but also depends on drivers' preferences. Therefore, modeling unique turning behaviors considering dynamic road conditions in practice is challenging, if not impossible. We need to come up with a robust solution that can model drivers' turning behaviors effectively under practical driving scenarios.

TABLE I
DIFFERENT DRIVING CONDITIONS THAT DRIVERS MAY ENCOUNTER WHILE MAKING A TURN.

Pedestrian	*With Pedestrian*	*Without Pedestrian*	
Weather Condition	*sunny*	*Rainy*	*Snowy*
Traffic Condition	*Traffic(Light/Heavy)*	*No Traffic*	
Traffic Signs	*Traffic Light*	*Stop Sign*	*Yield*
Turn Sharpness	*Sharp*	*90 degree*	*U-Turn*

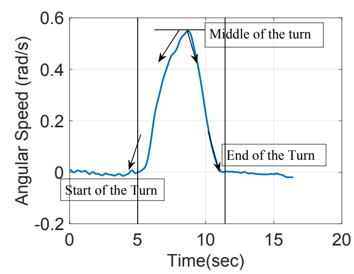


Fig. 5. Illustration of the Gyro-Z readings during a left turn and the start and end points.

Robust Driver Identification Using Mobile Devices. Another challenging part of the project is using a single smart device such as a smartphone to identify drivers. Although it is easy to obtain sensor data from smartphones, such sensor data is usually sensitive to noises caused by road conditions. Additionally, in order to get a good understanding of drivers' behaviors, we need fine-grained information from the sensor data, which requires high sampling rates in the sensors. However, having a high sample rate can be power and memory consuming for smartphones. Therefore, the design of our system should be able to mitigate various sensor noises and provide robust driver identification results when the sensors are running at a low sampling rate (e.g., ten samples/s).

Accurate Classification Using Single-turn Data. The last but not the least challenge in this research is to be able to identify drivers only using the data of one single turn. It means that we aim to identify drivers only with limited data. Mostly any driving session starts with coming out of a parking lot or leaving residential areas in which a driver experiences several turns of different types. We aim to identify drivers using very limited data which does not need a long time of driving data. Due to the short time duration of each turn, our system needs to model a driver's behavior and determine the driver's identity based on very few samples. We note that using turns for driver identification is practical. Usually, a driving session starts with coming out of a parking lot or leaving residential areas. Therefore, the driver experiences several turns of different types, which provide the opportunity for our driver identification solution.

V. Fine-grained Turn Segmentation

A. Coarse-grained Turn Detection

After obtaining the sensor data from the smartphone, our system first performs the noise filtering and coordinate alignment, which is detailed in Section VII. Then we conduct the Coarse-grained Turn Detection to identify turning events and extract the data segments containing the turning behaviors. The Coarse-grained Turn Detection has three steps. First, we determine a turn using a threshold on Gyro-Z amplitude [16]. Through empirical study with 12 drivers, we set the threshold to 0.13rad/s, which gives over 99% accuracy of identify turning events. Using this threshold we can filter the events such as lane changing.

Next, we determine the start and end points of a turn As illustrated in Figure 5, which are the points to reach zero before and after the peak in Gyro-Z readings. Since 90-degree turns are the most common type of turns in practice, especially in residential areas and city environments, our system focuses on using these turns and filter the turns with other angles. Therefore, in the third step, we find the 90-degree turns by examining the turning angle that is calculated by accumulating the Gyro-Z readings between the start and end points. If the turning angle is between 70 degrees and 100 degrees, we consider the turn is 90 degrees and use it for driver identification.

B. Turn Type Classification.

After determining the coarse-grained segment we classify the turns into "left/right" and "with/out stop events" categories. This is a necessary step because we find that drivers can have different behaviors in left and right turns. For example, drivers need to cross a lane to complete a left lane, they have more space to accelerate compared to right turns. Moreover, when drivers encounter events like stop signs, their turning behaviors are unlikely to be the same as when they turn without stop. In particular, we use Gyro-Z to determine left/right turns, which has been used in existing work [16]. To determine if a turn encounters any stop events or not, we need to consider various conditions that can cause the driver to stop in turns as shown in Table I. To simplify the problem, we define the driving conditions that require a driver to stop the vehicle as *Stop Events* and divide turns into two categories, *"Turn with Stop Events"* and *"Turns Without Stop Events"* based on whether the turns involve stop events or not. For example, turns with driving conditions such as *"Pedestrian, Stop sign, and Red traffic light"* are categorized as the *"Turns with Stop Events"*. Otherwise, they are categorized as the *"Turns Without Stop Events"*. Note that we focus on the turns with regular stop events, such as waiting for pedestrians, stop signs, and red traffic lights. We do not consider turns with random stop events caused by road traffic or accidents, which usually occur with small probability in practice.

Specifically, we check whether the turn involves stop events or not based on a threshold on the calculated speed before the start point of a turn. Our system tracks the speed of the vehicle by accumulating the Acc-Y when the vehicle starts moving [17]. Intuitively, all the scenarios in the "*Turns with Stop Event*" category make a driver to drive slowly or stops before the turn. Therefore, when the speed of the vehicle at the start point of a turn is lower than a threshold, we consider the turn involves the stop events and belongs to the category of Turn with Stop Events. Otherwise, the turn is classified as the Turn without Stop Events. Through our experimental results, we find that the speed 3km/h is a good threshold that can help separate most of the turns with stop events from the turns without stop events.

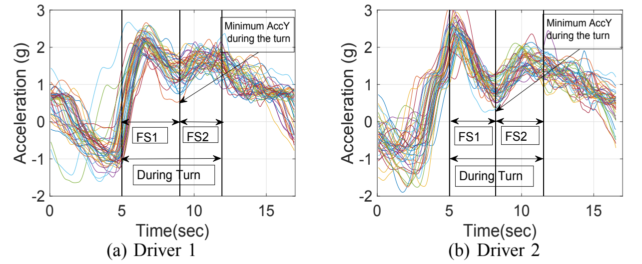


Fig. 6. Illustration of fine-grained segmentation for turns with stop events on Acc-Y of different drivers.

C. Behavior-oriented Turn Segmentation

General Segmentation. There are several driving behaviors involved when people making turns including braking, acceleration and rotating the steering wheel. One major behavior is steering wheel maneuver. We define the *During-Turn* segment to facilitate fine-grained analysis on this steering wheel maneuver behaviors in addition to other behaviors such as braking and acceleration. In order to have a more depth analysis on turning behaviors, we also define *Before-Turn* and *After-Turn* segments which can capture the unique braking and acceleration patterns of drivers when approaching and leaving a turn, respectively. More specifically, we empirically determine the *Before-Turn* segment is the 5s before the start point of a turn, and the *After-Turn* segment is the 5s after the end point of a turn, and the segment between the start and end point of a turn is defined as the *During-Turn* segment. This segmentation captures how the driver approaches the turn, how he makes the turn and leaves it.

Finer-Grained Segmentation for Turns with Stop Events. The general segmentation is applied to all turns, but for turns with stop events we need more fine-grained segmentation to capture drivers' unique turning behaviors. Compared to the turns without stop events, drivers on turns with stop events have less freedom because they need to brake to make a stop before turns and accelerate during and after turns. Therefore, we need to analyze the turns with stop events in a finer-grained manner to achieve high driver identification accuracy.

Particularly, we divide the during-turn segment into two finer-grained segments (denoted as FS1 and FS2) to capture the distinct driving behaviors when drivers make turns after stop events. FS1 is defined as the period from the start point of the turn to the minimum Acc-Y in the during-turn segment. FS2 is defined as the period from the minimum Acc-Y to the end point of the turn. Figure 6 illustrates the fine-grained segmentation on the turns with stop events from two different drivers. As we can see, the two drivers get to their minimum Acc-Y at different times with different intensity, and our finer-grained segments can dynamically adapt to each

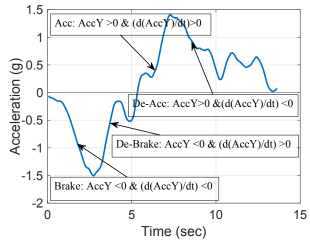


Fig. 7. Illustration of the turning behaviors defined on Acc-Y.

drivers behaviors, capturing the drivers unique behaviors at its maximum resolution. For instance, we can capture how the driver increases its speed to the maximum value. We can also capture how the driver controls the speed for the rest of the turn.

VI. Driver Identification Using Turning Behaviors

A. Behavior Definition on Turns

We divide two main behaviors (i.e., braking and acceleration) into more detailed ones that can give information of how the driver changes the speed as follows: Increasing/Decreasing Acceleration, Increasing/Decreasing Braking, and timing and intensity for each behavior. The definition of these behaviors are listed below.

- Increasing Acceleration (Acc): This behavior is defined as when the driver presses the gas pedal with ascending pressure.
- Decreasing Acceleration (De-Acc): This behavior is defined as when the driver presses the gas pedal with descending pressure.
- Decreasing Braking(De-Brake): This is defined as when the driver is pressing the brake pedal with descending pressure.
- Increasing Braking(Brake): This is defined as when the driver presses the brake pedal with ascending pressure.
- The intensity of the defined behaviors is defined as how hard the driver presses the gas or brake pedal.
- The consistency of the changes of the pressure of the gas and brake pedal.

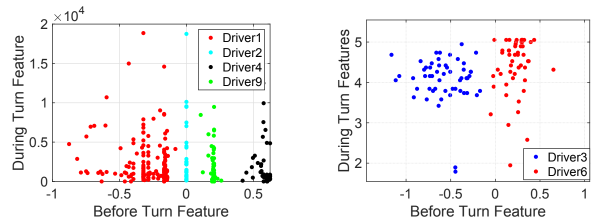
These behaviors can be derived from the *Acc-Y* and its first derivative.When a driver is pressing the gas pedal the acceleration reading is positive, and when the driver is pressing the brake pedal the acceleration reading places in the negative side. Moreover, based on the pressure on the gas/brake pedal the acceleration can be increasing or decreasing. For example, if the driver is an aggressive driver and tends to brake suddenly we observe a sudden sharp negative decrease in the acceleration reading, which also is shown in [18]. Figure 7 shows these behaviors on *Acc-Y*. By defining these behaviors, we extract features based on the type and the segment of a turn.

B. Turning Behavior Feature Extraction

In order to analyze the sensors data in fined-grained manner and extract features that can reflect the drivers unique behaviors each turn is divided into three different segments(e.g. *before, during, and after*-turn). Different types of sensors should be considered for feature extraction at different turn's segment. Since the acceleration data is changed through out the whole turn, it is used for feature extraction in all segments. However, *Gyro-Z* has nonzero values only during turn segment, this sensor reading is not considered for feature extraction in *before* and *after* the turn. For analyzing *Acc-Y* before a turn, different behaviors such as *Acc*, *De-Acc*, *Brake*, and *De-Brake* can be defined in this segment. These behaviors can be analyzed by their value, timing, and their changes. Figure 8 shows examples of different features in identifying different drivers. Specifically, Figure 8 (a) shows the features extracted before and during a right turn for four different drivers. Figure 8 (b) shows the features extracted before and during a left turn for two different drivers. Those extracted features are quiet helpful in identifying different drivers. Table II summarizes the features that we extracted from the each type of sensor Acc-Y and Gyro-Z. In total 200 features are extracted from all segments.

TABLE II
FEATURES EXTRACTED FROM ACC-Y AND GYRO-Z IN DIFFERENT CRITICAL TURNING STAGES.

Before-Turn
(Acc-Y) Acc, De-Acc,Brake, De-Brake:(value, time,changes), statistical features of each behavior and their first derivative, i.e., mean, max, min, mode, var.
During-Turn
(Acc-Y) Acc, De-Acc, Brake, De-Brake:(value, time,changes), statistical features of each behavior and their first derivative, i.e., mean, max, min, mode, var. ***(Gyro-Z)*** statistical features of the changes in the rotation angle from zero to 45-degree(middle point of a 90-degrees) and from 45-degrees to 90-degree
After-Turn
(Acc-Y) Acc, De-Acc, Brake, De-Brake:(value, time,changes), statistical features of each behavior and their first derivative, i.e., mean, max, min, mode, var.



(a) Four different drivers (b) Driver 3 & Driver 6

Fig. 8. Example of different drivers and corresponding features.

C. Driver Identity Classification

Gradient Boosting Tree based Classification. For feature classification, Gradient Boosting Tree (GBT) algorithm is being used [19]. GBT is chosen because it is famous for its robustness to different types of features with different scales. In other words, we do not need to normalize or whiten the feature data before classification which is required for other classifiers such as Support Vector Machine (SVM).

GBT involves three elements. A loss function that can be optimized, a weak learner to make the predictions, and an additive model to add weak learners to minimize the loss functions. The loss function depends on type of the problem. For classification, the logarithmic loss can be used. Decision tress are used as the weak learner in GBT. Trees are added once at a time to the existing trees in the model. Specifically the GBT tries to find a function to minimize the loss function. A fixed number of trees are added or training stops when the loss reaches the acceptable level.

In our study, we use the GBT implementation from the SQBlib library. The loss function is ”logloss”, with the shrinkage factor as 0.1 and 200 iterations. These GBT parameters optimized in term of accuracy based on our empirical data. After determining the loss function, a binary classifier is being built for each segment for each driver. We build binary classifier, because it can have higher accuracy in distinguishing one class versus other classes, whereas a multi-class classifier would have relatively lower accuracy in classifying multiple classes.

Single-turn Classification. Next, we discuss how we make the classifier for driver identification. As mentioned before, each turn is divided into different segments. For example, if we have N segmentsS_1, S_2, ... , S_n, we build D_n binary classifier Gradient Boosted Tree at each segment which D_n is the number of drivers. One turn from the test set is chosen randomly for driver identification. Each segment results a driver's id denoted as D_i. We integrate the results from all the segments to decide the driver's identity. In particular we use the mode function as equation 1.

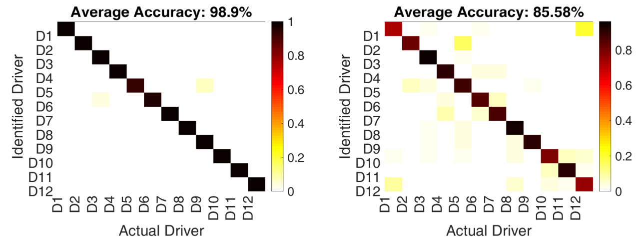
$$final_{DriverId} = mode(D_1, D_2, ..., D_N). \qquad (1)$$

The final result of the system is the *DriverID* which is resulted from most of the segments.

Multiple-turns Classification. Intuitively, we can use more than one turn for driver identification since it is common to encounter multiple turns in daily driving experiences. In this work, we use odd number of turns(*1, 3, 5, ...*) for driver identification. The intuition is that we conduct the majority vote concept over the results from multiple turns as well. For example, if we consider N_T turns in which (N_T mod $2 \neq 0$) for classification, the $final_{DriverId}$ can be resulted using the $mode$ function on the resulted from different turns as describe before. The final result of the system is the *DriverID* which is resulted from most of the turns. This process can reduce the driver identification error and increase the robustness of the system.

VII. Data Pre-processing

Coordinate Alignment. Commonly, a smartphone can be placed in any arbitrary position in a car. In other word, the phone's coordinate system may not be aligned with the vehicle's coordinate system. Therefore, the smartphone sensor data must be rotated to the vehicle's coordinate system to describe the driving behaviors. We adopt a rotation matrix R [16] to rotate the smartphone's sensor data to the vehicle's coordinate system. Specifically, the sensor data are rotated to the vehicle's coordinate system by using $\overrightarrow{V} = R \times \overrightarrow{P}$, where $\overrightarrow{P}$ is the sensor data vector in the smartphone's coordinate system, and $\overrightarrow{V}$ is the sensor data vector in vehicle's coordinate system. By using the coordinate alignment, the smartphone's in-vehicle position does not impact the performance of our system. We note that our system is not designed for the scenarios where the driver uses his/her phone while driving, which is illegal in most states and very reckless.



(a) Without stop events (b) With stop events

Fig. 9. Confusion matrix of identifying 12 drivers using a single right turn.

Noise Filtering. After rotating the smartphone's coordinate system, the next step is to remove the noises from the sensor data. The sensor data may have noises due to bad road conditions such as potholes and non-smooth surfaces. Such conditions may cause sudden high-power values in the sensor data. To mitigate these noises on the road, we first remove the bursts of high-power signals from the sensors data by using an outlier filter [20]. Then, we apply a moving average filter to remove the noises caused by the non-smooth road conditions. We empirically determine the size of the window used in the moving average filter to be 4 samples.

VIII. System Evaluation

A. Experimental Methodology

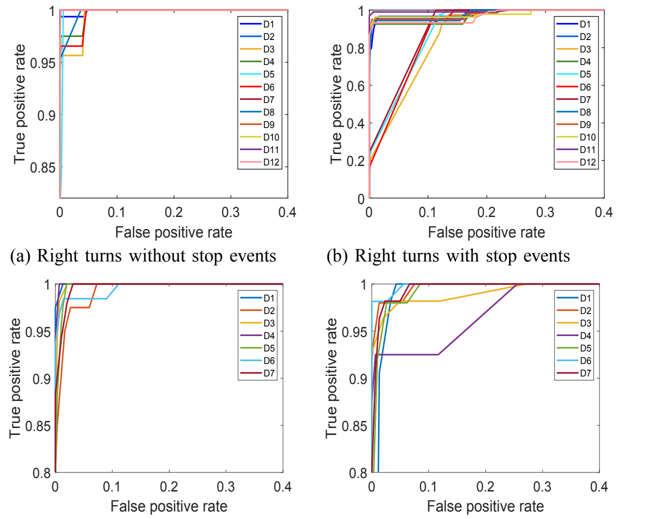
Experimental Setup. We conducted real driving experiments in our study. In particular, we have 12 drivers performing right and 7 drivers performing left turns. The experiments are conducted in the residential area, which contains two types of turns (i.e., right and left turns) with and without stop events. The stop events include stop signs and waiting for pedestrians and passing cars. We conduct the experiments at different times of different days, which cover different traffic conditions (e.g., light-traffic during weekends and heavy-traffic after working hours) and weather conditions (e.g., sunny and rainy days). In total, we extracted more than 2800 turns in our experiments. The experiments involve three cars of different models (i.e., one Dodge Caliber 2007, one Ford Taurus 2010, and one Jeep Explorer 2015). We develop an Android application for collecting the motion sensor data together with GPS locations. A smartphone (LG Nexus 5X phone) is fixed on the dashboard of the car running the data collection application during the experiments. The sampling rate of the sensors are set to 10 samples/second and data is processed using our system implemented by MATLAB offline.

B. Evaluation metrics.

True Positive Rate. The percentage of the testing turns from the target driver that are correctly classified as from that target driver.

False Positive Rate. The percentage of the testing turns from other drivers that are mistakenly classified as from the target driver.

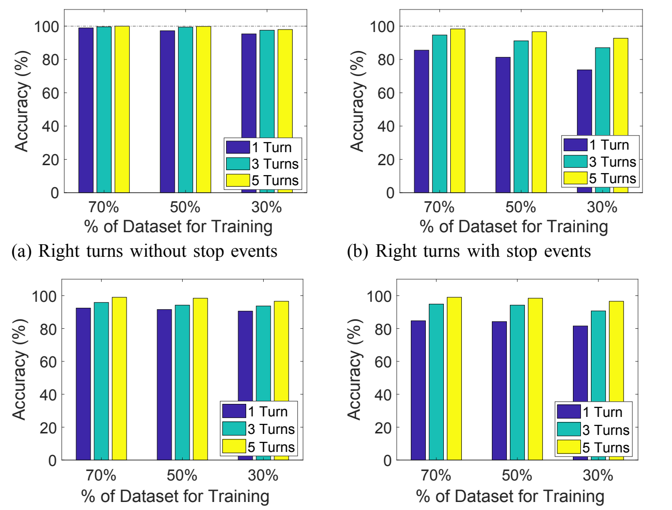
Receiver Operating Characteristic (ROC). ROC curve shows the trade-off between the True Positive Rate and the



(a) Right turns without stop events (b) Right turns with stop events

(c) Left turns without stop events (d) Left turns with stop events

Fig. 10. Impact study: ROC of different types of turns from different drivers.



(a) Right turns without stop events (b) Right turns with stop events

(c) Left turns without stop events (d) Left turns with stop events

Fig. 11. Impact study: average accuracy of different sizes of training and testing data.

False Positive Rate. The more the ROC curve close to the point (0, 1), the better the performance.

Confusion Matrix. The degree of color darkness in the matrix corresponds to the percentage of correctly classified turns.

Accuracy. The average of True Positive Rate of classifying all the drivers.

C. Driver Identification Performance

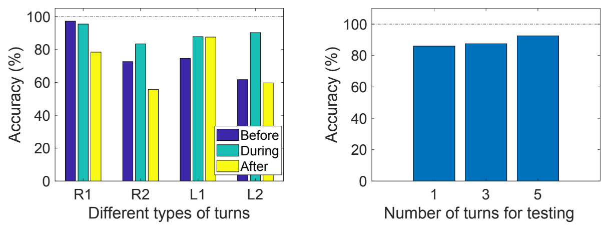
Figure 9 depicts the confusion matrix for identifying 12 different drivers on right turns with and without stop events, respectively. Each entry D_{ij} denotes the percentage of the turns conducted by drivers i was classified as from driver j. The diagonal entries show the average accuracy of identifying each driver. In our work, the training and testing data are not overlapped and from the same data set that includes multiple instances from several different turns. In particular, 70% of the turns of each driver are used for training and only one turn from the rest of the turns is used for testing. 100 combinations of turns are chosen as the training data randomly from each driver. It should be mentioned that we use the same size of training data for each drivers to ensure the results are not biased. The training size is determined by the minimum number of turns performed by a driver. We can see that the average driver identification accuracy is 98.91% and 85% for turns without and with the stop events, respectively. The results confirm that it is promising to use commodity smartphone to distinguish different drivers.

D. Impact of Various Factors

Impact of Turn Types. We next study the impact of different types of turns on the performance. As shown in Figure 10, different types of turns (i.e., *"right turns with/without stop events"* and *"left turns with/without stop events"*) have different ROC curves in identifying drivers. This is because drivers have different behaviors on different types of turns with different situations. In particular, we observe that ROC curve of turns with stop events is farther to the point (0, 1) than the turns without stop events. The reason is that different drivers tend to have more similar behaviors when turning with a stop event. These same behaviors result in worse performance in distinguishing different drivers. In addition, we also find that the True Positive Rate of identifying drivers on left turns is lower than on right turns. The reason is that when making a left turn, the drivers need to adjust their behaviours more cautiously to avoid the upcoming cars, the pedestrians, and bicycle rider from both left and right side [21]. All of these factors decrease the consistency of the driver's behavior on left turns than right turns. Those results show that our system could maintain the decent performance even though the different types of turns have some impacts.

Impact of Training Size and Number of Turns for Testing. Figure11 shows the performance of the system on different types of turns under different sizes of training set with different number of turns for testing. Particularly, we choose percentages of 70%, 50%, and 30% of each turn category data set for training the classifier, and adopt 1, 3, and 5 turns from the corresponding testing set for testing. The number of training turns per drivers for right turns without stop events is 23, for right turns with stop events is 36, and for left turns with/without stop events is 40 turns. We observe that our system can achieve the accuracy of 95.33% among 12 drivers on right turns without stop events when using only 30% turn data (i.e., 6 turns/driver) for training and only 1 turn for testing. As the size of the training or the number of turns for testing increases, the performance of the system improves as well. When having 70% turn data (i.e., 16 turns/driver) for training and 5 turns for testing, we can achieve 100% accuracy among 12 drivers for right turns without stop events. The results indicate that our system can achieve a very good accuracy with limited size of training and turns for testing in identifying different drivers, which ensures the convenience for usage on smartphones.

Impact of Different Segments of a Turn. In this section, we analyze the driver identification results in terms of the different segments of a turn. Figure 12 shows the system performance on four different types turns (i.e., right turns



(a) Accuracy for different turn types (b) Accuracy for different cars

Fig. 12. Impact study: average accuracy of differ types of turns and different cars.

without stop event *(R1)*, right turns with stop events *(R2)*, left turns without stop events *(L1)*, and left turns with stop events *(L2)*). All these results are calculated using 70% of the data set for training and only 1 turn for testing. We observe that on average, the results corresponding to before/after-turn stages are worse than the during-turn stage. Based on our experiments, the reason is that the driver can encounter more dynamic driving events before/after he performs the turn. While during the turn, the driver is not usually encountering any specific events. Those results show that the segments during a turn is more suitable for driver identification.

Impact of Different Types of Cars. We further analyze drivers' data from two different sedans (i.e., Dodge and Ford) to demonstrate that our system is car-independent. In particular, data collected from one care is used for training and 1, 3, and 5 random turns from the data set collected from second car are chosen for testing. As shown in Figure 12, the accuracy for classifying these two drivers is 85%, 87.5% and 92.5% respectively. This result shows that our system can identify different drivers with good accuracy even though their data for training and testing is coming from different cars, which suggests that our driver identification system is independent of vehicle types.

IX. Conclusion

This paper presents a low-cost and robust solution for identifying drivers using smartphones. Different from existing work using general driving behaviors, we focus on using drivers' turning behavior, which is the most complicated behavior on the road, to facilitate accurate driver identification with low training effort. We classify turns into different categories by considering the turns orientations and stop events. Furthermore, based on our extensive study of turning behaviors, we design the fine-grained segmentation to ensure the system can capture the distinct turning behaviors of different drivers under different traffic conditions. By analyzing the sensor data in each fine-grained segment, we determine the turning behavior related features and develop a Gradient boosting tree (GBT) based classifier for driver identification. Extensive experiments with more than 2800 turns collected from 12 drivers demonstrate that our system can identify different drivers with high accuracy and low false-positive rate under various real driving scenarios.

Acknowledgment

This work was partially supported by the National Science Foundation Grants CNS1814590, CNS1954959, CNS1566455, CCF1909963, CCF2000480, and ARO Grant W911NF-18-1-0221.

References

[1] D. Hallac, A. Sharang, R. Stahlmann, A. Lamprecht, M. Huber, M. Roehder, J. Leskovec *et al.*, "Driver identification using automobile sensor data from a single turn," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2016, pp. 953–958.

[2] M. Enev, A. Takakuwa, K. Koscher, and T. Kohno, "Automobile driver fingerprinting," *Proceedings on Privacy Enhancing Technologies*, vol. 2016, no. 1, pp. 34–50, 2016.

[3] H. Zhang, V. M. Patel, M. Fathy, and R. Chellappa, "Touch gesture-based active user authentication using dictionaries," in *2015 IEEE Winter Conference on Applications of Computer Vision*. IEEE, 2015, pp. 207–214.

[4] S. Furnell, N. Clarke, and S. Karatzouni, "Beyond the pin: Enhancing user authentication for mobile devices," *Computer fraud & security*, vol. 2008, no. 8, pp. 12–17, 2008.

[5] K. O. Bailey, J. S. Okolica, and G. L. Peterson, "User identification and authentication using multi-modal behavioral biometrics," *Computers & Security*, vol. 43, pp. 77–89, 2014.

[6] C. Zhang, M. Patel, S. Buthpitiya, K. Lyons, B. Harrison, and G. D. Abowd, "Driver classification based on driving behaviors," in *Proceedings of the 21st International Conference on Intelligent User Interfaces*. ACM, 2016, pp. 80–84.

[7] T. Wakita, K. Ozawa, C. Miyajima, K. Igarashi, K. Itou, K. Takeda, and F. Itakura, "Driver identification using driving behavior signals," *IEICE TRANSACTIONS on Information and Systems*, vol. 89, no. 3, pp. 1188–1194, 2006.

[8] S. Choi, J. Kim, D. Kwak, P. Angkititrakul, and J. H. Hansen, "Analysis and classification of driver behavior using in-vehicle can-bus information," in *Biennial workshop on DSP for in-vehicle and mobile systems*, 2007, pp. 17–19.

[9] M. Van Ly, S. Martin, and M. M. Trivedi, "Driver classification and driving style recognition using inertial sensors," in *2013 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2013, pp. 1040–1045.

[10] B. Wallace, R. Goubran, F. Knoefel, S. Marshall, M. Porter, and A. Smith, "Driver unique acceleration behaviours and stability over two years," in *2016 IEEE International Congress on Big Data (BigData Congress)*. IEEE, 2016, pp. 230–235.

[11] S. Ezzini, I. Berrada, and M. Ghogho, "Who is behind the wheel? driver identification and fingerprinting," *Journal of Big Data*, vol. 5, no. 1, p. 9, 2018.

[12] C.-H. Yang, C.-C. Chang, and D. Liang, "A novel gmm-based behavioral modeling approach for smartwatch-based driver authentication," *Sensors*, vol. 18, no. 4, p. 1007, 2018.

[13] K. Igarashi, K. Takeda, F. Itakura, and H. Abut, "Is our driving behavior unique?" in *DSP for in-Vehicle and Mobile Systems*. Springer, 2005, pp. 257–274.

[14] X. Zhang, X. Zhao, and J. Rong, "A study of individual characteristics of driving behavior based on hidden markov model," *Sensors & Transducers*, vol. 167, no. 3, p. 194, 2014.

[15] J. F. Júnior, E. Carvalho, B. V. Ferreira, C. de Souza, Y. Suhara, A. Pentland, and G. Pessin, "Driver behavior profiling: An investigation with different smartphone sensors and machine learning," *PLoS one*, vol. 12, no. 4, p. e0174959, 2017.

[16] Y. Wang, J. Yang, H. Liu, Y. Chen, M. Gruteser, and R. P. Martin, "Sensing vehicle dynamics for determining driver phone use," in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 2013, pp. 41–54.

[17] J. Yu, H. Zhu, H. Han, Y. J. Chen, J. Yang, Y. Zhu, Z. Chen, G. Xue, and M. Li, "Senspeed: Sensing driving conditions to estimate vehicle speed in urban environments," *IEEE Transactions on Mobile Computing*, vol. 15, no. 1, pp. 202–216, 2015.

[18] J. Yu, Z. Chen, Y. Zhu, Y. J. Chen, L. Kong, and M. Li, "Fine-grained abnormal driving behaviors detection and identification with smartphones," *IEEE transactions on mobile computing*, vol. 16, no. 8, pp. 2198–2212, 2016.

[19] C. Becker, R. Rigamonti, V. Lepetit, and P. Fua, "Supervised feature learning for curvilinear structure segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2013, pp. 526–533.

[20] S. Verboven and M. Hubert, "Libra: a matlab library for robust analysis," *Chemometrics and intelligent laboratory systems*, vol. 75, no. 2, pp. 127–136, 2005.

[21] S. Clearly. (2017) Why is a left turn one of the most dangerous driving maneuvers and should it be banned? [Online]. Available: https://www.seanclearypa.com/blog/left-turn-dangerous-maneuvers-banned/