This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Smartphone-Based Adaptive Driving Maneuver Detection: A Large-Scale Evaluation Study

German Castignani, Thierry Derrmann, Raphaël Frank, and Thomas Engel

***Abstract*— The proliferation of connected mobile devices together with advances in their sensing capacity has enabled a new distributed telematics platform. In particular, smartphones can be used as driving sensors to identify individual driver behavior and risky maneuvers. However, in order to estimate driver behavior with smartphones, the system must deal with different vehicle characteristics. This is the main limitation of existing sensing platforms, which are principally based on fixed thresholds for different sensing parameters. In this paper, we propose an adaptive driving maneuver detection mechanism that iteratively builds a statistical model of the driver, vehicle, and smartphone combination using a multivariate normal model. By means of experimentation over a test track and public roads, we first explore the capacity of different sensor input combinations to detect risky driving maneuvers, and we propose a training mechanism that adapts the profiling model to the vehicle, driver, and road topology. A large-scale evaluation study is conducted, showing that the model for maneuver detection and scoring is able to adapt to different drivers, vehicles, and road conditions.**

***Index Terms*— Driving maneuver detection, anomaly detection.**

I. INTRODUCTION

MONITORING driving activities is gaining increasing attention. In this context, ubiquitous computing is becoming a promising alternative to traditional telematics systems. The increasing availability of advanced smartphones equipped with a large range of sensors (e.g., GPS, accelerometer, gyroscope, magnetometer) and network interfaces allows the development of connected applications that can be used to analyze and evaluate driving behavior in real time.

Two main application fields have been identified for which such systems could generate new revenue streams. The first includes fleet management systems. Traditionally, dedicated embedded systems have been mounted inside vehicles to monitor driving activities. It is expected that in the near future such systems will become obsolete as dedicated mobile applications for smartphones can be used for the same purpose. Further, real-time vehicle information gathered through a wireless On-Board Diagnostics (OBD-II) adapter connected to the smartphone would allow collecting more advanced metrics such as fuel consumption, RPM or vehicle fault codes. Examples of such applications include Automatic, Dash or Enerfy, which are publicly available for Android and iOS. The primary objective of these systems is to monitor the location and condition of corporate vehicles and motivate the driver to behave more efficiently (e.g., reduce the energy footprint, reduce travel time).

The second application field is the insurance telematics market, including *Pay As You Drive* (PAYD) or *Usage Based Insurance* (UBI). By using smartphones, the insurance companies do not have to provide, mount and maintain expensive monitoring hardware. Also, individual drivers are more likely to use their personal mobile device as a telematics system instead of a black box that they cannot control [1].

As opportunities are growing, research has focused on how mobile phone sensors can be exploited to provide meaningful data that accurately describes the risk profile of a driver. Several studies based on the Virginia Tech 100-Car data set [2] have shown the high correlation that exists between sensor data and driving riskiness (e.g., involvement in crashes and near-crashes) [3], [4]. The common approach to measure driver riskiness is to compute a single score (typically between 0 and 100) that represents the driving skills of a user. This is not a trivial task since many factors need to be taken into account to obtain fair and comparable results. Current approaches typically rely on fixed thresholds to identify risky driving maneuvers (see Section II). The most common maneuvers that can be detected using mobile phone sensors are: rapid acceleration, hard braking, speeding and aggressive cornering. However, considering the large number of different vehicles with a variety of performance and mechanical characteristics, it is clear that a static profiling approach will not provide a fair, representative and comparable scoring metric for every driver and vehicle combination.

In this article, we propose an adaptive driving maneuver detection method that relies on a Multivariate Normal distribution (MVN) to build a statistical model of the user's driving characteristics. We first introduce a feature set obtained by fusing the output of different mobile phone sensors which can potentially be used to detect and classify lateral and longitudinal driving maneuvers. Next, we describe how the continuous training of the model is performed. By means of an extensive field experiment conducted on a test track to create labelled datasets, we evaluate how different feature sets impact classification and estimate the hyperparameters of the learning algorithm. We propose a risk function that can be used to dynamically adjust the score of the driver by taking into

Manuscript received May 25, 2016; revised October 28, 2016; accepted December 24, 2016. This work was supported by Fonds National de la Recherche, Luxembourg, under Grant POC/13/07. The Associate Editor for this paper was M. Brackstone.

The authors are with the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg, L-2721 Luxembourg City, Luxembourg (e-mail: German.Castignani@uni.lu; Thierry.Derrmann@uni.lu; Raphael.Frank@uni.lu; Thomas.Engel@uni.lu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITS.2016.2646760

account the number, severity and sequence of detected events. In order to validate the scoring function, we conducted a large scale experiment, capturing more than 2.8 million kilometers of driving data from around 4,800 distinct drivers in the greater region of Luxembourg.

The remainder of this article is organized as follows. In Section II, we introduce the related work on smartphone-based driver profiling mechanisms. In Section III, we present the theoretical aspects of the proposed MVN-based model, including the input definition, feature transformation and training mechanism. In Section IV, we present a model parameterization study for the event classifier based on experimental data collected with smartphones in both a controlled environment (test track) and public roads. A large scale validation study for the proposed scoring function is presented in Section V. Finally, in Section VI, we conclude the paper and present perspectives for the future work.

II. Related Work

Several methodologies of driver profiling and maneuver classification have been proposed in the past, and with the rising market penetration of smartphones, the potential user base of smartphone-based techniques has increased recently. Engelbrecht *et al.* [5] provide a survey giving an overview of different smartphone-based sensing applications inside vehicles. More specifically, these applications range from the detection of maneuvers, drunk driving and accidents to driver assistance systems. The authors stress that the adoption of these systems can be slow, an issue which we address in this article through an incentivized, gamified study. Given the focus of our work on maneuver detection, we mainly discuss the existing platforms on this topic in the remainder of this section. They typically involve the integrated Inertial Measurement Units (IMU) along with satellite-based positioning (GNSS/GPS) data.

In a GNSS-only study, Wahlström *et al.* [6] propose a dangerous cornering detection system based on Kalman-filtered GNSS data, and show that it can be sufficient for the detection of the majority of cornering events in their evaluation by means of a study with three smartphones. As they report 40% false alerts or missed maneuvers, they discuss the possibility of taking IMU data into consideration to improve the proposed test statistic and detection accuracy.

In this vein, Paefgen *et al.* [7] propose a technique making use of both GPS and IMU data, while relying on predefined thresholds (e.g., an accelerometer output of $0.1g$ for acceleration and braking events and $0.2g$ for steering). The authors present a measurement study to compare event detection using smartphone sensors to those detected using a fixed-position IMU. They observe that the obtained event count distribution matches different statistical distributions in smartphones and IMUs; this was mainly due to variations in smartphone-to-car fixing and positioning inside the vehicle, which can dramatically affect the pre-established thresholds. However, the authors find only weak correlations between smartphones and IMU-based events and describe some possible solutions to enhance smartphone event detection, including a dynamic calibration process [8].

Johnson and Trivedi [9] present a Dynamic Time Warping (DTW) based driver profile algorithm, MIROAD, using smartphone sensors and GPS. They evaluate the performance of different sensor fusion sets to detect lateral and longitudinal movements. By evaluating over 200 driving events, the authors show that the sensor fusion set composed of the x-axis (i.e., gravity axis) rotation rate, y-axis (i.e., lateral movements) acceleration and pitch, provides the best classification performance using DTW. This study relies on maneuver templates that generalize to different cars, but it is not clear whether other phones and drivers would yield similar classification results.

Eren *et al.* [10] propose a driver profiling algorithm that distinguishes between risky and safe drivers, using a supervised learning approach. As sensing data, they consider IMU data to detect start and end times of driving events (e.g., sudden maneuvers, aggressive steering, braking or acceleration) using moving averages and empirical thresholds. The authors compute the similarity of each event to template data (i.e., risky and safe event patterns that had been previously collected) using Dynamic Time Warping (DTW) and apply Bayesian classification to decide whether a driver is risky or safe. They present an evaluation study for fifteen drivers using iPhones and fixed departure and arrival points, showing a successful classification rate of 93.3 %. However, using their method to compute event template data appears to be unfeasible for a large and heterogeneous sensing environment, considering different types of vehicles and devices.

Joubert *et al.* [11] use in-vehicle data recorder (IVDR) data, more precisely velocity and accelerometer data, from 124 drivers (driving different vehicles) and build a risk space over the full data set, i.e. an aggregate distribution over all drivers. They use lateral, longitudinal and vertical motion features. Their approach is population-based, i.e. a single, global model is built centrally, yielding identical risk estimates at identical feature values, independent of the drivers and cars. It is a matter of discussion whether a car's performance (in particular engine and brakes) should impact the definition of which acceleration values are to be considered safe and unsafe.

In contrast to these studies, we want to propose a model that generalizes to different driver, car and smartphone combinations, using unsupervised learning of both GPS and IMU derived features. We learn the distribution of these features for each individual's driving behavior, car and phone combination, and evaluate it online, directly on the phone. In previous work [12], we described a set of experiments to analyze the performance of smartphones to detect risky driving events. In particular, we introduced a sensor fusion algorithm and a Fuzzy System to detect aggressive steering, acceleration and braking maneuvers. We evaluated the system in a small experiment using a single car and smartphone combination. By contrast, in this article, we avoid relying on a prior training set specific to the used car and phone, but make use of some of the signal processing techniques developed. To reflect the main advantage–the adaptability–of our unsupervised model, we evaluate it in a large-scale field study with 4,800 drivers using different cars and smartphones. We compare the driver scores to a corresponding survey and eco-driving scores, which we preferred over a direct comparison to previous, supervised models that would have required labeled driving maneuver data for arbitrary car/phone combinations.

III. MVN-Based Driving Maneuver Detection

In this section, we describe how to use a Multivariate Normal (MVN) model to detect risky driving maneuvers and profile drivers. In the research literature, we can find several examples of anomaly detection problems that have been addressed using MVNs [13], [14]. However, to the best of our knowledge, this is the first time an MVN approach has been considered for risky driving maneuver detection. As introduced in Section II, existing solutions use driver profiling algorithms that identify high sensor outputs or pre-established patterns in order to declare a risky driving maneuver, and therefore decrease the driver's score. However, given the heterogeneity of mobile devices and vehicles, such a profiling system needs to adapt its parameters to each individual context. Traditional fixed threshold or supervised Machine Learning techniques (e.g., Neural Networks, Naive Bayes, SVM) may be suitable for designing such a driver profiling mechanism but require previous knowledge of the training samples, indicating the expected values of the input variables at the exact time a risky maneuver is performed. Consequently, the generation of such a training data set for risky driving maneuvers for many different devices and vehicles becomes unfeasible. Moreover, we cannot expect target users of the driver profiling platform to manually label risky driving maneuvers in a controlled training phase. Finally, an MVN-based anomaly detection seems to fit well with the problem we are modeling. First, it provides a high flexibility by dynamically adapting to different driving conditions with a periodic update of its parameters. Also, given that it does not require any preliminary knowledge on the driving maneuvers it has to detect, it can be applied to any type of vehicle and mobile device.

In general, the MVN model has many advantages over other approaches for anomaly detection: (i) There is no need to establish rule sets or thresholds to detect risky maneuvers. The probability of an observation enables the assessment of driver behavior in a continuous fashion rather than binary classification (i.e., distinguishing between a risky and a non-risky driving maneuver). (ii) The model can be estimated efficiently on a mobile phone and can be continuously retrained to improve the parameter estimates. (iii) It is easy to interpret the model, since the sampling distribution probability density can easily be represented visually, e.g. by contour plots. (iv) Some of the input variables (e.g. GPS acceleration and bearing variation) are normally distributed and thus lend themselves naturally to this kind of model.

A. General Architecture

The general architecture of the proposed model can be divided into two phases. In the *initialization* phase, the mobile device starts collecting input variables that are then transformed into a two-dimensional feature space that represents the axes of movement of the vehicle. These features are then stored in a training buffer that will finally serve to generate the first MVN model using the Maximum Likelihood Estimator (MLE). After the initialization phase, the current MVN model is adapted to new data samples regularly in the *retraining* phase and updated, as shown in Fig. 1, while also being used for the maneuver detection. As input variables, we consider motion sensor variables that are available in modern smartphones. The first component of the input vector $o = [\sigma(j), \mu(y), \Delta v, \Delta b, v]$ is the standard deviation of the jerk, $\sigma(j)$. The jerk is calculated as the time derivative of the norm of the acceleration vector using the device's accelerometer. This variable is able to provide a better fit to aggressive driving maneuvers than the raw acceleration, which can be greatly affected by vibrations due to speed. Then, we consider the average yaw rate, $\mu(y)$, the angular velocity measured with the device's magnetometer. Finally, we consider GPS input: the speed, v, the speed variation, Δv, and the bearing variation, Δb.

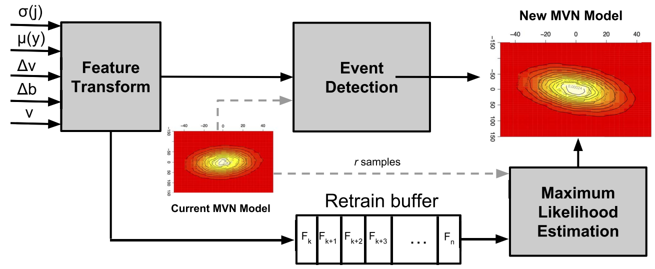


Fig. 1. Maneuver detection and retraining phase.

Note that the GPS variables provide a much lower sampling rate than motion sensor variables (i.e., 1 Hz against 20 to 200 Hz). All variables are synchronized in the mobile device by buffering motion sensing data between two consecutive location updates. The input variables $\sigma(j)$ and $\mu(y)$ are then calculated over this buffer. The input vector refresh rate is fixed at a rate of 10 Hz in order to provide an accurate short-term view of potential maneuvers.

The proposed MVN model considers a two-dimensional feature space $\mathbf{X}$ to obtain an interpretable model of longitudinal and latitudinal features. We define a feature $\mathbf{x} \in \mathbf{X}$ as a function of the input data o having the following properties: i) extreme values of the feature represent risky or abnormal driving maneuvers, ii) feature distribution can be approximated by a Gaussian distribution. In Section IV, we propose different feature sets based on the input variables in o and we evaluate their suitability for identifying risky driving maneuvers.

B. Model Definition and Training Methodology

The MVN model can be defined by the multivariate Gaussian probability density function in Eq. 1:

$$p(\mathbf{x} \mid \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\tfrac{1}{2} (\mathbf{x} - \mu)^T \, \Sigma^{-1} (\mathbf{x} - \mu)\right) \tag{1}$$

$p(\mathbf{x} \mid \mu, \Sigma)$ is the multivariate Gaussian probability density function, where $\mathbf{x}$ denotes the features that we consider, d denotes the dimension of the feature space, and $\theta = (\mu, \Sigma)$

denotes the parameters of the distribution. Note that in our approach, we want to dynamically estimate θ on each individual mobile device. Using the probability density function with the estimated θ, we can calculate the likelihood of a given observation o. Based on the likelihood, we can distinguish between a common observation (i.e., that corresponds to normal driving behavior) and anomalies (i.e., an abnormal driving maneuver).

The estimation of the model parameters implies a training process. The system is initially trained after m input samples and then a retraining phase is triggered every n subsequent input samples in order to update the model parameters. In both the initial training and the retraining phases, we estimate the set of parameters θ using MLE shown in Eq. 2.

$$\begin{cases} \widehat{\mu} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(\mathbf{x}_i) \\ \widehat{\Sigma} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})^{\mathrm{T}} \end{cases} \tag{2}$$

As for the retraining phase (see Fig. 1), the model is built by considering not only n new samples collected since the previous model update but also a number of samples r that are resampled from the previously computed model. The MLE update is computed over this sample set and a new, updated model is obtained.

However, updating the model is a sensitive operation: there is a trade-off between the performance of the risky maneuver detection and the adaptability of the model. In other words, even though a frequent retraining strategy (i.e., a low n and a low r) results in a rapid adaptation of the model to new conditions (e.g., a change in the driver's driving style or road conditions), it can greatly impact the quality of maneuver detection by incorporating short-term data samples into the model. The objective is always to avoid updating the model with low-variance input samples, i.e., samples that reflect a monotonous driving behavior, like driving on a straight line at constant speed. In order to verify that the model update is useful, we compute the determinant of the covariance matrix of the new n data samples.

In more detail, we use a criterion based on the determinant of the covariance matrix of the newly-collected samples (n) and the current model resampled data (r), as shown in Eq. 3. The rule used to update the current model is expressed in Eq. 4, in which we simply compare the ratio u to a predefined limit. This predefined limit is empirically defined in Section IV.

$$u := \frac{det(\Sigma_n)}{det(\Sigma_r)} \tag{3}$$

$$\begin{cases} u \geq u_{lim} : \text{update} \\ u < u_{lim} : \text{continue} \end{cases} \tag{4}$$

Using this rule, we can guarantee that we do not degrade the sampling distribution by fitting data with very low variance. The newly-computed model reflects driver behaviour and can adapt to car and driving style changes, as its parameters are periodically re-estimated. Thus, the anomaly detection can recover from abrupt changes in the environment.

C. Driving Maneuver Detection

For every given input o and its corresponding features $\mathbf{x}$, the proposed model outputs a metric for the severity (or risk factor), that can later be used to score the driver. To this end, we propose a cut-off quantile value, Q_{lim}, which marks the quantile limit between very common observations and those we consider relevant for anomaly detection. Q_{lim} defines a cut-off probability value, p_{lim}, which can be used to classify the samples. Let p_{lim} be the model-dependent inverse CDF value of Q_{lim}, and $Q(\mathbf{x})$ the (approximate) quantile of the feature transform $\mathbf{x}$ of observation o, we can then propose the severity metric of Eq. 5.

$$s(\mathbf{x}) := \begin{cases} 0 & \text{if } p(\mathbf{x} \mid \theta) \geq p_{lim} \\ 1 - \dfrac{Q(\mathbf{x})}{Q_{lim}} & \text{if } p(\mathbf{x} \mid \theta) < p_{lim} \end{cases} \tag{5}$$

Eq. 5 defines a normalized metric of how anomalous an observation is. Note that the value of Q_{lim} (and correspondingly, p_{lim}) shifts the limit quantile of observations of interest. If a lower value of Q_{lim} is chosen, more observations will be considered to be anomalies but these additional observations will have lower severities. In this article, we set $Q_{lim} := 0.01$, or approximately $\mu \pm 2.5\sigma$, but lower values of Q_{lim} would yield comparable results (with additional reported outliers having low severities). We chose this value in accordance to the three-sigma rule, while slightly reducing the limit quantile to capture samples leading up to outliers as well.

$$D_m(\mathbf{x_1}, \mathbf{x_2}) = \sqrt{(\mathbf{x_1} - \mathbf{x_2})^T \Sigma^{-1} (\mathbf{x_1} - \mathbf{x_2})} \tag{6}$$

$$s(\mathbf{x}) := \begin{cases} 0 & \text{if } D_m(\mathbf{x}, \mu) < D_{lim} \\ 1 - \dfrac{e^{-\frac{D_m(\mathbf{x},\mu)^2}{2}}}{Q_{lim}} & \text{if } D_m(\mathbf{x}, \mu) > D_{lim} \end{cases} \tag{7}$$

$$with\ D_{lim} := \sqrt{-2 \cdot \log(1 - Q_{lim})}$$

In the bivariate case, we can rewrite the severity using the Mahalanobis distance (Eq. 6) [15]. Eq. 7 shows that we can then compute the severity only using this distance. Since we are only interested in the low probability regions of the severity, we can thus avoid sampling and constructing a large Empirical Cumulative Distribution Function to approximate the quantiles. This improves the computation precision of the metric because the need for approximation is eliminated.

IV. Model Parameterization Based on Closed-Loop Experiments

In this section, we explore the design parameters of the MVN driver profiling model based on real experiments that we carried out on a test track and public roads. In the testbed, we create labelled data by logging when risky driving maneuvers were performed, which we can then use to choose the design parameters of our model.

A. Testbed Setup

The experiments were carried out using a Toyota Prius and a Samsung Galaxy S4 (GT-I9505) mounted in a car holder.

The first design parameter we need to consider is the feature set. To this end, we carried out a number of driver

TABLE I
FEATURE SETS

	FS_1	FS_2	FS_3
f_1	Δv	$\Delta v \cdot \sigma(j)$	$\Delta v \cdot \sigma(j) \cdot v$
f_2	Δb	$\Delta b \cdot \mu(y)$	$\Delta b \cdot \mu(y) \cdot v$

maneuvers in a controlled environment in two different driving modes, namely *calm* and *aggressive*. We focused on three baseline scenarios: (1) acceleration, braking and steering, (2) slalom maneuvers, and (3) U-turns. All these maneuvers were performed in the CFC Luxembourg [16], a large driving circuit in the north of Luxembourg, and have been repeated several times, alternating between calm and aggressive driving patterns. For the acceleration, braking and steering scenario, we selected a 550 m long oval track inside the main track, composed of two straight lines of 240 m each and two curves. In this case we alternated six aggressive and calm laps (see Fig. 3a). For the slalom scenario, we have alternated slalom steering phases at low and high speed (25 km/h and 35 km/h respectively) on a straight line with obstacles (see Fig. 3b). Finally, the U-turn maneuvers were made on a short oval track, doing an aggressive U-turn (at 45 km/h) in the first curve and a calm U-turn (at 25 km/h) in the second curve (see Fig. 3c).

The second design parameter is the choice of the retraining strategy, which is defined as the number of new samples to consider for retraining (r) and the optimal update criterion limit (u_{lim}). To determine this, we collected a much larger data set that included urban roads and highways. During this trip, which lasted 40 minutes and traversed a 30 km path, we considered a mix of driving environments. The first half of the trip includes urban and suburban driving (i.e., maximum speed of 70 km/h), while the second half of the trip used a highway, where we attained a maximum speed of 150 km/h. During the whole trip, we triggered thirty risky maneuvers (hard acceleration, braking, steering and slalom) whose start and end time were manually logged in the mobile device by slightly pressing the touchscreen at the beginning and end of the maneuver.

B. Feature Selection

In order to identify the most suitable feature set, we have used the test track driving maneuvers data set and computed offline the performance of the different feature sets in terms of the output severity of the MVN classifier. In order to obtain an interpretable model that can make use of the Mahalanobis distance (cf. Sec.III-C), we have focused on three two-dimensional candidate feature sets. They represent the possible risky maneuvers of the driver in the longitudinal and lateral axes of movement.

1) Candidate Feature Sets: Table I summarizes the components (f_1 and f_2) of the three feature sets (FS). FS_1 simply consists of the GPS input data. This is the baseline feature set, and we expect the features to be nearly independent of each other, and thus not mandatorily require a multivariate model. The second feature set that we consider consists of the product of the GPS input data and the corresponding sensor outputs,

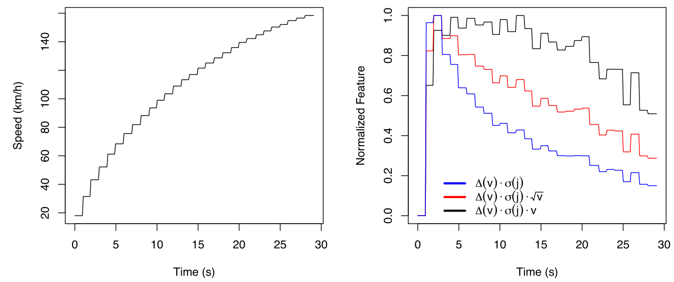


Fig. 2. Mitigating heteroscedasticity of acceleration.

i.e. $\sigma(j)$ and $\mu(y)$. Hence, the features will indicate if there is constructive consensus between the input values. The third feature set incorporates the current speed (v) into the product, so as to mitigate the heteroscedasticity (i.e. heterogeneous variance) of the features. In more detail, Fig. 2 shows the evolution of the speed at maximum possible acceleration of a Toyota Prius in a straight line. To do this experiment, we started from 20 km/h and we fully pushed down the accelerator pedal until the maximum speed (around 150 km/h) was attained. We can observe the decay of Δv for increasing speed, resulting in a reduced variance of the $\Delta v \cdot \sigma(j)$ term in the feature. In order to mitigate this effect, we added a new term to the product in the third feature. As shown in Fig. 2, we finally chose to use speed as a linear function, since we did not wish to overfit the profile of this particular car, and to avoid penalizing more powerful cars with a more linear acceleration curve. Note that we want the model to capture the changing variance of acceleration in relation to speed without becoming overly sensitive to the (relatively low) acceleration at high speeds. Note that we approximated the distributions using a multivariate normal, but that speed generally follows a multimodal distribution. However, our empirical data shows that the product of the two first factors of each feature $\Delta v \cdot \sigma(j)$ and $\Delta b \cdot \mu(y)$ outweighs the speed factor and the distribution remains approximately normal, which is sufficient for anomaly detection purposes. Also note that the choice of this feature set impacts the training strategy, since it benefits from continuous updates of the distribution to the current speed profile on unvarying trips.

2) Event Detection Performance: Figures 3a, 3b and 3c show the performance of the different feature sets in the three scenarios. To measure performance, we show the severity $s(\mathbf{x})$ output by the MVN classifier during the experiments. Note that to initially train our classifier, we drove 25 km around the test track, in different modes of driving (aggressive, normal and calm). The zones where the driver performed risky maneuvers are highlighted in red in Fig 3. In general, we can observe that FS_2 and FS_3 provide a more accurate severity output during such maneuvers. Also, during the calm driving phases (i.e., those without red highlight), FS_3 provides fewer false positives, not triggering any high severity during calm driving.

C. Training Strategy and Hyperparameters

1) Model Update Frequency: In order to improve the performance of the model, it is important to set up the correct

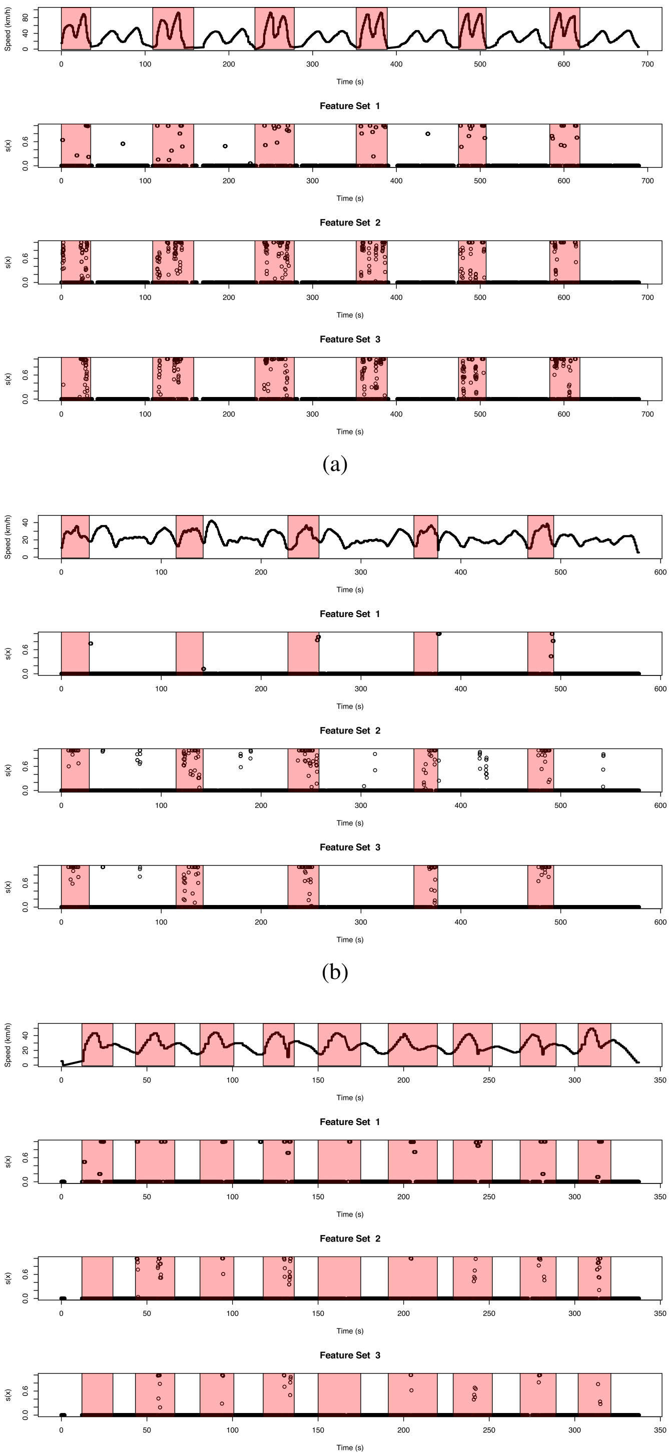


Fig. 3. Detection performance: Severities computed from different feature sets on a labelled dataset. (a) Acceleration, braking and steering. (b) Slalom. (c) U-turns.

training rates, i.e., how often the model is re-estimated and how many samples are drawn from the current model and combined with the new observations. In the following, we show the performance of the maneuver detection by considering a sensitivity analysis of the training parameters. For this analysis, we consider the second data set, consisting of 30 risky driving maneuvers on public roads. We replayed the complete data set offline, considering a variable number of new samples (n) and a variable number of samples drawn from the previous distribution (r), both varying between 100 and 3,000, with a step of 100, producing 900 total combinations. The main objective of this analysis is to estimate the performance of the event detection over the whole data set for the different combinations of n and r, and to determine optimal values for those parameters. As the performance metric, we consider the proportion of samples with non-zero severity ($s(\mathbf{x}) > 0$) falling between the start and end time of an event, i.e. the recall of the classifier. Fig. 4a shows the results of this analysis as a heatmap. Based on Fig. 4a, we observe that the performance of the classifier degrades with a low number of new samples, (i.e., $n < 700$). We can see that the best training strategy makes use of a mixture of both the previous distribution samples and new observations. We can also infer that there is a trade-off between the update frequency and the performance of the classifier. A very frequent update rule degrades the performance of the classifier, since it is very likely that the samples generated during this short retrain phase are insufficiently representative of the driving style. A lower update frequency, on the other hand, may not react to sudden changes in the scenario and lead to wrong classification. For the remainder of the analysis, we use $n = 1200$ and $r = 2900$, since they provide the best classification recall.

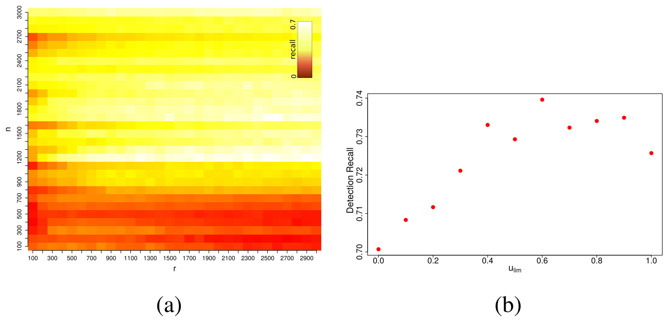
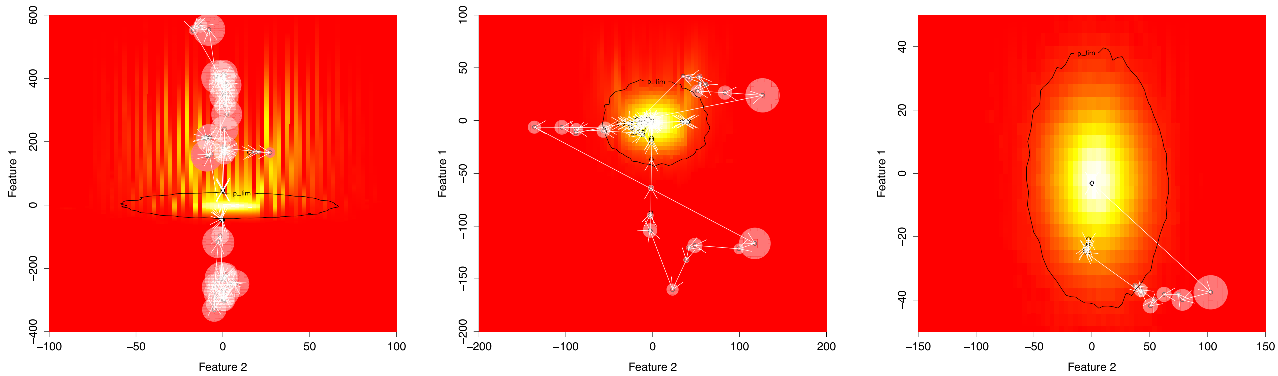


Fig. 4. Event detection recall. (a) Recall wrt. n and r. (b) Recall wrt. u_{lim}.

2) Model Update Criterion: In the previous analysis, we computed the performance of the classifier for different combinations of r and n by always using a simple update rule, which unconditionally triggers an update of the model every n samples. However, in some cases, these newly collected samples may not be representative of typical driving behavior, e.g. during constant-speed driving in a straight line on a highway. Since such low-variance data sets may bias the model, we added a criterion to detect and potentially reject them, and thus selectively update the model. As described in Section III-B, we introduce u as the quotient between the determinant of the covariance matrix of the new data samples and the determinant of the covariance matrix of the previously computed model. This is a relative measure of the variance of the new data with respect to the data that has served to build the current model. Fig. 4b illustrates the performance of the classifier for different values of u_{lim}. The additional criterion updates the model only if $u \geq u_{lim}$. We observe that the maximum performance obtained in the



(a) (b) (c)

Fig. 5. Risky driving maneuvers in the feature space, different maneuvers on the same model (axes scaled). (a) Acceleration and braking. (b) Slalom. (c) U-turn.

results presented in Fig. 4a (which sets $u_{lim} = 0$) can be improved by increasing u_{lim}. We observe the best performance for $0.5 < u_{lim} < 0.9$, which implies updating the model only if the variance of the new data set is greater than at least half the variance of the previously computed model. We base our analysis on $u_{lim} = 0.6$, where we have observed optimum recall.

D. Maneuver Riskiness Based on Severity

So far, we have identified the parameters of the proposed driving maneuver detection. For many applications, it is sensible to provide a metric that describes the level of riskiness of a certain maneuver for driving scores or the computation of profiles. Based on the severity metric, we can express how anomalous a observation $\mathbf{x}$ is with respect to the current model. However, in order to analyze riskiness, we need to explore how the consecutive observations corresponding to an event evolve in the two-dimensional feature space over time. Using the severity metric $s(\mathbf{x})$ defined above, together with the Mahalanobis distance (see Eq. 6), we derive the risk function of Eq. 8.

$$R(\mathbf{x}) := \frac{s(\mathbf{x})}{2}\left(\frac{1-\cos\alpha}{2} + \frac{2}{\pi} atan\Big(D_m\left(\mathbf{x_i}, \mathbf{x_{i-1}}\right)\Big)\right) \quad (8)$$

The definition of Eq. 8 is motivated by the idea that large variations in the feature space during an anomaly should be more strongly penalized. Regular anomalies manifest themselves as a temporal cluster of similar values, whereas more extreme maneuvers display more variability and cover larger distances within the feature space, so the angle between consecutive samples and/or the distance between them are large. In order to create a normalized risk metric in $[0, 1]$, we normalize the arctangent of the Mahalanobis distance and also evaluate the Haversine function of the relative angle α between consecutive samples $(\mathbf{x_{i-1}}, \mathbf{x_i})$ in the feature space. These values represent the similarity of two consecutive samples and are weighted using our severity metric $(s(\mathbf{x}))$. Thus, the risk function reflects the trajectory of the anomaly within the feature space.

The risk function is accumulated over a trip and can serve as a basis for driver scoring. Using this metric, risky maneuvers of a longer duration are more comparable to short maneuvers (e.g. sudden lane changes), as similar consecutive samples' risk values are attenuated by the distance and angle coefficient. Fig. 5 illustrates a set of example risky maneuvers extracted from the baseline events recorded at the test track. We show the MVN model in a heatmap representation and a sequence of consecutive observations connected with arrows. Note that the figures have different scales but show the same model, and that the observations $\mathbf{x}$ are represented by white dots of size proportional to their risk function value $R(\mathbf{x})$. We consider maneuvers with variance in both dimensions (e.g. steering while accelerating) to represent a greater risk than the elementary maneuvers in one dimension (e.g. just steering). The slalom trace shows that the risk function amplifies abrupt changes in the features. The U-turn trace shows that the risk function value is high for the first observations, which are considered to be anomalies, and that the subsequent anomalies give lower risk values, as they are similar and belong to the same driving event. Overall, this shows that the risk function allows an adequate evaluation the risk of individual maneuvers.

V. Large-Scale Evaluation

A. Implementation and Deployment

The MVN classifier and scoring function were integrated into a publicly available Android and iOS application in close partnership with a local car insurance company. The application allows individual drivers to detect driving events and submit their trips to a server for score computation. A gamification layer, including a variety of badges and social network interactions, was added in order to encourage drivers to contribute to the platform. In the mobile application, an MVN model is built and retrained on each mobile device. Users record their driving and use the MVN model to classify driving events, as described in Section III. The output of the MVN classifier is sent to a remote server at the end of each trip. The server then calculates a drive score based on the density of MVN events, i.e., the ratio between the risk computed using Eq. 8 and the distance driven. Trips with negligible risk density (lower than a constant minimum threshold)

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

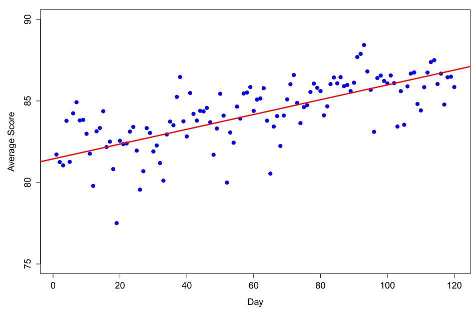


Fig. 6. Evolution of average score.

are assigned a drive score of 100 points drive score. Drive score linearly decreases (down to zero) with increasing risk density. This drive score is then adjusted using a variety of contextual information, including historical weather and daylight information in order to more heavily penalize events occurring in bad environmental conditions. For every trip recorded, the driver gets a number of game points that are accumulated for each trip. The number of game points on each trip is proportional to the drive score, the distance driven and the frequency of usage of the application.

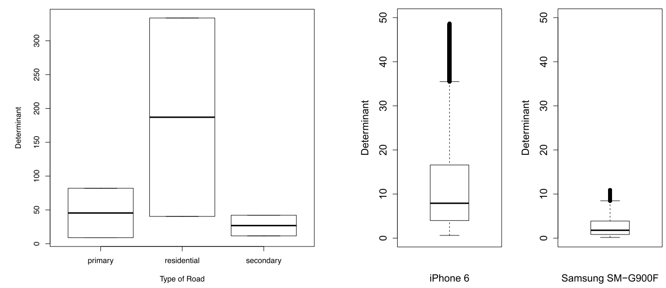
B. Collected Evaluation Dataset

Launched in early March 2015, the application collected, during the first four months, more than 150,000 individual trips covering 2,800,000 kilometers from 4,800 distinct drivers. Using our MVN model, we detected more than 600,000 risky maneuvers and computed around 280,000 model updates.

As a result, after 120 days of usage of the application, we were able to observe the evolution of the average score for the whole community of users. In Fig. 6, the average score calculated over the total number of users is shown for the first 120 days of the data collection campaign. We can observe that the average score improves roughly linearly over time. We attribute this score shift to users' behavior changes triggered by the gamified application and users' motivation to improve their scores (cf. sec.V-D).

C. Model Adaptation

As explained in Section IV-C, the MVN classifier is updated frequently in order to adapt to changes in the environment. One of the goals of the retrain phase is to let the MVN model adapt to different road conditions. As may be expected, lateral and longitudinal movements when driving in residential areas are different from those when driving on primary or secondary national roads. In order to illustrate the effect of different road conditions on the model, Fig. 7a shows the average determinant of the co-variance matrix of the current model, $det(\Sigma_r)$, depending on the road type. In order to obtain road type information, we looked up OpenStreetMap (OSM) road classes for every location that corresponds to each model update. We then associated the corresponding road class to the MVN model determinant. Fig. 7a shows that the determinant of the co-variance matrix of the MVN models computed on residential roads is around four times higher than that observed on primary and secondary roads, where driving is smoother, without extreme cornering maneuvers. Also, building a unique model on each enrolled device mitigates the effect of having different sampling rates and resolutions for motion sensors and GPS. Hence we can expect different sensing data quality from different device brands and models. This is observed in Fig. 7b, which shows the determinants of iPhone 6 and Samsung Galaxy S5 (SM-G900F) using box-plots. We observe much larger dispersion and higher median determinant for the Samsung Galaxy S5 compared to iPhone 6.



(a) (b)

Fig. 7. Model covariance determinant wrt. context and devices. (a) det vs. type of roads. (b) det for iPhone6 and GS5.

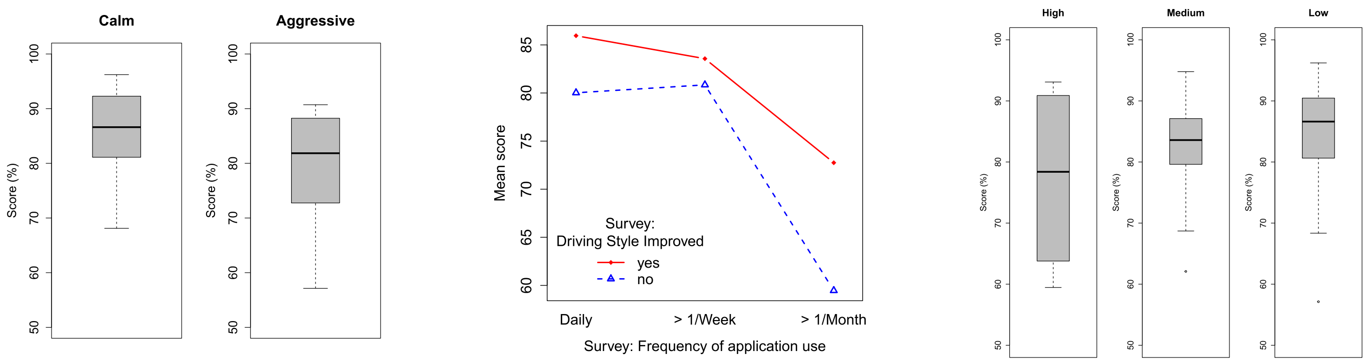
D. Score Reliability

1) Survey-Based Score Evaluation: In order to validate the reliability of the computed scores, we conducted an online survey of our community of users. It consisted of a set of questions designed to identify users' subjective opinions of their own driving. We obtained 52 sets of answers from active users and ran analysis of variance (ANOVA) tests on the results with respect to the participants' mean scores. The analysis yielded significant mean differences of scores between respondents on the question on subjective driving style. Fig. 8a shows the dispersion of score values for users who declare themselves in the survey to be either calm or aggressive drivers. The analysis of variance between these two groups rejects the null hypothesis, showing that there is evidence that the expected scores in both categories (i.e., calm and aggressive) differ. The p-value for the test was 0.0154.

Further, ANOVA indicated significant interaction between two survey questions: whether users think their driving behavior changed and how often they used the application. Fig. 8b shows the interaction between the answers to these two questions and the mean score of the groups. We can observe that users who declared an improvement in their driving and used the application frequently achieved the best mean scores, reflecting that the driver profiles were able to capture the improvement in driving style (which could also be observed earlier in Fig. 6). This also supports the idea that gamification can improve user behavior given frequent use of the application.

Additionally, we asked whether the planned introduction of speed radars was likely going to affect the respondent's driving. This question intends to differentiate between calm and aggressive drivers in terms of speeding: a driver declaring that the introduction of speed radars will affect his driving

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



(a) (b) (c)

Fig. 8. Score evaluation using survey results. (a) Subjective driving style estimation and corresponding score. (b) Interaction between application use frequency and subjective driving improvement w.r.t. scores. (c) Propensity for speeding and corresponding score.

TABLE II
ECO-DRIVING FACTORS AND THEIR CORRELATION TO SCORES

	Formula	ρ
Relative Positive Acceleration	$\frac{1}{d}\int(v\cdot a^{+})dt$	-0.553
Positive Kinetic Energy	$\frac{1}{d}\cdot\sum v_f^2 - v_i^2 \mid a>0$	-0.552
Deceleration Factor	$\overline{(a^{-})}$	0.491
Extreme Acceleration Factor	$\frac{1}{d}\cdot card(a>2.5\,m/s^2)$	-0.397

style implicitly accepts that he tends to occasionally exceed speed limits. In Fig. 8c we show the relation of the computed drive score against the propensity for exceeding speed limits. We observe that the computed drive score tends to increase for drivers with a low propensity for exceeding speed limits.

2) Eco-Driving Factors: Additionally, we have evaluated the score metric of drivers against known eco-driving factors that have been used to characterize driver behaviour and aggressiveness in studies [17], [18]. *Positive Kinetic Energy* represents the driver behavior during acceleration process, where v_i and v_f indicate the speed at the beginning and end of an acceleration phase. PKE is computed as the density over distance d of the sum of squared speed variation during acceleration phases. *Relative Positive Acceleration* models driver's power demand, producing high values with increasing agressiveness and fuel efficiency. The *Deceleration Factor* is simply the average deceleration value over a trip. The *Extreme Acceleration Factor* evaluates how often a driver surpasses a certain acceleration threshold (cardinality of samples with $a > 2.5m/s^2$) over a distance d.

For each trip in the study, we evaluated different eco-driving factors and computed their correlation ρ to the corresponding trip scores, shown in Table II. The score metric reflects these different driving factors with moderate to good correlation, and together with the results of the survey, this shows that it is a good indicator of driver behavior.

VI. CONCLUSION AND PERSPECTIVES

In this article, we have described an MVN model to detect risky driving maneuvers using smartphone sensors and GPS data. Rather than detecting maneuvers based on fixed thresholds or supervised learning methods requiring labelled driving data, the system allows the driver style to be dynamically fitted in a multivariate Gaussian model that is frequently updated in order to adapt to changing driving conditions. The main advantage of such a system is that the model is computed for each individual mobile device, vehicle and driver, avoiding any dependency on a priori training data. Through experimentation, we have analyzed the performance of the system in terms of maneuver detection and adaptability. We have also proposed a metric to measure the relative riskiness of observed maneuvers, and a strategy to avoid model degrading through selective updates. We have implemented and deployed our proposed model in a mobile application and collected driving traces from more than 4,800 users. The results confirm that the model adapts well to different road conditions and device types. Through an extensive survey, we confirm that the proposed scoring function accurately represents users' perceived driving styles, with possible applications in insurance telematics, car-pooling and ride-sharing.

In future work, we want to evaluate further parameter estimation methods to enhance our current model, in particular the robust Minimum Covariance Determinant and Minimum Volume Ellipsoid estimators. Furthermore, we plan to evaluate Multivariate Gaussian Mixture Models to account for the heteroscedasticity of the data by using different models depending on the road topology and corresponding velocities.

ACKNOWLEDGMENT

The authors would like to thank IEE Luxembourg, for providing the test car for the field experiment. The authors would also like to thank the staff of the Centre de Formation pour Conducteurs in Colmar-Berg, Luxembourg for giving them access to the training track.

REFERENCES

[1] Deloitte. (2014). *Overcoming Speed Bumps On the Road To Telematics*, accessed Oct. 2016.

[2] V. L. Neale, T. A. Dingus, S. G. Klauer, J. D. Sudweeks, and M. J. Goodman, "An overview of the 100-car naturalistic study and findings," Jun. 2005.

[3] G. Fitch, S. Lee, S. G. Klauer, J. Hankey, J. Sudweeks, and T. Dingus, "Analysis of lane-change crashes and near-crashes," Tech. Rep. DOT HS 811 147, 2009.

[4] F. Guo, S. G. Klauer, M. T. McGill, and T. A. Dingus, "Evaluating the relationship between near-crashes and crashes: Can near-crashes serve as a surrogate safety metric for crashes?" Tech. Rep. DOT HS 811, 2010, vol. 382.

[5] J. Engelbrecht, M. J. Booysen, G.-J. van Rooyen, and F. J. Bruwer, "Survey of smartphone-based sensing in vehicles for intelligent transportation system applications," *IET Intell. Transp. Syst.*, vol. 9, no. 10, pp. 924–935, Oct. 2015.

[6] J. Wahlström, I. Skog, and P. Händel, "Detection of dangerous cornering in GNSS-data-driven insurance telematics," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 3073–3083, Dec. 2015.

[7] J. Paefgen, F. Kehr, Y. Zhai, and F. Michahelles, "Driving behavior analysis with smartphones: Insights from a controlled field study," in *Proc. 11th Int. Conf. Mobile Ubiquitous Multimedia (MUM)*, New York, NY, USA, Dec. 2012, pp. 36:1–36:8.

[8] J. Almazán, L. M. Bergasa, J. Yebes, R. Barea, and R. Arroyo, "Full auto-calibration of a smartphone on board a vehicle using IMU and GPS embedded sensors," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2013, pp. 1374–1380.

[9] D. A. Johnson and M. M. Trivedi, "Driving style recognition using a smartphone as a sensor platform," in *Proc. 14th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2011, pp. 1609–1615.

[10] H. Eren, S. Makinist, E. Akin, and A. Yilmaz, "Estimating driving behavior by a smartphone," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2012, pp. 234–239.

[11] J. W. Joubert, D. de Beer, and N. de Koker, "Combining accelerometer data and contextual variables to evaluate the risk of driver behaviour," *Transp. Res. F, Traffic Psychol. Behaviour*, vol. 41, pp. 80–96, Aug. 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1369847816300791

[12] G. Castignani, T. Derrmann, R. Frank, and T. Engel, "Driver behavior profiling using smartphones: A low-cost platform for driver monitoring," *IEEE Intell. Transp. Syst. Mag.*, vol. 7, no. 1, pp. 91–102, 2015.

[13] C. C. Aggarwal, "Outlier analysis," in *Data Mining*. Springer, 2015, pp. 237–263.

[14] Y. L. Tong, *The Multivariate Normal Distribution*. Springer, 2012.

[15] P. C. Mahalanobis, "On the generalized distance in statistics," in *Proc. Nat. Inst. Sci. India*, vol. 2. 1936, pp. 49–55.

[16] Centre de Formation des Conducteurs (CFC). (2014). *Colmar Berg, Luxembourg*, accessed on Jul. 2014. [Online]. Available: http://www.cfc.lu/pages/english/about-us/the-centre.php

[17] C. Andrieu and G. S. Pierre, "Evaluation of ecodriving performances and teaching method: Comparing training and simple advice," *Proc. EJTIR*, vol. 14, no. 3, pp. 201–213, 2014.

[18] E. Ericsson, "Independent driving pattern factors and their influence on fuel-use and exhaust emission factors," *Transp. Res. D, Transp. Environ.*, vol. 6, no. 5, pp. 325–345, Sep. 2001.

German Castignani received the Computer Sciences Engineering degree from University of Buenos Aires, FIUBA, Argentina, and the Ph.D. degree in computer sciences from Institut Mines-Telecom, Telecom Bretagne, Rennes, France, under the direction of Prof. X. Lagrange and the supervision of Dr. N. Montavont. He is currently a Research Associate with the Interdisciplinary Center for Security, Reliability and Trust, University of Luxembourg. He is involved in several research projects at the VehicularLab, including handover management in the IEEE 802.11 V2I communications, mobile sensing, and crowdsourcing for driving assistance systems. He has authored over 15 scientific publications in conference proceedings and international peer-reviewed journals. His research interests include vehicular and wireless networks, mobility management, intelligent transportation systems, and advanced driving assistance systems.

Thierry Derrmann received the Computer Sciences Diploma (Dipl.-Inf.) degree from the Karlsruhe Institute of Technology, Germany, in 2011. He is currently working toward the Ph.D. degree with the Interdisciplinary Center for Security, Reliability and Trust, University of Luxembourg. During his Diploma thesis, he was involved in the design of filtering algorithms using GPUs for applications in the field of mass spectrometry. His Ph.D. project on multimodal mobility models is funded by the National Research Fund of Luxembourg. His research interests are related to data analysis, traffic optimization, and mobility applications and stochastic modeling.

Raphaël Frank received the master's degree in cryptography and network security from University Joseph Fourier, Grenoble, France, in 2006 and the Ph.D. degree in computer science from University of Luxembourg in 2010. He was a Visiting Scholar with University of California at Los Angeles, where he conducted research on data routing protocols for vehicular networks. He is currently a Research Scientist with the Interdisciplinary Center for Security, Reliability and Trust, University of Luxembourg, and the Head of the VehicularLab. He is currently involved in several European and national research projects. He has authored or co-authored over 20 scientific papers, published in peer-reviewed international journals and conferences. His research interests include vehicular networks and mobile computing. He is a member of the Car-2-Car Communication Consortium and the IPv6 Forum Luxembourg.

Thomas Engel received the degree and the Dr.rer.nat. degree in physics and computer science from University of Saarbruecken, Germany, in 1992 and 1996, respectively. Between 1996 and 2003, he was a member of the Board of Directors, a Joint Founder, and the Vice-Director of the Fraunhofer-guided Institute for Telematics e.V., Trier, Germany, where he was coresponsible for the scientific orientation and development of the institute. Since 2002, he has been a Professor with the IST/University of Luxembourg. Since 2009, he has also been the Deputy Director of SnT, the Interdisciplinary Center for Security Reliability and Trust, University of Luxembourg. He is currently a Professor in computer networks and telecommunications with University of Luxembourg. His SECAN-Lab Team deals with performance, privacy, and identity handling in next generation networks. As a member of the European Security Research Advisory Board, European Commission, Brussels, he advised the Commission on the structure, content, and implementation of the FP7 Security Research Program. He is a member of the Information and Communication Security Panel ICS of NATO and a Civil High-Level Expert for Electronic Communications (representing Europe) of NATO CEP/CCPC. He is Speaker of the Regional Group Trier/Luxemburg of the German Society for Computer Science. Since 2009, he has been the Chairman of the National IPv6 Council Luxembourg.