Need Data for Driver Behaviour Analysis? Presenting the Public UAH-DriveSet

Eduardo Romera, Luis M. Bergasa, Roberto Arroyo

***Abstract*— Driving analysis is a recent topic of interest due to the growing safety concerns in vehicles. However, the lack of publicly available driving data currently limits the progress on this field. Machine learning techniques could highly enhance research, but they rely on large amounts of data which are difficult and very costly to obtain through Naturalistic Driving Studies (NDSs), resulting in limited accessibility to the general research community. Additionally, the proliferation of smartphones has provided a cheap and easy-to-deploy platform for driver behavior sensing, but existing applications do not provide open access to their data. For these reasons, this paper presents the UAH-DriveSet, a public dataset that allows deep driving analysis by providing a large amount of data captured by our driving monitoring app DriveSafe. The application is run by 6 different drivers and vehicles, performing 3 different behaviors (normal, drowsy and aggressive) on two types of roads (motorway and secondary road), resulting in more than 500 minutes of naturalistic driving with its associated raw data and processed semantic information, together with the video recordings of the trips. This work also introduces a tool that helps to plot the data and display the trip videos simultaneously, in order to ease data analytics. The UAH-DriveSet is available at: http://www.robesafe.com/personal/eduardo.romera/uah-driveset**

I. Introduction

Driver behavior analysis is an emerging trend that suits the needs of multiple markets. The most traditional is automation, where detecting inattentive or aggressive driving behaviors is essential to improve safety in vehicles [1] or to switch control in semi-autonomous vehicles [2]. Another potential market is car insurance, which has been interested in monitoring driving activities in order to provide fair insurance premiums to its customers [3]. A third one is fleet management market, where logistics fleet administrators need to know how their vehicles are used and how their drivers behave in order to mitigate potential risks and reduce operational costs [4].

While driver profiling is an interesting research topic, the lack of available data currently limits the progress on this field. Machine learning techniques could highly enhance research, but they rely on large amounts of data, which can be obtained using three different methodologies [5]: 1) Driving style questionnaire: every driver evaluates their own driver behavior in self-reported scales. Thus, this approach reports few data and generates subjective measurements. 2) Driving simulators: this strategy allows an accurate control of the conditions of the experiment, thus facilitating the task of identifying cause-effect relations through lots of data. However, the artificial nature of the environment can lead to conclusions that are not easily translated into real world situations. 3) Real vehicles: the data captured in a real vehicle during typical day-to-day driving session provides the highest levels of accuracy, and it is the most objective of the three methods. This approach is known as Naturalistic Driving Study (NDS) in the recent literature [6]. Some representative examples are the 100-car study in [7] and the more recent Strategic Highway Research Program (SHRP2) [8]. Unfortunately, NDSs are very costly because instrumented vehicles are needed during long-term tests using multiple users. In addition, tasks involved with identifying and coding relevant epochs (semantic data) are very complicated and usually hand-made with little or no automatic means, which limits its accessibility to the general research community.

Fig. 1. DriveSafe and video-recorder on one of the testers windshield.

The proliferation of smartphones and mobile devices embedding different types of sensors has provided a cheap and easy-to-deploy platform for driver behavior sensing in a naturalistic way, offering a low-cost alternative to the instrumented vehicles, where precision losses are compensated with communication flexibility and crowdsource capabilities. A good survey of smartphone-based sensing for ITS applications can be found in [9]. However, none of them provides open access to their data, making performance comparison very difficult.

DriveSafe [10] [11] is a driver safety app for iPhones (launched by the authors in 2013) that infers drowsy and aggressive driving behaviors giving corresponding feedback to drivers and scoring their driving. In this paper, we take advantage of our app to present a new public dataset, named UAH-DriveSet, in order to push forward the performance of

*This work was supported by the University of Alcalá (UAH) through a FPI grant, the Spanish MINECO through the project SmartElderlyCar (TRA2015-70501-C2-1-R) and the Community of Madrid through the project RoboCity2030 III-CM project (S2013/MIT-2748).

The authors are with the Department of Electronics, UAH. Alcalá de Henares, Madrid, Spain. e-mail: {eduardo.romera, bergasa, roberto.arroyo}@depeca.uah.es

Fig. 2. DriveSafe running on a secondary road.

driver profiling studies using naturalistic data in a similar way that KITTI [12] carried out for the urban scene understanding topic or the PASCAL VOC [13] for the visual detection and segmentation one.

UAH-DriveSet provides a large amount of data obtained from 6 different drivers and vehicles, that simulated 3 different behaviors (normal, drowsy and aggressive) on two types of roads (motorway and secondary road), which results in more than 500 minutes of naturalistic driving with its associated raw and processed data, together with the video recordings of the trips. Processed data includes maneuvers recognition (acceleration, braking, turning, lane weaving, lane drifting, over-speeding, car following) and driving style estimation (normal, drowsy and aggressive), as the steps toward automating the process of extracting semantic information from the raw measurements, vital for data reduction in NDSs. This work also presents a tool that helps to review the data and recorded videos simultaneously, in order to ease data analysis and comparison with future proposals using this dataset.

II. Setup and Methodology

The UAH-DriveSet has been recorded by using the smartphone application DriveSafe, which uses all the available sensors on the smartphone (inertial sensors, GPS, camera and internet access) to log and recognize driving maneuvers and infer behaviors from them.

The tests were performed on the vehicles of the drivers by placing two phones on their windshield. Fig. 1 shows the setup reproduced on each tester. An iPhone with DriveSafe App is set on the center of the windshield, with the rear camera aiming at the road. The app has its own simple calibration stage at the start in order to set the phone perpendicular to the ground and align the vehicle with the inertial axes. A second phone is set closely on its right in order to record a video of the whole route. At the beginning of the designed routes, both the recorder and the DriveSafe App are started and the testers perform each full route without interfering with the phones.

The test bed is shown in Table I. It is composed by 6 different users of different ages and with different types of vehicles, including a fully electric car.

TABLE I
List of drivers and vehicles that performed the tests.

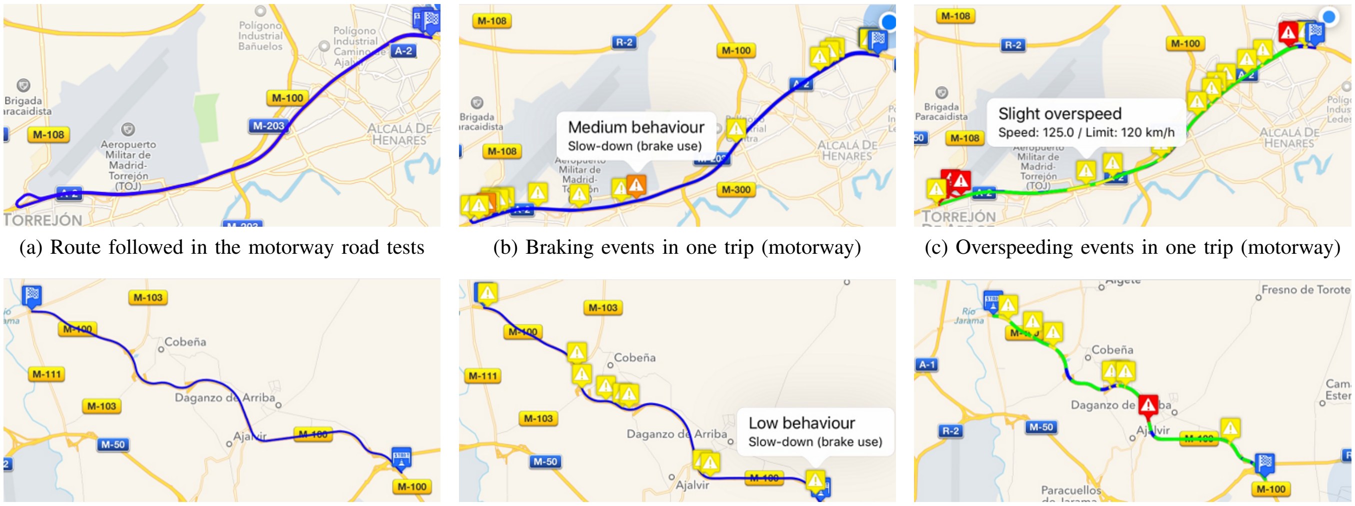
Driver	Genre	Age range	Vehicle Model	Fuel type
D1	Male	40-50	Audi Q5 (2014)	Diesel
D2	Male	20-30	Mercedes B180 (2013)	Diesel
D3	Male	20-30	Citröen C4 (2015)	Diesel
D4	Female	30-40	Kia Picanto (2004)	Gasoline
D5	Male	30-40	Opel Astra (2007)	Gasoline
D6	Male	40-50	Citröen C-Zero (2011)	Electric

Each driver repeats pre-designated routes by simulating a series of different behaviors: normal, drowsy and aggressive driving. In the case of normal driving, the tester is only told to drive as he usually does. In the drowsy case, the driver is told to simulate slight sleepiness, which normally results in sporadic unawareness of the road scene. Finally, in the case of aggressive driving, the driver is told to push to the limit his aggressiveness (without putting the vehicle at risk), which normally results in impatience and brusqueness while driving. These are the only indications that are given to perform the routes since the start. The co-pilot is in charge of the tests safety and he does not interfere by giving any additional instruction during the trips, except in cases of extreme risk during the maneuvers.

The two different routes covered in the tests are shown in Fig. 3. Both are roads from the Community of Madrid (Spain), close to the city of Alcalá de Henares. The first route is in its majority a “motorway” type of road, composed of between 2 and 4 lanes on each direction and around 120km/h of maximum allowed speed. The second route mostly covers a “secondary” type of road, composed of principally 1 lane on each direction and around 90km/h of maximum speed. Each driver performed three trips on the motorway road (round-trip, around 25km each), simulating each of the three behaviors, and four trips on the secondary road (one-way, around 16km each), which consist of: departure as normal, return as normal, departure as aggressive and return as drowsy. The electric vehicle (D6) performed all the motorway routes, but only one normal and the drowsy one in the case of the secondary road, due to problems related to lack of autonomy.

III. The UAH-DriveSet files

DriveSafe captures plenty of information of each route in the form of both raw measurements and processed signals (semantic information), such as the image captured by the rear camera. All this data has been gathered into files to create the UAH-DriveSet, which are described in this section. The dataset is split into folders for each of the drivers. Within these folders, each full route performed with a different behavior is stored in a subfolder with the following name format: “Date(YYYYMMDDhhmmss)-Distance(Km)-Driver-Behavior-Road”. These subfolders contain the video recorded during the route and 9 data files that are further described in the subsequent sections. These files contain different variables disposed on columns, where the first column is always a “timestamp” that represents the seconds since the start of the route, which allows to synchronize between the different files and the corresponding video.



(a) Route followed in the motorway road tests (b) Braking events in one trip (motorway) (c) Overspeeding events in one trip (motorway)

(d) Route followed in the secondary road tests (e) Braking events in one trip (secondary) (f) Overspeeding events in one trip (secondary)

Fig. 3. Some examples of the dataset routes reviewed in the map viewer of DriveSafe. Coloured tags in (b) and (e) indicate dangerous braking maneuvers, and (c) and (f) indicate dangerous overspeeding sections. Tag colour depends on the risk of the event (yellow equals low, orange medium and red high).

A. Raw real-time data

Two files, whose name starts by “RAW”, contain measurements obtained directly by the phone sensors. These are the inertial sensors (accelerometers and gyroscopes) and the GPS, aside from the camera images that have been made available on the videos. Both files are described below:

- **RAW_GPS** contains the data collected from GPS, at 1Hz. The contents of each column are listed below:
  1) Timestamp (seconds)
  2) Speed (km/h)
  3) Latitude coordinate (degrees)
  4) Longitude coordinate (degrees)
  5) Altitude (meters)
  6) Vertical accuracy (degrees)
  7) Horizontal accuracy (degrees)
  8) Course (degrees)
  9) Difcourse: course variation (degrees)
- **RAW_ACCELEROMETERS** contains all the data collected from the inertial sensors, at 10Hz (reduced from the phone 100Hz by taking the mean of every 10 samples). The iPhone is fixed on the windshield at the start of the route, so the axes are the same during the whole trip. These are aligned in the calibration process of DriveSafe, being Y aligned with the lateral axis of the vehicle (reflects turnings) and Z with the longitudinal axis (positive value reflects an acceleration, negative a braking). The accelerometers measurements are also logged filtered by a Kalman Filter (KF). The contents of each column are:
  1) Timestamp (seconds)
  2) Boolean of system activated (1 if >50km/h)
  3) Acceleration in X (Gs)
  4) Acceleration in Y (Gs)
  5) Acceleration in Z (Gs)
  6) Acceleration in X filtered by KF (Gs)
  7) Acceleration in Y filtered by KF (Gs)
  8) Acceleration in Z filtered by KF (Gs)
  9) Roll (degrees)
  10) Pitch (degrees)
  11) Yaw (degrees)

B. Processed data as continuous variables

Three files, whose name starts by “PROC”, contain variables that are processed in real-time by DriveSafe App in a continuous way. From the rear camera, DriveSafe processes maneuvers with respect to the driving lane and the ahead vehicles. From the internet connection, DriveSafe obtains and processes road information that is collected from road APIs such as OpenStreetMap (OSM). All this is gathered in three files:

- **PROC_LANE_DETECTION** contains relevant data processed from vision according to the road and ego-motion model described in [10], at around variable 30 Hz (FPS). The contents of each column are:
  1) Timestamp (seconds)
  2) X: car position relative to lane center (meters)
  3) Phi: car angle relative to lane curvature (degrees)
  4) W: road width (meters)
  5) State of the lane det. algorithm [-1=calibrating, 0=initializing, 1=undetected, 2=detected/running]
- **PROC_VEHICLE_DETECTION** contains important data processed from the vehicle detector module (described in [11]), at around variable 10 Hz (FPS). The contents of each column are listed below:
  1) Timestamp (seconds)
  2) Distance to ahead vehicle in current lane (meters) [value -1 means no car is detected in front]
  3) Time of impact to ahead vehicle (seconds) [distance related to own speed]
  4) Number of detected vehicles in this frame (traffic)
  5) GPS speed (km/h) [same as in RAW_GPS]

- **PROC_OPENSTREETMAP_DATA** contains important data processed by accessing internet to a map information API such as OSM. Frequency depends on the server response, around 1 Hz. The contents of each column are:
  1) Timestamp (seconds)
  2) Maximum allowed speed of current road (km/h)
  3) Reliability of obtained maxspeed (0=unknown, 1=reliable, 2=used previously obtained maxspeed, 3=estimated by type of road)
  4) Type of road (motorway, trunk, secondary...)
  5) Number of lanes in current road
  6) Estimated current lane (1=right lane, 2=first left lane, 3=second left lane, etc) [experimental]
  7) GPS Latitude used to query OSM (degrees)
  8) GPS Longitude used to query OSM (degrees)
  9) OSM delay to answer query (seconds)
  10) GPS speed (km/h) [same as in RAW_GPS]

C. Processed data as events

DriveSafe also detects individual events produced during the driving. From the accelerometers, the app detects sudden accelerations, brakings and turning events. From the lane detection, the app detects the lane changes. These are called "EVENTS" instead of "PROC" as they are not continuously saved but only stored when each event is produced. Both files are described below.

- **EVENTS_LIST_LANE_CHANGES** contains the list of lane changes detected during the route. A lane change is assumed as irregular when it is performed too fast or too slowly. The contents of each column are:
  1) Timestamp (seconds)
  2) Type [+ indicates right and - left, 1 indicates normal lane change and 2 slow lane change]
  3) GPS Latitude of the event (degrees)
  4) GPS Longitude of the event (degrees)
  5) Duration of the lane change (seconds) [measured since the car position is near the lane marks]
  6) Time threshold to consider irregular change (secs.) [slow if change duration is over this threshold and fast if duration is lower than threshold/3]
- **EVENTS_INERTIAL** contains a list of the inertial events detected during the route: brakings, turnings and accelerations. We detect 3 different levels for each one, according to the thresholds described in [10]. However, if the speed is less than 50Km/h (see boolean of system activated in RAW_ACCELEROMETERS), the events are not saved in this list. Its columns are:
  1) Timestamp (seconds)
  2) Type (1=braking, 2=turning, 3=acceleration)
  3) Level (1=low, 2=medium, 3=high)
  4) GPS Latitude of the event
  5) GPS Longitude of the event
  6) Date of the event in YYYYMMDDhhmmss format

D. Semantic information

From all the mentioned variables, DriveSafe evaluates a series of maneuvers in a mid-level semantic step and scores them according to a general behavior pattern obtained in a heuristic way (see Section IV and Table II). Driver behavior is rated among 3 classes (normal, drowsy and aggressive) by using this semantic data jointly with additional road, user and traffic information. All these results are logged in two files, which contain a wide number of variables that cannot fit in this document due to space constraints, so their detailed structure is made available in a *readme* file within the dataset:

- **SEMANTIC_FINAL**: contains maneuver set and their final scores, behavior ratios and other relevant info.
- **SEMANTIC_ONLINE**: contains the real-time sequence of maneuvers and behavior scores estimated during the route, as shown to the user in DriveSafe.

E. DriveSet reader

As there are several variables and files for every route and syncing them with the recorded video may suppose difficulties, a reader tool has been made available with the dataset. This tool allows to select each of the routes in order to simultaneously reproduce the associated video and plot a selection of variables synced in real-time within an user interface (see Fig. 4). This tool can be used to find patterns in the driving behaviors by reviewing all the variables available in the dataset together with the videos that show what did actually happen during the tests. For example, the user may analyze how is a car-following maneuver in an aggressive behavior by reviewing the real-time plot of the distance that the drivers keep with respect to the ahead vehicle.

IV. Results

DriveSafe does not only log data, but also performs real-time maneuver detection, scoring and behavior analysis. Although the description of all the developed algorithms is out of the scope of this paper, the processed maneuver scores and estimated behavior ratios are presented in this section in order to allow the comparison for possible future users of the dataset. Table II contains relevant semantic data obtained for each of the routes and drivers of the dataset. Each entry has some short info of the route, like the exact duration in minutes and kilometers and the average and maximum speed obtained in the route. The "Maneuver scores" part contains the scores given by DriveSafe App for each of the analyzed maneuvers: Accelerations, Brakings, Turnings, Lane-Weaving, Lane-Drifting, Overspeeding and Carfollowing, where the minimum of them is marked in bold letters. The "Behavior" part contains the ratios estimated by DriveSafe for the 3 evaluated classes (normal, drowsy or aggressive). The predominant one is marked in bold letters. All the scores and ratios are in base 10.

The scores for maneuvers detected from inertial events (Accelerations, Brakings and Turnings) normally depend on each driver profile and the road conditions (e.g. traffic density) instead of depending on a specific behavior. However, aggressive driving may result in brusqueness and this is

TABLE II

COMPLETE LIST OF DRIVESAFE SCORES PER SIMULATED BEHAVIOUR, PERFORMED ROUTE AND DRIVER.

State	Driver	Duration		Speed (Km/h)		Maneuver scores							Behavior		
		Time	Km	Avg	Max	Acc	Bra	Tur	Weav	Drift	Overs	Carfoll	Nor	Drow	Agg
Normal (Motorway)	D1	14m.	25	107	131	10	9.7	8.7	9.3	**7.9**	9.4	9.8	**6.8**	1.4	1.8
	D2	15m.	26	98	127	9.9	9.9	**7.2**	10	7.5	9.6	9.3	**6.8**	1.5	1.7
	D3	15m.	26	101	122	10	9.9	9.4	9.4	**8.1**	9.7	9.8	**7.3**	1.3	1.4
	D4	16m.	25	91	120	9.9	9.9	9.7	10	**8.9**	9.9	9.9	**8.2**	0.6	1.2
	D5	15m.	25	99	120	9.0	9.4	**7.8**	10	8.0	9.3	9.1	**6.8**	1.2	2.0
	D6	17m.	25	89	104	9.7	9.7	**3.5**	10	8.7	9.8	9.7	**8.0**	0.8	1.2
Drowsy (Motorway)	D1	15m.	25	97	113	10	3.8	6.9	**2.6**	4.3	9.7	9.7	3.2	**5.6**	1.2
	D2	15m.	25	98	122	9.4	4.8	7.8	5.2	**4.7**	9.7	9.4	**4.2**	4.1	1.6
	D3	16m.	26	91	129	9.8	10	7.9	**1.5**	5.2	9.7	9.9	2.6	**6.0**	1.4
	D4	17m.	25	88	106	9.9	9.8	8.7	**4.1**	4.6	9.0	9.9	3.8	**4.6**	1.6
	D5	18m.	25	83	96	8.6	4.2	8.2	**0.9**	3.1	9.5	9.9	1.8	**6.8**	1.3
	D6	17m.	25	84	99	9.6	9.2	**1.8**	3.9	4.8	7.1	9.9	2.5	**4.7**	2.8
Aggressive (Motorway)	D1	12m.	24	120	148	10	7.0	8.1	10	8.5	**6.1**	9.1	**5.1**	0.9	4.0
	D2	14m.	26	107	147	6.6	5.9	6.6	9.2	5.7	6.7	**2.1**	1.2	2.7	**6.1**
	D3	13m.	26	110	146	9.1	**0.0**	9.4	10	8.0	6.9	6.5	**5.4**	1.2	3.4
	D4	15m.	25	97	130	6.8	**2.7**	8.5	9.0	8.6	8.3	3.3	3.7	1.0	**5.3**
	D5	13m.	25	114	147	7.8	2.4	1.3	10	7.7	6.1	**0.3**	1.3	1.4	**7.3**
	D6	15m.	25	101	127	6.4	5.3	**0.0**	10	8.9	8.4	4.4	**4.8**	0.6	4.6
Normal[1] (Secondary)	D1	10m.	16	96	116	10	10	8.7	10	**6.3**	7.3	9.8	**6.4**	1.5	2.1
	D2	10m.	16	91	103	9.9	10	10	10	**7.4**	7.8	9.9	**6.2**	1.5	2.3
	D3	11m.	16	85	97	9.9	10	10	10	**6.9**	9.6	9.8	**6.9**	1.9	1.2
	D4	11m.	16	82	101	10	10	9.5	10	**8.8**	9.6	10	**9.1**	0.7	0.2
	D5	11m.	16	84	102	9.4	9.9	9.5	10	**7.3**	9.4	8.9	**7.6**	1.6	0.8
	D6	13m.	16	75	90	9.9	9.7	**4.5**	10	9.2	9.9	10	**9.5**	0.4	0.0
Drowsy (Secondary)	D1	8m.	13	94	107	10	4.9	6.6	10	**2.8**	7.7	10	3.3	**4.3**	2.4
	D2	10m.	16	91	110	8.8	3.8	8.1	**0.0**	4.1	8.5	9.6	0.9	**7.2**	1.9
	D3	10m.	17	91	118	10	9.4	9.5	**0.0**	4.0	8.1	9.9	0.7	**7.2**	2.1
	D4	11m.	17	87	102	9.9	9.1	8.1	**2.0**	3.9	9.4	9.9	1.8	**6.0**	2.2
	D5	11m.	16	84	100	10	9.7	4.8	10	**1.4**	9.8	9.2	3.4	**5.1**	1.5
	D6	12m.	16	80	94	8.7	8.8	2.5	**0.0**	4.6	9.9	10	1.4	**7.1**	1.5
Aggressive (Secondary)	D1	8m.	16	112	132	10	2.9	5.7	10	5.9	**0.0**	9.5	0.5	2.4	**7.1**
	D2	10m.	16	96	119	7.2	3.7	10	10	5.8	**0.2**	0.7	0.0	2.5	**8.7**
	D3	11m.	16	87	119	8.4	8.2	8.6	10	6.4	7.3	**1.5**	1.5	2.1	**6.4**
	D4	10m.	16	89	113	6.8	8.0	10	10	6.9	8.0	**2.3**	1.8	1.9	**6.3**
	D5	7m.	12	100	147	9.0	0.1	6.2	10	5.0	**0.0**	4.6	0.0	3.0	**8.0**

reflected in the inertial scores, which are lower in general for all drivers on both aggressive routes (motorway and secondary). Smaller vehicles (like D6 electric small car) are also usually more brusque on turnings and brakings, as demonstrated in [14]. The Lane-weaving maneuver analysis scores the irregularities in switching between lanes, which can be produced when the driver is momentarily not aware of the road (slow change) or when the driver is being brusque (fast change). The Lane-drifting evaluates the capacity of the driver to continue centered on its own lane. Swinging around the sides of the lane instead of keeping a straight way also reflects unawareness of the road. Both scores are generally lower on the drowsy routes. Overspeeding evaluates the capacity of driving under the legally allowed speed. This score degrades depending on how much and how long the driver surpassed the allowed maximum speed. Car-following evaluates if the driver keeps a safety distance to ahead vehicle on its own lane. This score highly degrades if the driver performs dangerous actions like tailgating (i.e. keeping too close to ahead vehicle). Both last scores are generally lower on the aggressive routes.

The behavior ratios are inferred from the mid-level semantic maneuvers and other variables. Normal state is the lack of the other two states, so it is high if no drowsiness or aggressiveness is detected, and 0 if the sum of these two surpasses 10. The "Drowsy" ratio reflects the sleepiness of the driver during the route, and the "Aggressive" ratio reflects the aggressiveness of the driver. The results presented in Table II show that DriveSafe correctly detects the predominance of each behavior. On the case of secondary road, it detected the behavior correctly on all routes (100%). On the motorway case, it detected correctly the behavior predominance for 78% of the trips, and in the rest it was very close to achieving it. For instance, on drowsy-motorway for D2 it gave 4.2 to normal and 4.1 to drowsy, the correct one. For the aggressive case on motorway, on the trips that it incorrectly labeled as "Normal", the score for aggressive was also high (e.g. D6 with 4.8 normal vs 4.6 aggressive). The errors on these aggressive cases are mostly due to the fact that the concept of aggressiveness is very subjective. The only indication given to the drivers was "to try to simulate an aggressive behavior". Therefore, while some drivers had no problem in performing dangerous aggressive maneuvers, others limited a bit their aggressiveness due to the high associated risk, mostly due to high traffic (on the motorway case). Moreover, some drivers reflect aggressiveness only on one indicator such as overspeeding, without being brusque on the inertial movements for instance, which explains the

[1] Only one of the normal-secondary routes (departure) has been displayed in the table due to lack of space, but both are available in the dataset.

Fig. 4. Screenshot of the UAH-DriveSet reader, a publicly available tool to perform analysis on the dataset by reviewing variables and videos simultaneously.

high differences in the aggressive scores between drivers.

V. Conclusion and Future Work

In this paper, we have presented the UAH-DriveSet, a publicly available set of driving data that has been recorded with our smartphone app DriveSafe by various drivers in different environments and behaviors. With machine learning in mind, this dataset contributes a large amount of public data to facilitate future research possibilities in the field of driving analysis. It contains more than 500 minutes of naturalistic driving tests in which DriveSafe has logged and processed several types of variables that have been made available together with video recordings of each route. Additionally, semantic information obtained from our driving analysis has been provided in order to allow future comparison of analysis techniques. We also contribute a tool to display the trip videos while plotting the variables within an user interface, which facilitates the task of analyzing patterns.

The dataset will allow future works in driving analysis like the work presented in [14], and future enhancements in the techniques applied by DriveSafe app to score and analyze the drivers. These algorithms will be made available in additional works in the near future, supposing our complete proposal for driving analysis. Additionally, future research could involve performing more tests to expand the UAH-DriveSet with more vehicles, drivers and road environments.

References

[1] L. M. Bergasa, J. Nuevo, M. Sotelo, R. Barea, and M. E. Lopez, “Real-time system for monitoring driver vigilance,” *IEEE Trans. on ITS*, vol. 7, no. 1, pp. 63–77, 2006.

[2] R. Vasudevan, V. Shia, Y. Gao, R. Cervera, R. Bajcsy, and F. Borrelli, “Safe semi-autonomous control with enhanced driver modeling,” in *IEEE American Control Conf. (ACC)*, 2012, pp. 2896–2903.

[3] C. Troncoso, G. Danezis, E. Kosta, J. Balasch, and B. Preneel, “Pripayd: Privacy-friendly pay-as-you-drive insurance,” *IEEE Trans. on Dependable and Secure Comp.*, vol. 8, no. 5, pp. 742–755, 2011.

[4] G. Castignani, T. Derrmann, R. Frank, and T. Engel, “Driver behavior profiling using smartphones: A low-cost platform for driver monitoring,” *IEEE ITS Magazine*, vol. 7, no. 1, pp. 91–102, 2015.

[5] A. B. Rodriguez Gonzalez, M. R. Wilby, J. J. Vinagre, and C. Sanchez, “Modeling and detecting aggressiveness from driving signals,” *IEEE Trans. on ITS*, vol. 15, no. 4, pp. 1419–1428, 2014.

[6] R. K. Satzoda and M. M. Trivedi, “Drive analysis using vehicle dynamics and vision-based lane semantics,” *IEEE Trans. on ITS*, vol. 16, no. 1, pp. 9–18, 2015.

[7] V. L. Neale, T. A. Dingus, S. G. Klauer, J. Sudweeks, and M. Goodman, “An overview of the 100-car naturalistic study and findings,” *NHTSA Technical Report*, no. 05-0400, 2005.

[8] G. A. Davis and J. Hourdos, *Development of analysis methods using recent data*. Transportation Research Board, 2012.

[9] J. Engelbrecht, M. J. Booysen, G.-J. Rooyen, and F. J. Bruwer, “Survey of smartphone-based sensing in vehicles for intelligent transportation system applications,” *IET ITS*, vol. 9, no. 10, pp. 924–935, 2015.

[10] L. M. Bergasa, D. Almería, J. Almazán, J. J. Yebes, and R. Arroyo, “DriveSafe: an app for alerting inattentive drivers and scoring driving behaviors,” in *IEEE Intel. Vehicles Symp. (IV)*, 2014, pp. 240–245.

[11] E. Romera, L. M. Bergasa, and R. Arroyo, “A real-time multi-scale vehicle detection and tracking approach for smartphones,” in *IEEE Intel.t Transportation Systems Conf. (ITSC)*, 2015, pp. 1298–1303.

[12] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the kitti vision benchmark suite,” in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 3354–3361.

[13] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, “The Pascal visual object classes (VOC) challenge,” *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.

[14] C. Arroyo, L. Bergasa, and E. Romera, “Adaptive fuzzy classifier to detect driving events from the inertial sensors of a smartphone,” in *IEEE Int. Conf. on Intelligent Transportation Systems (ITSC), accepted for presentation*, 2016.