The PREVENTION dataset:
a novel benchmark for PREdiction of VEhicles iNTentIONs

R. Izquierdo, A. Quintanar, I. Parra, D. Fernández-Llorca, and M. A. Sotelo

Abstract— Recent advances in autonomous driving have shown the importance of endowing self-driving cars with the ability of predicting the intentions and future trajectories of other traffic participants. In this paper, we introduce the PREVENTION dataset, which provides a large number of accurate and detailed annotations of vehicles trajectories, categories, lanes, and events, including cut-in, cut-out, left/right lane changes, and hazardous maneuvers. Data is collected from 6 sensors of different nature (LiDAR, radar, and cameras), providing both redundancy and complementarity, using an instrumented vehicle driven under naturalistic conditions. The dataset contains 356 minutes, corresponding to 540 km of distance traveled, including more than 4M detections, and more than 3K trajectories. Each vehicle is unequivocally identified with a unique id and the corresponding image, LiDAR and radar coordinates. No other public dataset provides such a rich amount of data on different road scenarios and critical situations and such a long-range coverage around the ego-vehicle (up to 80 m) using a redundant sensor set-up and providing enhanced lane-change annotations of surrounding vehicles. The dataset is ready to develop learning and inference algorithms for predicting vehicles intentions and future trajectories, including inter-vehicle interactions.

Fig. 1.   PREVENTION benchmark outline.

I. Introduction and Related Work

Self-driving cars have experienced a booming development in the past years, both in terms of driving complexity and safety. However, recent research works have demonstrated that autonomous driving systems require advanced prediction of intentions of other traffic agents, such as cars and pedestrians, in order to make further progress in the field. Thus, self-driving cars need to be able to predict potential hazards that involve a deeper understanding of the complex driving behaviors of other human-driven cars, including inter-vehicle interactions. By doing so, the level of safety and autonomy of self-driving cars will be improved, making SAE level 4 a reachable objective for car makers. In this context, the need of data from naturalistic driving scenarios can be considered as one of the most relevant requirements for identifying, modeling and understanding driver behaviors [1] as well as for accelerating the verification and validation of automated vehicles [2]. A considerable effort has been made during the last decade to collect data from equipped vehicles driven under naturalistic conditions, covering different driving tasks, such as car following, lane change, lane departure, cut-in maneuvers, etc., and using different in-vehicle sensors, such as cameras, 2D laser scanners, radars, CAN-Bus signals, GPS devices, etc. [1]. Predicting
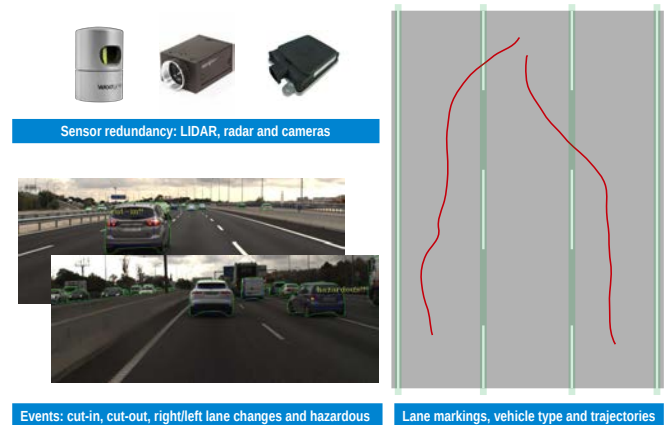
R. Izquierdo, A. Quintanar, I. Parra, D. Fernández-Llorca and M. A. Sotelo are with the Computer Engineering Department, Universidad de Alcalá, Alcalá de Henares, Spain ruben.izquierdo@uah.es

the intentions of vehicles is of paramount importance in highways and urban environments when it comes to getting the most out of autonomous vehicles in such a way that they can mimic human driving in complex traffic situations. For example, an autonomous vehicle trying to enter a congested highway might never find the gap that, according to its programming, is needed to merge the traffic and safely enter the lane unless an advanced system for prediction of intentions is implemented. Such systems can endow autonomous cars with the capability to imitate human driving and find their way to enter the highway by means of prediction and incremental interaction, very much in the way human drivers do. Making such predictions in an accurate way is an extremely complex task, given that thousands of real examples are needed, showing as many different situations as possible, in order to learn from experience what the most likely vehicle maneuvers and reactions are in different critical circumstances. For such purpose, current datasets do not offer solid material to researchers, given that they are mostly oriented to recognition and segmentation learning tasks, but not to prediction. Some limited intention prediction systems have been developed with the little available data focused on ego-motion prediction [3], [4], [5] or based on an external point of view [6], [7]. However, some recently released datasets contain dense traffic data aiming at supporting the development of intelligent systems for vehicle detection and tracking. Thus, the NGSIM HW101 dataset [8] contains 45 minutes of images recorded from a building using 8 synchronized cameras at 10 Hz covering a stretch of 5-lane highway 640 meters long. Similarly, the NGSIM I80 dataset [9] provides 45 minutes of images recorded from a building

using 7 synchronized cameras at 10 Hz covering a stretch of 6-lane highway 500 meters long. The HighD dataset [10] provides aerial images obtained with a drone covering 420 meters of a straight stretch of highway. This dataset contains 16.5 hours of recorded data in 4 different locations on 2-lane and 3-lane highways. However, due to the intrinsic limitations of cameras for obtaining accurate range measurements at long distance, these approaches are somehow limited to frameworks where accuracy is not critical (e.g. traffic flow models). The development of accurate trajectory prediction systems for autonomous or assisted driving requires on-road data collection in order to gather as much accurate data as possible. In this line, the PKU dataset [11] was released in 2017 by Peking University and the PSA Group. It contains 170 minutes of publicly available data gathered along the 4th ring road around the city of Beijing, using a vehicle equipped with 4 2D-LiDARs covering a region of 40 meters around the vehicle. The PKU dataset does not contain information regarding the road lane markings, the number of road lanes, or the relative positioning of the ego-vehicle. More recently, the ApolloScape dataset [12] has been released in 2018 by Baidu Research, containing data obtained in urban environments from 4 cameras (2 forward and 2 rear) and 2 Laser scanners using a vehicle driving at 30 km/h. ApolloScape is currently one of the most complete datasets in the state-of-the-art given that it provides 2D and 3D information of all detected objects. However, it does not contain radar data, making detections more sensitive to failure in adverse weather conditions and highway scenarios (apart from that, the benefit of redundant operation by analyzing the data from LiDAR, vision, and radar is not possible). In addition, it does not provide labeled tracking information (IDs and tracklets) for all detected objects. This information is necessary to evaluate the future trajectories (predictions) of vehicles and Vulnerable Road Users (VRUs). No critical traffic situations (from the prediction standpoint) are specially labeled or highlighted in any of the previously mentioned datasets.

In conclusion, it becomes clear that a new, specialized dataset is needed for vehicle prediction purpose. Such a dataset must contain thousands of examples obtained in real driving conditions in different environments (highway and urban areas). In this paper, we present a new dataset -PREVENTION- for the prediction of vehicles intentions, featuring the following characteristics:

- Data from 6 sensors of different nature (laser, radar, and vision) are provided, contributing to redundancy and fault-tolerant development. Measurements from the 6 sensors are time-synced and cross-calibrated [13] [14]. The vehicle used to create the dataset is equipped with 3 radars (1 narrow field of view long-range radar and 2 broad field of view radars), 1 Velodyne HDL-32E, and 2 high-resolution cameras, forward and rear-looking, respectively, providing high frame rate data. Readings obtained from these 6 sensors are combined with data coming from GPS, IMU, and CAN bus.

- Positions of all vehicles around the ego-vehicle are accurately labeled in a semi-automatic process [13] (including parked vehicles) and made available together with their respective vehicle IDs. Fusion between the appropriate sensors is carried out in order to obtain as much an accurate positioning as possible, both in lateral and longitudinal dimensions.

- Surrounding data are provided in a range of at least 80 meters around the ego-vehicle (up to 200 meters in the frontal area). This allows for developing a safety area around the ego-vehicle in which all vehicles entering or leaving such area are carefully located and tracked in order to accurately predict their most likely trajectories.

- Road lane markings are included in the dataset, providing the relative positioning of all vehicles on the road (lateral positioning and orientation), including the ego-vehicle, as well as the number and type of road lanes present on the road. This information is essential for enhancing road scene understanding and for providing contextual framing.

- The PREVENTION dataset also offers the possibility of automatically extracting selected data corresponding to certain types of critical maneuvers that are of special interest for prediction purpose. Such is the case of overtaking or lane change maneuvers. For example, users can extract data sequences in which a fast vehicle driving on the right-most lane makes a change to the center lane when approaching a slow vehicle on the same lane, causing a cut-in on the ego-vehicle trajectory (assuming that the ego-vehicle is driving on the center lane). Another critical situation is that of a vehicle aggressively entering the highway from a ramp lane, cutting-in the trajectory of the ego-vehicle. These maneuvers are extremely interesting for prediction purpose given that they cause potentially dangerous situations that can be predicted based on contextual information and experience. By extracting selected sequences containing critical maneuvers, such as lane changes and overtaking, learning can concentrate on such scenarios leading to more advanced prediction systems that learn critical situations from real experience.

To date, no other public dataset in the state of the art provides such a rich amount of data on different road scenarios and critical situations and such a long-range coverage around the ego-vehicle using a redundant sensor set-up.

The rest of the paper is organized as follows. Section II provides a thorough description of the methodology followed to create the dataset. Section III discusses the different pros and cons in detail. Finally, section IV concludes the paper and provides some insights into the future developments.

II. METHODOLOGY

In this section the methodology employed to build the dataset is detailed. First, the sensors used to perceive the environment and the ego-state are enumerated. Secondly, spatial calibration and time synchronization procedures are explained. Then, the driving environment and the driving

style are described, followed by the data logging process and the information data labeling. Finally, web access and the data format description are provided.

A. Sensors Setup

The image acquisition system consists of two Grasshopper3 cameras mounting 12.5 mm fixed focal length lens. The cameras cover a Field Of View (FOV) of 48° in the front and the back. The sensor is a SONY CMOS Bayer array with WUXGA (1920×1200) resolution that can be triggered up to 163Hz.

A Velodyne HDL-32E generates point clouds at a constant rate of 10 Hz. Each cloud is defined by an array of 3D points with 32 vertical and more than 2000 horizontal samples with all around coverage and +10° to -30° vertical FOV. The detection range of the LiDAR is up to 100 m with an error lower than 2 cm.

Three radars complete the perception system. A Continental ARS308 long-range radar is located centered in the front bumper with a detection range up to 200 m and a FOV up to 56°. Real-time scanning of tracked objects are provided at 15 Hz. Two Continental SRR208 blind-corner radars are installed in both corners of the front bumper with a detection range up to 50 m and a FOV up to 150°. Tracked objects information is provided at a rate of approximately 33 Hz.

The localization task is performed by a Differential Global Navigation Satellite System (DGNSS) Trimble Net R9 Geospatial with Real Time Kinematic (RTK) capability. Geographical coordinates are generated at 20 Hz with differential corrections received through the 3G/4G network and a Bluetooth connection.

The Controller Area Network (CAN) bus of the vehicle is constantly monitored and many variables such steering position, braking pressure, throttle position, speed, acceleration, gear, etc. are available and logged.

Finally, an Inertial Measurement Unit (IMU) MPU6050 complements the localization task. This low-grade IMU in combination with the CAN bus and the GNSS data enables better localization and ego-state estimations by means of an Extended Kalman Filter (EKF) [15] and a dynamic vehicle model [16].

B. Sensors time-sync and cross-calibration

The perception system is composed of sensors of three different nature: vision, LiDAR, and radar. The spatio-temporal relationships between measures are critical when sensor fusion techniques are applied. Focused on this, time synchronization mechanisms and sensor calibration procedures are deployed.

Some sensors produce a non-controllable data output such a radar, LiDAR, IMU, or CAN bus. Others, like cameras, are actively triggered and the data output is known and expected. Two different approaches have been used as a time synchronization mechanism to cover both kinds of data streams. The clock of the recording computers is synchronized in a common time reference by means of a GPS Pulse Per Second (PPS) signal and a Network Time Protocol (NTP)
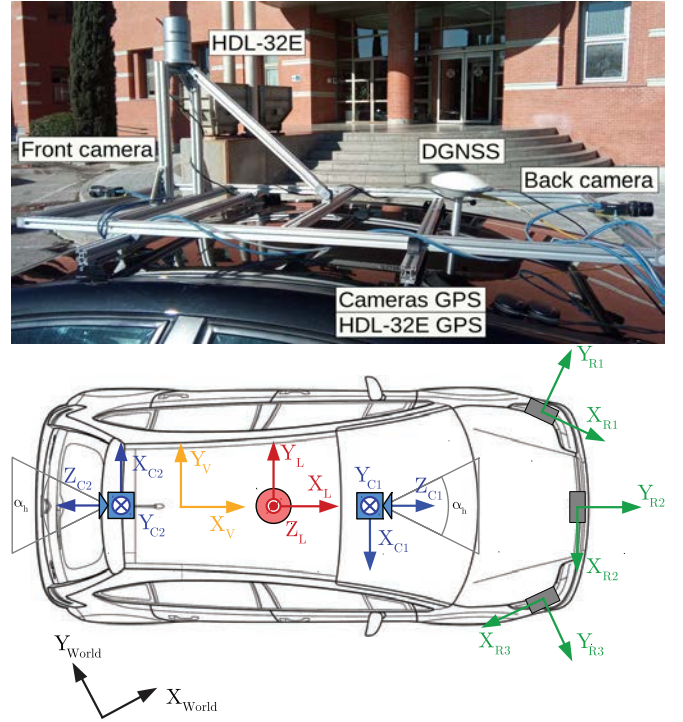

Fig. 2. Sensor's reference systems in the vehicle frame. All the reference systems are Cartesian right-handed systems.

server. Thus, different recording computers are capable of adding a common time stamp to data coming from different sensors at different locations.

Cameras must be externally triggered, and consequently, the data output is actively generated. Cameras are individually triggered when the LiDAR points in the same direction where each camera is pointing to. This guarantees minimum point cloud distortion in the area covered by the cameras even when the point cloud is not corrected with the ego-vehicle movement. A dedicated computer develops the triggering task and stores two different timestamp values: the time when the triggered signal is sent from the computer and the time when the acknowledgment is received from the camera. Both values are within the DGNSS time reference system.

Spatial calibration between sensors has been carried out in order to enable sensor fusion capabilities. The ego-vehicle reference system is located in the middle of the vehicle over the rear axis, where the DGNSS is placed. The x-axis and y-axis matches with ego-vehicle's forward and left movement directions, consequently, z-axis points up according to a Cartesian right-handed system. Three groups of sensors: radar, camera, and LiDAR have been calibrated to transform all the available information coming from the environment into a common reference system. Fig. 2 shows all the reference systems defined in the vehicle.

Cameras are intrinsically calibrated and extrinsically w.r.t the LiDAR according to the procedure described in [13]. This endows the system with the ability of generating 3D information mixing camera and LiDAR detections. The high

resolution of the camera is combined with the accurate range measures of the LiDAR to achieve precise detections. Radars are extrinsically calibrated w.r.t the vehicle reference system with a procedure based on a Digital Map with known high-sensitive radar elements such as traffic signals and light poles [14]. Extrinsic calibration between LiDAR and vehicle reference system is defined by a constant transformation matrix composed with a translation vector. The LiDAR is mounted in a fixed position over a structure built in parallel to the vehicle axes.

C. Driving environments and Driving Style

The PREVENTION dataset contains both urban and highway scenarios, but it is mainly oriented to predicting intentions and trajectories on highway environments. Three different people drove the car to generate the data. Drivers were instructed to arrive at the destination following the traffic rules. Drivers used the cruise control at their will. A total length of 6 hours and more than 500 km were recorded in 5 different days and 3 different areas. Table I summarizes the dataset details. The A2 and A3 are both 3 lane highway areas with straight stretches mostly. The M-50 and M-40 are the outer and middle rings around the city of Madrid respectively, which have three or more lanes. The recordings were made during the central hours of the day to avoid rough traffic, however, traffic jams and congested traffic can be found in the dataset. Moreover, the sun position is optimal for image acquisition at these hours. Some of the records cover the same driven areas. By doing so, different behaviors and interactions can be observed at the same location with different points of view. This dataset is not created for localization algorithms, however, these algorithms can be developed and tested with multiple records of the same areas.

TABLE I

DATASET MAIN FEATURES

Record #	Area	Date	Length	Distance
1	A2	21th Jun	18 min	47 km
2	A2, M50, A3	19th Jul	59 min	83 km
3	A2, M50, A3	24th Jul	57 min	86 km
4	A2, M-40	18th Oct	108 min	149 km
5	A2, M-40	22th Nov	114 min	175 km
Total	–	–	**356 min**	**540 km**

D. Data-logging

Data-logging is carried out by three different computers. The main computer is the control computer of the vehicle, which is in charge of reading data coming from CAN bus, radars, IMU, and DGNSS. This computer generates the ego-vehicle log with the raw data and the time when it was received. A second computer stores the images coming from both cameras. The data flow can reach up to 6 Gbps when the cameras are triggered at their maximum rate. However, in this application, they are triggered at the LiDAR spinning rate, close to 10 Hz generating a data flow of 360 Mbps which can be supported for long periods of time. The last computer

is dedicated to read the LiDAR input and generate the trigger for the cameras. The custom cloud video file with the LiDAR measures is generated by this computer, as well as the log files with the triggering and acknowledgment timestamps of each image.

E. Data-labeling

Focusing on vehicle intention and trajectory prediction tasks, six types of labels are provided in this dataset.

The first one is the segmentation of the relevant actors in the scene, these are *cars*, *trucks*, *buses*, *motorcycles*, *bicycles*, and *pedestrians*. These labels are automatically generated using the Detectron framework [17]. To do so, the top-class state-of-the-art Mask-R-CNN [18] model with a ResNet-101 [19] backbone is used as an instance segmentation engine. The raw output detections are provided as bounding boxes and contours, moreover, a temporal integration of the detections is provided. First the detections with a confidence value lower than 0.5 are filtered out. Then a non-maximal suppression algorithm is applied. Finally, a Hungarian Matrix algorithm uses the modified Intersection over Union (mIoU) as the inverse of the distance (Eq. 1 where A_1 and A_2 are the evaluated areas) to establish a temporal association between detections and assigning the same id to them.

$$\text{mIoU} = A_1 \cap A_2 / \min\{A_1, A_2\} \tag{1}$$

In the next step, vehicles trajectories are labeled on the images. High-accurate lateral movements of vehicles are crucial for prediction tasks. Each vehicle in the scene is labeled with a unique id and tracked since it appears until it leaves the image or it becomes irrelevant for the scene understanding, i.e. when another physically separated lane is reached. Manual annotations in a total of more than 6 hours is an intensive task. A semi-automatic process is carried out by manually selecting symmetrically-placed key points in the vehicle frame. These key points are automatically tracked by a Median Flow tracking algorithm [13]. The tracking process is supervised and any deviation is corrected. The bounding box that fits the initial key points and its center along the frames are provided to represent the position and lateral movement of the vehicle.

Image labels and LiDAR detections are used together to produce vehicle trajectories. The positions of the vehicles on each frame are computed using the lateral image coordinates and the LiDAR detection range. These positions are provided in the camera and the LiDAR reference systems.

Knowing the structure of the road is of paramount importance. The relative positioning of surrounding vehicles w.r.t. the road lanes enhances the scene understanding. A custom lane detection system [20] detects and tracks each individual road lane markings. The images are analyzed into a BEV perspective and vehicle detections are removed to prevent failures in the process. Lanes are modeled as a 2nd order polynomial and their coefficients represent lateral distance, angular misalignment, and lane curvature.

The relative positioning of the ego-vehicle w.r.t the road surface is sometimes useful, i.e. for BEV pitch or height

correction, and obstacle or vehicle detection. The ground plane coefficients have been computed using a RANSAC algorithm and the point cloud as input. A cube with 20 meters edges is defined to select the points used to segment the plane.

Lane changes are presented as events defined by the time, or equivalently the frame when the center of the vehicle crosses the line between two lanes. They are evaluated using the reconstructed trajectories and the road structure information and also added manually.

F. Web access

The dataset is publicly accessible at *http://prevention-dataset.invett.es*. For simplicity, all the files of each drive have been packaged in two files due to the large size of the raw data. Direct download of the raw data is available, and, for unstable connections, an alternative by-parts download method has been released as well. The post-processed data and the labels can be downloaded independently.

G. Data-format

The database format is structured as follows, where X, Y, N, and M are used to define the record number, the drive order in each record, the camera identifier, and the radar identifier respectively. Fig. 3 shows the directory tree of a record.

Extrinsic and intrinsic sensor calibration files are provided for each record.

Transformations between the camera reference system and the image plane are enabled as it is described in Eq. 2 where u and v are pixel-coordinates in the undistorted image. The intrinsic camera calibration file contains the image resolution expressed as $[width, \ height]$, intrinsic camera parameters as $[f_x, \ sk, \ c_x, \ f_y, \ c_y]$ and distortion coefficients as $[k_1, \ k_2, \ k_3, \ p_1, \ p_2]$.

$$w \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & sk & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} p_C \qquad (2)$$

The extrinsic calibration parameters are provided as a 3D rotation matrix $\mathbf{R}$ defined in a row order plus a translation vector $\mathbf{t}$. A homogeneous transformation matrix $\mathbf{T}$ is generated with them as it is shown in Eq. 3. Inverse transformations are possible by using $\mathbf{T}^{-1}$.

$$\mathbf{T} = \begin{bmatrix} \mathbf{R}_{3\times3} & \mathbf{t}_{3\times1} \\ \mathbf{0}_{1\times3} & 1 \end{bmatrix} \qquad (3)$$

The extrinsic camera calibration file contains the parameters needed to build the transformation matrix to convert points from the LiDAR to the camera reference system. The extrinsic radar calibration files provide the parameters of the transformation matrix to transform points from the radar reference system to the vehicle reference system. Transforming points from radar to vehicle reference system assumes that radar detections are at $z = 0$. Finally, the extrinsic Velodyne calibration file enables the transformation of points from the Velodyne to the vehicle reference system.

```
DATABASE
 └── RecordX
      └── DriveY
           ├── detection_cameraN
           │    ├── detections.txt
           │    ├── detections_filtered.txt
           │    ├── detections_tracked.txt
           │    ├── labels.txt
           │    ├── lanes.txt
           │    ├── lane_changes.txt
           │    └── trajectories.txt
           ├── detection_cloud
           │    └── ground_coefficients.txt
           ├── detection_radar
           │    └── detections_radarM.txt
           ├── logs
           │    ├── log_ego-vehicle.txt
           │    └── log_cameraN.txt
           ├── pcap_radar.pcapng
           ├── video_cameraN.raw
           └── video_velodyne.bin
      ├── cameraN_extrinsic_calibration.dat
      ├── cameraN_intrinsic_calibration.dat
      ├── radarM_extrinsic_calibration.dat
      └── velodyne_extrinsic_calibration.dat
```
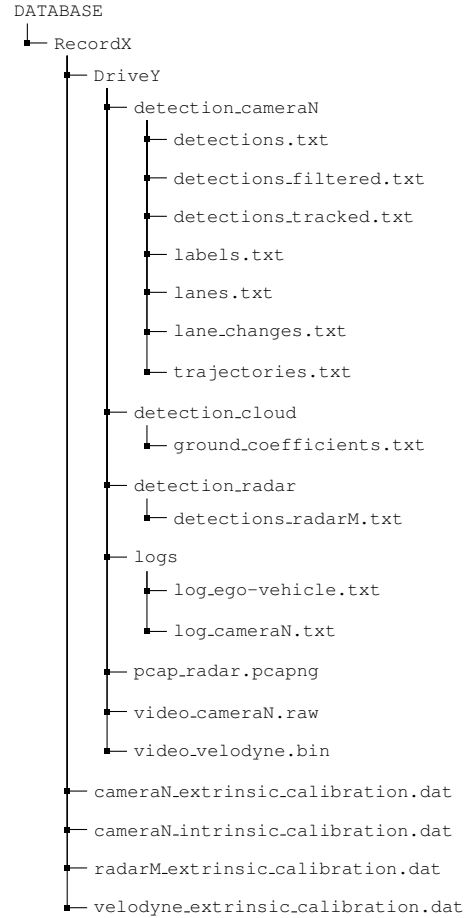
Fig. 3. Dataset Format.

Each drive contains the raw sensor inputs. Two raw videos and a copy of the Velodyne and radars data transmissions are provided for each drive. The original image height has been reduced to half dimension by removing top and bottom equal bands with irrelevant information to keep the file size smaller. Each single channel image is formed by 1920×600 bytes in a *BayerBG* pattern codification. Velodyne data is codified in a custom cloud video data format. Each *frame* has a fixed size and stores the cloud number, each triggering time, the 3D information, and the returned intensity. The cloud video structure definition is provided with the cloud video file. Radar data is provided in parsed format in a text file due to the complex parsing procedure.

Folder `detection_cameraN` contains different information extracted directly from or related to images:

- `detections.txt` provides bounding boxes and contours of objects in the scene automatically generated by a CNN. `detections_filtered.txt` are the same detections with a minimum confidence value of 0.5 and non-maximal suppression. `detections_tracked.txt` is the result of a temporal tracking of the filtered detections, assigning a unique id to each object along the frames. Data is stored as a sequence of $[frame, id, class, x_i, y_i, x_f, y_f, conf,$

n] values followed by n tuples of x and y coordinates that represents the contour.

- `labels.txt` is a sequence of [$frame$, id, x, y, $width$, $height$] values for each manual annotation that describes an area centered in the vehicle.
- `lanes.txt` is a sequence of [$frame$, n, c_0, c_1, c_2] values that represents the n lane lines in the scene as a 2$^{\text{nd}}$ order polynomial (Eq. 4).

$$y = c_2 x^2 + c_1 x + c_0 \qquad (4)$$

- `lane_change.txt` is a list of four values [id, $type$, $frame$, $val1$, $val2$] that indicates a type of lane change that could be *left* (1) or *right* (2) performed by the vehicle with that id at that frame. The parameter $val1$ is used for time-lapse events annotations, such as beginning of the lane changes. Lane change can be also labeled as *cut-in* (1), *cut-out* (2) when the vehicle arrives to or leaves the ego-lane or *none* (0) otherwise. This information is codified in $val2$. Moreover, some anomalous circumstances are labeled as *hazardous* (3), e.g. extreme lane changes, stopped vehicles on the shoulder, or emergency vehicle overtaking. These situations are maintained in the time and are labeled as the initial and final frame making use of the $val1$ parameter. Events related with pedestrians such as *zebra crossing* are also labeled (4) defining the frame when the pedestrian starts to cross or he/she is seen for the first time. The end of the zebra crossing event is defined using parameter $val1$.
- `trajectories.txt` is a sequence of [$frame$, id, x_c, y_c, z_c, x_l, y_l, z_l] values that represents the position of a vehicle in the camera and the LiDAR reference system.

Folder `detection_cloud` holds the file `ground_coefficients.txt` where the coefficients of the principal plane are stored as a sequence of [$frame$, A, B, C, D] values of an scalar plane equation (Eq. 5) expressed in m^{-1} and relative to the Velodyne reference system.

$$Ax + By + Cz + D = 0 \qquad (5)$$

File `detections_radarM.txt` contains parsed radar data. Wide-range (1 and 3) and long-range (2) radars are different models, thus provide mostly common but also specific information. Each radar sends a transmission with a snapshot of 25 or 40 objects for the wide-range and the long-range respectively.

- Object #: number of the object in the transmission.
- ID: unique id (wide-range only).
- y, x: longitudinal and lateral distances in meters.
- v_y, v_x: longitudinal and lateral speeds in m/s.
- RCS: radar cross section.
- LT/PoE: life time in seconds or probability of existence (wide-range/long-range).
- t: reception time in microseconds.

Simple C/C++ and MATLAB examples are provided in the dataset website to load, use, and show the available
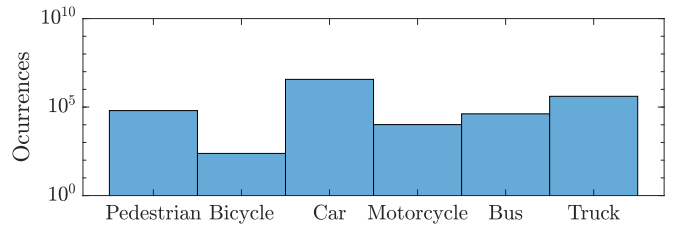

Fig. 4.   Histogram of classes occurrences in the dataset.

information. As an example of this visual information Fig. 7 shows a sequence of integrated information of point clouds, images, radar, and automatic and manual annotations. The top row shows the acquisitions of the front camera and the bottom row the back camera. Both rows show a forward time sequence from left to right. Fig. 8 is a collage of some of the events that could be found in the dataset such as cut-in and cut-out maneuvers, lane changes to both sides, including entrance and exit ramps, hazardous situations, and pedestrian zebra crossings.

III. DISCUSSION

The dataset presented in this work represents a big effort to contribute to the ongoing research works in inter-vehicle interactions such as vehicle trajectory and intention prediction. Some of the benefits of this dataset are highlighted below.

There are no other lane-change prediction oriented datasets which include radar detections. Radars are one of the most reliable sensors for object detection across varied weather and lighting conditions. Furthermore, radars are one of the sensors with the longer detection range. Radar detections are of utmost importance not just for intention or vehicle trajectory prediction, but also for world modeling.

This dataset contains more than 4 million detections, including vehicles and pedestrians. More than 3.5 million of these detections are cars and 0.5 million are trucks. Pedestrian, motorcycle, bicycle, and bus classes are a minority. This is explained because most of the recording time was on highways which is where pedestrians and bicycles seldom appear. Trucks, buses, and motorcycles are obviously a minority with respect to cars, according to the vehicle pool. Additional efforts can be made to increase the pedestrian and bicycle detections and to extend the use of this dataset to pedestrian path predictions. Fig. 4 shows the proportion of each class in the dataset, note that the y-axis is logarithmic.

Manual labeling effort is focused on trajectories generation. Because of this, the lateral movement of the vehicles has been manually labeled. To relax the intensity of the labeling task, labelers used a semiautomatic labeling tool based on manual anchors and an automatic tracking system. Fig. 5 shows the histogram of the trajectories duration. The last bin (90-100 seconds) is greater than the previous ones because the duration of the trajectories has been truncated to 100 seconds for a more detailed representation. There are a few trajectories with a duration up to 300 seconds.

The accurate lateral positioning of the vehicles is used to generate precise trajectories, both longitudinal and lateral,
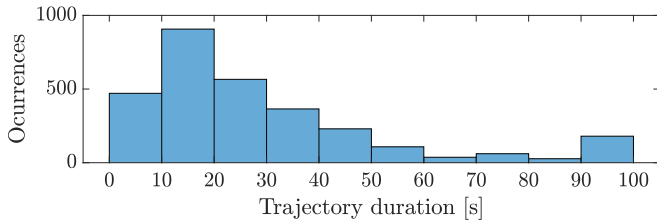
Fig. 5. Histogram of trajectories' duration.

fusing the available information such as LiDAR and radar range detections. There are more than 3000 trajectories with a total of 1.3 million manual annotations that identify unequivocally each vehicle with a unique id and the image coordinates that represent its positions along the frames. Fig. 6 shows a 2D histogram local trajectory reconstructions for a single recording. Warm colors represent more populated areas near the ego vehicle. The surrounding vehicles are detected and their trajectories are precisely reconstructed both in the rear and in the front of the ego-vehicle using the LiDAR and the cameras. Radar detections are also used to generate trajectories in the areas where the LiDAR cannot provide information.

More than 900 lane changes have been manually labeled. Table II shows statistics of the different type of lane changes for each record.

TABLE II
LANE CHANGE STATISTICS

Record #	1	2	3	4	5
Left LC	22	36	46	139	170
Right LC	51	48	47	175	178
Mean frames per LC	40.6078				
Mean time per LC	3.76 s				

Apart from the benefits of this dataset, there are some issues that limit its scope, improving or working around them could contribute to reaching better results.

The detection range is up to 200 meters, but only on the front side of the vehicle. The detection range in the rear part is conditioned by the LiDAR because there are not radars in the rear of the vehicle. The LiDAR range is limited up to 100 meters, however, in practice, the range is reduced up to 70 meters due to the geometrical configuration of the layers. Detections up to 100 meters range are only produced with high vehicles, such as trucks or buses. These negative issues could be easily solved installing radars in the rear part of the vehicle and replacing the LiDAR by another one featuring more layers and/or a longer detection range.

IV. CONCLUSIONS AND FUTURE WORK

As a conclusion, we would like to highlight that more than six hours of naturalistic urban and highway drivings with sensor redundancy have been recorded. Different raw data sources such as cameras, LiDAR and radars are provided as well as manual annotations, automatic detections, and high-level information. This data can be widely used in
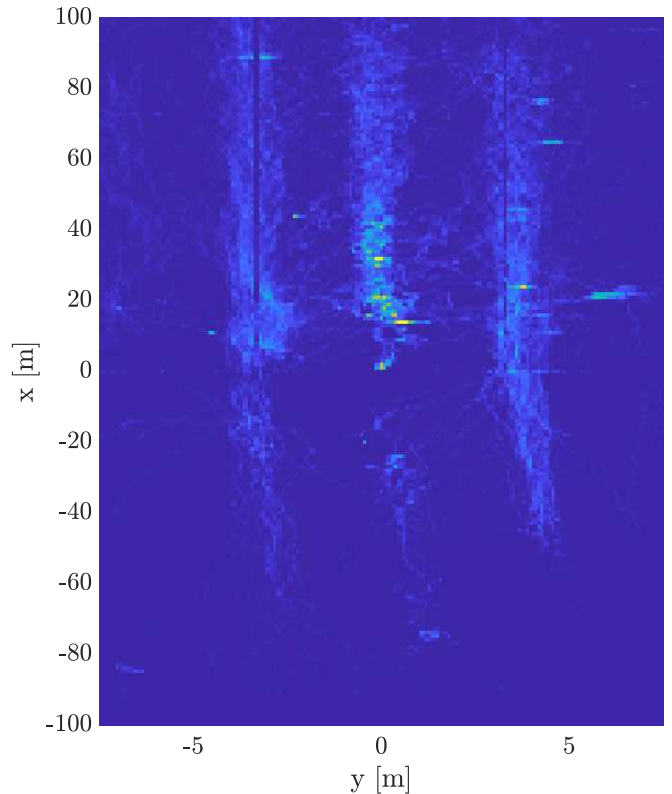

Fig. 6. Heat map of accumulated vehicle occupancy. Occupancy is computed using LiDAR plus cameras and radar detections. The areas covered by the radars, such as the frontal area shows a higher occupancy density caused by the double sensor detection.

different research contexts, focusing on intention and vehicle trajectory prediction. According to this, a special effort has been made to label all the actions that appear in the scene, in which the ego-vehicle can be involved/affected or not such as all type of lane changes and potential hazards.

As future work, the most important task would be to continue extending the dataset, enhancing occurrences of minority classes or relevant events for the driver intention or vehicle trajectory predictions. Following this, a trajectory or intention prediction benchmark can be set up with specific sequences to challenge the scientific community.
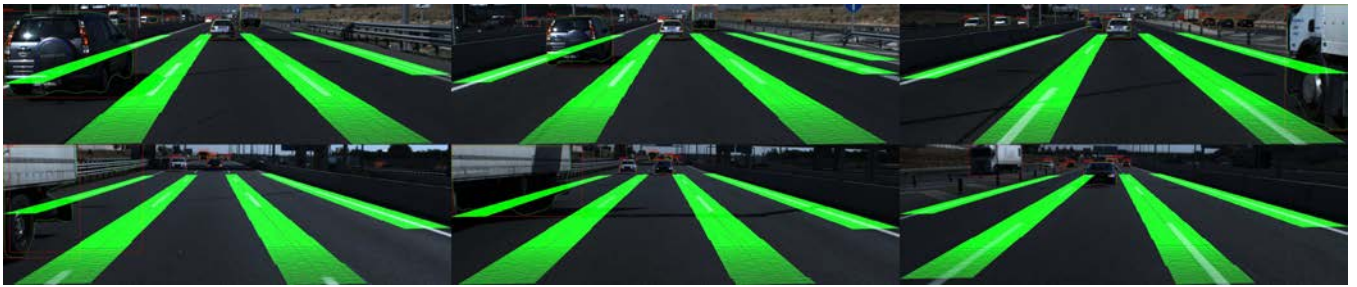
ACKNOWLEDGMENT

Fig. 7. Sequence of vehicle detections and tracking. Top images are front and bottom images are back. From left to right are progressing in time.

Fig. 8. Example of Different occurrences in the dataset. From left to right and top to bottom, cut-in, cut-out, left-lane change, right-lane change, hazardous event and pedestrian crossing.

REFERENCES

[1] C. Hermes, C. Wohler, K. Schenk, and F. Kummert, "Long-term vehicle motion prediction," in *2009 IEEE Intelligent Vehicles Symposium*, June 2009, pp. 652–657.

[2] R. S. Tomar, S. Verma, and G. S. Tomar, "Prediction of lane change trajectories through neural network," in *2010 International Conference on Computational Intelligence and Communication Networks*, Nov 2010, pp. 249–253.

[3] R. Schubert, C. Adam, M. Obst, N. Mattern, V. Leonhardt, and G. Wanielik, "Empirical evaluation of vehicular models for ego motion estimation," in *Intelligent Vehicles Symposium (IV), 2011 IEEE*. IEEE, 2011, pp. 534–539.

[4] R. Izquierdo, I. Parra, J. Munoz-Bulnes, D. Fernández-Llorca, and M. A. Sotelo, "Vehicle trajectory and lane change prediction using ann and svm classifiers," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, Oct 2017, pp. 1–6.

[5] V. Leonhardt and G. Wanielik, "Neural network for lane change prediction assessing driving situation, driver behavior and vehicle movement," in *Intelligent Transportation Systems (ITSC), 2017 IEEE 20th International Conference on*. IEEE, 2017, pp. 1–6.

[6] S. Yoon and D. Kum, "The multilayer perceptron approach to lateral motion prediction of surrounding vehicles for autonomous vehicles," in *2016 IEEE Intelligent Vehicles Symposium (IV)*, June 2016, pp. 1307–1312.

[7] Y. Hu, W. Zhan, and M. Tomizuka, "A framework for probabilistic generic traffic scene prediction," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 2790–2796.

[8] J. Colyar and J. Halkias, "Ngsim - us highway 101 dataset," Jan 2007, https://www.fhwa.dot.gov/publications/research/operations/07030/07030.pdf [Online; accessed Jan. 9 2019].

[9] J. Halkias and J. Colyar, https://www.fhwa.dot.gov/publications/research/operations/06137/06137.pdf, December 2006, nGSIM - Interstate 80 Freeway Dataset [Online; accessed Jan. 9 2019].

[10] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems," in *2018 IEEE 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018.

[11] H. Zhao, C. Wang, Y. Lin, F. Guillemard, S. Geronimi, and F. Aioun, "On-road vehicle trajectory collection and scene-based lane change analysis: Part i," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 1, pp. 192–205, Jan 2017, http://poss.pku.edu.cn/download [Online; accessed Jan. 9 2019].

[12] X. Huang, X. Cheng, Q. Geng, B. Cao, D. Zhou, P. Wang, Y. Lin, and R. Yang, "The apolloscape dataset for autonomous driving," *arXiv preprint arXiv:1803.06184*, 2018.

[13] R. Izquierdo, I. Parra, C. Salinas, D. Fernández-Llorca, and M. A. Sotelo, "Semi-automatic high-accuracy labelling tool for multi-modal long-range sensor dataset," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, June 2018, pp. 1786–1791.

[14] R. Izquierdo, I. Parra, D. Fernández-Llorca, and M. A. Sotelo, "Multi-radar self-calibration method using high-definition digital maps for autonomous driving," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Nov 2018, pp. 2197–2202.

[15] G. L. Smith, S. F. Schmidt, and L. A. McGee, *Application of statistical filter theory to the optimal estimation of position and velocity on board a circumlunar vehicle*. National Aeronautics and Space Administration, 1962.

[16] R. Pepy, A. Lambert, and H. Mounier, "Path planning using a dynamic vehicle model," in *2006 2nd International Conference on Information Communication Technologies*, vol. 1, April 2006, pp. 781–786.

[17] R. Girshick, I. Radosavovic, G. Gkioxari, P. Dollár, and K. He, "Detectron," https://github.com/facebookresearch/detectron, 2018.

[18] K. He, G. Gkioxari, P. Dollár, and R. B. Girshick, "Mask R-CNN," *CoRR*, vol. abs/1703.06870, 2017. [Online]. Available: http://arxiv.org/abs/1703.06870

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 770–778.

[20] C. Fernández, R. Izquierdo, D. F. Llorca, and M. A. Sotelo, "Road curb and lanes detection for autonomous driving on urban scenarios," in *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Oct 2014, pp. 1964–1969.