Identifying Driver Behaviors using Trajectory Features for Vehicle Navigation

Ernest Cheung[1], Aniket Bera[1], Emily Kubin[2], Kurt Gray[2], and Dinesh Manocha[3]
Website/Video: http://gamma.cs.unc.edu/TDBM

Abstract— We present a novel approach to automatically identify driver behaviors from vehicle trajectories and use them for safe navigation of autonomous vehicles. We propose a novel set of features that can be easily extracted from car trajectories. We derive a data-driven mapping between these features and six driver behaviors using an elaborate web-based user study. We also compute a summarized score indicating a level of awareness that is needed while driving next to other vehicles. We also incorporate our algorithm into a vehicle navigation simulation system and demonstrate its benefits in terms of safer real-time navigation, while driving next to aggressive or dangerous drivers.

I. INTRODUCTION

Identifying dangerous drivers is crucial in developing safe autonomous driving algorithms and advanced driving assistant systems. The problem has been extensively studied in transportation and urban planning research [1]. However, prior work usually correlates driver' behaviors with their backgrounds (e.g., driver age, response to questionnaires, etc.). On the other hand, to develop autonomous vehicle systems, we need to understand the behavior of surrounding drivers using only the sensor data. As with to a human driver, an autonomous navigation algorithm that can predict other vehicle's driving behavior can navigate safely and efficiently avoid getting near dangerous drivers.

Prior work in transportation research [2], [1] often characterizes drivers using their levels of aggressiveness and carefulness. Several works in modeling pedestrian trajectories [3] and navigation [4] algorithms have applied psychological theory to capture human behavior. Current autonomous driving systems uses a range of different algorithms to process sensor data. Object detection and semantic understating methods are applied to obtain trajectory data [5]. Some work [6] uses end-to-end approaches to make navigation decisions from the sensor inputs (e.g. camera images, LIDAR data, etc.).

Main Results: We present a novel approach to automatically identifying driver behaviors from vehicle trajectories. We perform an extensive user study to learn the relationship and establish a mathematical mapping between extracted vehicular trajectories and the underlying driving behaviors: Trajectory to Driver Behavior Mapping (TDBM). TDBM enables a navigation algorithm to automatically classify the

driving behavior of other vehicles. We also demonstrate simulated scenarios where navigating with our improved navigation scheme is safer.

Our approach takes into account different trajectory features. We use five different features, which can be easily extracted from vehicle trajectories and used to classify driving behaviors. We show that selecting a subset of these features is more favorable than selecting the currently used ones to produce a strong regression model that maps to driving behaviors.

As compared to prior algorithms, our algorithm offers the following benefits:

1. Driving Behavior Computation: We present a data-driven algorithm to compute TDBM. We conducted a comprehensive user survey to establish a mapping between five features and six different driving behaviors. We further conduct factor analysis on the six behaviors, which are derived from two commonly studied behaviors: aggressiveness and carefulness. The results show that there exists a latent variable that can summarize these driving behaviors and that can be used to measure the level of awareness that one should have when driving next to a vehicle. In the same study, we examine how much attention a human would pay to such a vehicle when it is driving in different relative locations.

2. Improved Realtime Navigation: We compute the features and identify the driving behaviors using TDBM. We enhance an existing Autonomous Driving Algorithm [7] to navigate according to the neighboring drivers' behavior. Our navigation algorithm identifies potentially dangerous drivers in realtime and chooses a path that avoids potentially dangerous drivers.

An overview of our approach is shown in Figure 1. The rest of the paper is organized as follows. We give a brief overview of prior work in Section II. We introduce the new trajectory features that are used to identify the driver behaviors in Section III. We present our data-driven mapping algorithm (TDBM) in Section IV and use it for autonomous car navigation in Section V.

II. RELATED WORKS

A. Studies on Driving Behaviors

There has been a wide range of work studying drivers' behaviors in Social Psychology and Transportation. Feng et al. [2] proposed five driver characteristics (age, gender, year of driving experience, personality via blood test, and education level) and four environmental factors (weather, traffic situation, quality of road infrastructure, and other cars' behavior), and mapped them to 3 levels of aggressiveness

[1] Authors from the Department of Computer Science, University of North Carolina at Chapel Hill, USA
[2] Authors from the Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, USA
[3] Author from the Department of Computer Science and Electrical Computer Engg, University of Maryland at College Park
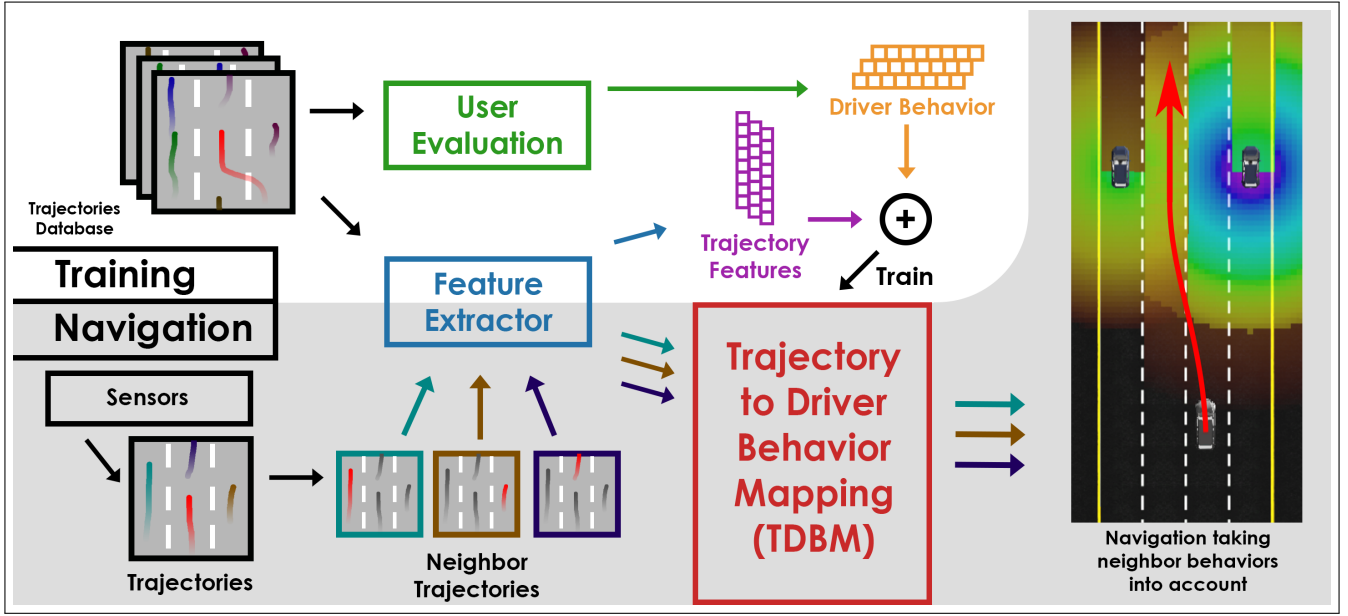
Fig. 1. Overview of our Algorithm: During the training of TDBM, we extract features from the trajectory database and conduct a user evaluation to find the mapping between them. During the navigation stage, we compute a set of trajectory and extract the features, then compute the driving behavior using TDBM. Finally, we plan for real-time navigation, taking into account these driver behaviors.

(driving safely, verbally abusing other drivers, and taking action against other drivers). Aljaafreh et al. [8] categorized driving behaviors into 4 classes: Below normal, Normal, Aggressive, and Very aggressive, in accordance to accelerometer data. Social Psychology studies [9], [10] have examined the aggressiveness according to the background of the driver, including age, gender, violation records, power of cars, occupation, etc. Moulouha et al. [11] designed a questionnaire on subjects' previous aggressive driving behavior, and concluded that these drivers also repeated those behaviors under a simulated environment. Meiring et al. [1] used several statistical reports to conclude that distracted behaviors and drunk behaviors are also serious threats to road safety. Many of the driver features used by these prior methods cannot be easily computed in new, unknown environments using current sensors. Our work uses trajectory data which can be extracted from sensor data in most autonomous driving systems.

B. Trajectories Features

Murphey et al. [12] conducted an analysis on the aggressiveness of drivers and found that longitudinal (changing lanes) jerk is more related to aggressiveness than progressive (along the lane) jerk (i.e. rate of change in acceleration). Mohamad et al. [13] detected abnormal driving styles using speed, acceleration, and steering wheel movement, which indicate direction of vehicles. Aufrere et al. [14] proposed a model-driven based approach to detect and track road edges. Qi et al. [15] studied driving styles with respect to speed and acceleration. Shi et al. [16] pointed out that deceleration is not very indicative of aggressiveness of drivers, but measurements of throttle opening, which is associated with acceleration, is more helpful in identifying aggressive drivers. Wang et al. [17] classified drivers into two categories,

aggressive and normal, using speed and throttle opening captured by a simulator. There are also some other works [18], [19] that leverage massive simulated data for training algorithms to extract trajectories of other obstacles (e.g. pedestrian).

Sadigh et al. [20] proposed a data-driven model based on Convex Markov Chains to predict whether a driver is paying attention while driving. There are considerable works on development of in-car smart systems to alert users when they are found driving distracted by indicating nearby vehicles [21], departures from lane markers and drivers' appearances and road conditions [22], and trajectories computed by camera, IMU and GPS [23].

Instead of directly analyzing real-world data, many methods model driving behaviors as input parameters to generate driving simulations. Treiber et al. [24] proposed a lane following model, that controls the speed of the car using desired velocity, minimum spacing, desired time headway, acceleration, and maximum breaking deceleration. Kesting et al. [25] proposed a lane changing model, that makes lane changing decisions based on the speed advantage gained and the speed disadvantage imposed on the other vehicles, using a franticness and a politeness factor. Choudhury et al. [26] proposed a complex lane changing model, composed of desired speed, desired time gap, jam distance, maximum acceleration, desired deceleration, coolness factor, minimum acceptable gap, etc.

We combine a set of selected features proposed by previous works in terms of behavior mapping and simulation with two new trajectory features, lane following metric and relative speed metric. Then, we use variable selection to select a subset of features that can produce a good regression model.

C. Autonomous Car Navigation

There is substantial work on autonomous vehicle navigation [27], [28], [29], [30], [31], [32]. Ziegler et al. [33] presented a navigation approach that is capable of navigating through the historic Bertha Benz route in Germany. Numerous navigation approaches [34], [35], [36], [37] have been proposed in the DAPRA Urban Grand Challenge and the Grand Cooperative Driving Challenge. Recent work proposed by Best et al. [7], AutonoVi, presented an improved navigation algorithm that takes into account dynamic lane changes, steering and acceleration planning, and various other factors. Our approach is complimentary to these methods and can be combined with them.

D. Adaptation to Human Drivers' Behavior

Sadigh et al. [38] observed that an autonomous car's action could also affect neighboring human drivers' behavior, and studied how humans will react when the autonomous car performs certain actions [39]. Huang et al. [40] presented techniques for making autonomous car actions easily understandable to humans drivers. They also proposed an active learning approach [41] to model human driving behavior by showing examples of how a human driver will pick their preference out of a given set of trajectories. While this stream of work went further to take into account how humans would react to an autonomous car's action, it also emphasized the importance of a robot navigating according to other drivers' behavior.

III. METHODOLOGY

In this section, we present the two novel trajectory features that are used to identify driver behaviors. We also compare their performance with other features and give an overview of driver behavior metrics used in our navigation algorithm.

A. Features

The goal of our work is to extract a set of trajectory features that can be mapped properly to driving behaviors. We assume that the trajectories have been extracted from the sensor data. Many of the previous works deal with different driver characteristics: driver background, accelerometer use, throttle opening, etc., which may not be available for an autonomous vehicle in new and uncertain environments. Moreover, in the simulation models described in Section II-B, a lot of features cannot be measured from trajectories with insufficient lane-changing samples: comfortable breaking deceleration, desired time headway, etc. Therefore, we derive some variants of features that can be easily extracted from the trajectories and summarize them in Table I. These features are further shortlisted with the results from a user study described in the next section.

1) Acceleration: As pointed out in several prior works [12], [13], [16], [17], acceleration is often correlated with driver aggressiveness. While previous studies [12] concluded that longitudinal jerk can reflect aggressiveness better than progressive jerk, our goal is to use features that also correlate with all the driving styles, instead of just aggressiveness. Therefore, we include both longitudinal jerk j_l and progressive jerk j_p in our computations.

Symbol	Notation	Description
f_0	v_{front}	Average relative speed to the car in front
f_1	v_{back}	Average relative speed to the car in the back
f_2	v_{left}	Average relative speed to cars in the left lane
f_3	v_{right}	Average relative speed to cars in the right lane
f_4	v_{nei}	Relative speed to neighbors
f_5	v_{avg}	Average velocity
f_6	s_{front}	Distance with front car
f_7	j_l	Longitudinal jerk
f_8	j_p	Progressive jerk
f_9	s_{center}	Lane following metric

TABLE I

We CONSIDERED TEN CANDIDATE FEATURES $f_0, .., f_9$ FOR SELECTION. FEATURES HIGHLIGHTED IN GREEN ARE SELECTED FOR MAPPING TO BEHAVIOR-METRICS ONLY, AND THOSE IN BLUE ARE SELECTED FOR MAPPING TO BOTH BEHAVIOR-METRICS AND ATTENTION METRICS.

2) Lane following: Previous work [23] proposed a metric measureing the extent of lane following that depends on the mean and standard deviation of lane drifting and lane weaving. We propose a feature that also depends on lane drifting, but distinguishes between drivers who keep drifting left and right within a lane and those who are driving straight but not along the center of the lane. Moreover, we compensate for the extent of lane drifting while performing lane changing to avoid capturing normal lane changing behaviors into this metric.

Given y_l, which is the center longitudinal position of the lane that the targeted car is in, and $y(t)$, which is the longitudinal position of the car at time t, we detect a lane changing event when the car has departed from one lane to the another and remained in the new lane for at least k seconds.

With a set of changing lane events happened at time t_i, $C = \{t_1, t_2, ..., t_n\}$, the lane drift metric $s_C(t)$ is measured as below:

$$s_C(t) = \begin{cases} 0, & \text{if } \exists t \in C \text{ s.t. } t \in [t-k, t+k], \\ y(t) - y_l, & \text{otherwise.} \end{cases} \quad (1)$$

We use a term that measures the previous τ seconds of rate of change in drifting to differentiate lane drifts from those drivers who are driving straight but off the center of the lane. Our overall lane following metric is illustrated in Figure 2 and defined as:

$$s_{center} = \int |s_C(t)| \left[\mu + \int_{t-\tau}^{t} |s'_\emptyset(t)| dt \right] dt, \quad (2)$$

where μ is a parameter that distinguish drivers who are driving off the center of the lane and those who are along.

3) Relative Speed: Relative speed has been used to evaluate the aggressiveness of drivers [15]. However, directly measuring the relative speed using v_{front}, v_{back}, v_{left} and v_{right} has many issues. First, such a feature sometimes does not exist as there may be no car next to the target car. Second, these features might not be directly related to the driving behavior of the car. While driving substantially faster than other cars would be perceived as aggression,
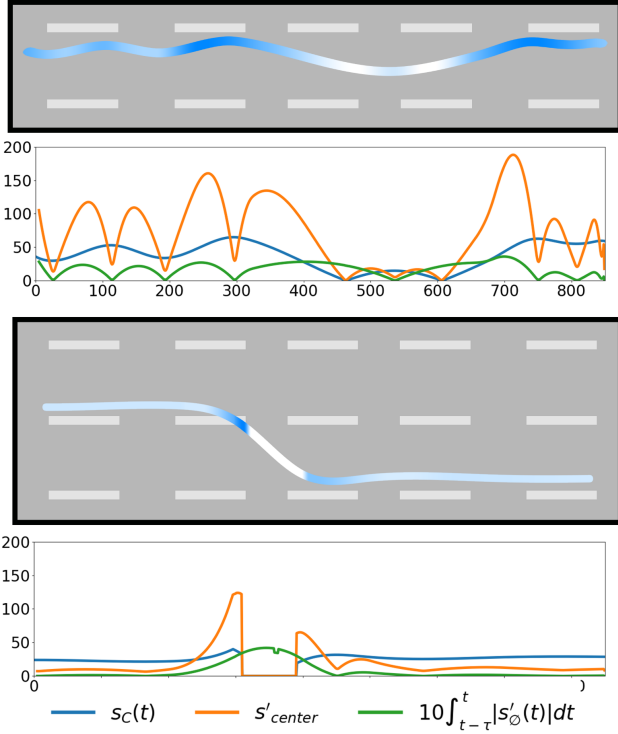
Fig. 2. Illustration of the lane drift metric ($|s_c(t)|$), and the lane following metric (s_{center}). The lane following metric for the trajectories above is the sum of the area under the plot of s'_{center}. The first example shows that our lane following metric (s_{center}) captures the 'drifting behavior' in the top example. The second example shows that our metric deliberately neglected the 'lane changing' phase, and has a significant lower magnitude for 'driving straight off the center', since this metric is designed to merely capture the dangerous 'drifting behavior'.

driving slower might not necessarily imply that the driver is non-aggressive. Third, computing such an average velocity requires knowledge about the trajectories and range of speeds of the neighboring vehicles. Given these considerations, we design the following metric to capture the relationship between the driving behavior and the relative speed with respect to neighboring cars:

$$ v_{nei} = \int \sum_{n \in N} \max\left(0, \frac{v(t) - v_n(t)}{dist(x(t), x_n(t))}\right) dt, \qquad (3) $$

where N denotes the set containing all neighboring cars within a large range (e.g., a one-mile radius). $x(t)$, $v(t)$, $x_n(t)$, $v_n(t)$ denote the position and the speed of the targeting car, and the position and the speed of the neighbor n, respectively.

B. Driving Behavior Metrics

As discussed in Section II-A, aggressiveness [2], [8], [42] and carefulness [1], [20], [43] are two metrics that have been used to identify road safety threats. Typically, social psychologists add related items into studies to leverage robustness and the observed effects. Therefore, we would like to evaluate four more driving behaviors: Reckless, Threatening, Cautious, and Timid. They are listed in Table II.

C. Attention Metrics

Observing different maneuvers of other drivers on the road can result in paying more attention to those drivers. However, the relative position of such drivers (with respect to the targeted vehicle) would affect the level of attention that one is paying to them. For instance, one would pay more attention to a vehicle in the front making frequent stops, as opposed to a following vehicle. We would like to understand how much attention a driver will pay to the targeted car when the user assumes that he or she is driving in different relative positions than the target. We study four different relative positions: preceding, following, adjacent to and far away from the targeted vehicle, also listed in Table II. These positions affect the level of attention one would pay when driving in that relative position.

Symbol	Description	Symbol	Level of Attention when
b_0	Aggressive	b_6	following the target
b_1	Reckless	b_7	preceding the target
b_2	Threatening	b_8	driving next to the target
b_3	Careful	b_9	far from the target
b_4	Cautious		
b_5	Timid		

TABLE II
SIX DRIVING BEHAVIOR METRICS (b_0, b_1, ...,b_5) AND
FOUR ATTENTION METRICS (b_6, b_7, b_8, b_9) USED IN TDBM

IV. DATA-DRIVEN MAPPING

We designed a user study, involving 100 participants to identify driver behaviors from videos rendered from the Interstate 80 Freeway Dataset [44]. The video dataset consists of 45 minutes of vehicle trajectories, captured in a 1650 feet section on I-80 in California, US. The videos were first annotated automatically using a proprietary code developed in the NGSIM program, and then manually checked and corrected. The raw videos provided in the dataset are low in quality and divided into seven different segments with different camera angles. Therefore, we have rendered the videos using a game engine, Unreal Engine, to provide a stable and consistent view for the users in the survey. The virtual cameras have a fixed transform to the targeted car, which is highlighted in red, and will follow it throughout the video.

Figure 3 shows snapshots of the videos used in the user study. The participants were asked to rate the six behaviors we described in Section III-B on a 7-point scale: {Strongly disagree, Disagree, Somewhat disagree, Neither agree or disagree, Somewhat agree, Agree, Strongly agree}. This was followed by another question on how much attention they would be paying if they were in different positions relative to the targeted car, as described in Section III-C, on a 5-point scale, where -2 indicates not at all, 0 indicates a moderate amount and 2 indicates a lot.

A. Data Pre-Processing

We perform data augmentation to make sure that the dataset has a sufficiently wide spectrum of driving behaviors corresponding to lane changes, fast moving cars, passing
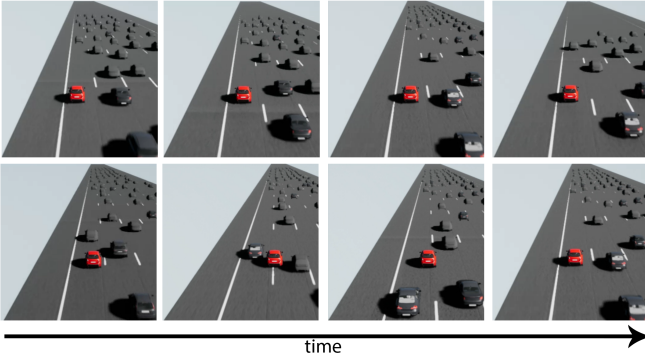
Fig. 3. Two example videos used in the user study. Participants are asked to rate the six driving behavior metrics and four attention metrics of the target car colored in red.

	v_{front}	v_{left}	v_{right}	v_{back}	v_{nei}	v_{avg}	s_{front}	j_l	j_p	s_{center}
Aggressive	-2.7	-2.08	-2.12	-1.78	-2.2	-0.65	-0.62	-1.45	-1.72	-2.05
Reckless	-2.7	-1.62	-3.55	-2.22	-1.35	-1.85	-1.38	-1.7	-3.0	-0.38
Threatening	-1.78	-1.92	-1.92	-2.92	-1.5	-1.72	-2.1	-2.1	-2.4	-0.45
Careful	-1.82	-2.52	-1.92	-2.92	-1.35	-2.58	-2.02	-2.2	-2.48	-0.45
Cautious	-2.05	-2.8	-2.2	-3.05	-1.48	-2.55	-2.0	-2.17	-2.95	-0.52
Timid	-2.17	-1.95	-2.4	-2.55	-1.52	-2.02	-2.0	-2.05	-2.85	-0.52
$Attention_{back}$	-3.35	-1.62	-2.8	-1.88	-1.58	-1.72	-1.88	-2.2	-2.9	-0.52
$Attention_{front}$	-2.62	-3.32	-2.75	-4.5	-1.78	-2.12	-2.33	-2.1	-2.48	-0.75
$Attention_{adj}$	-2.02	-2.12	-3.52	-2.35	-1.78	-1.32	-2.15	-2.5	-2.22	-1.08
$Attention_{far}$	-1.98	-2.4	-3.72	-2.45	-1.75	-1.78	-2.2	-2.48	-2.65	-0.9
Behavior Average	-2.2	-2.15	-2.35	-2.58	-1.57	-1.9	-1.69	-1.95	-2.57	-0.73
Attention Average	-2.49	-2.37	-3.2	-2.79	-1.72	-1.74	-2.14	-2.32	-2.56	-0.81
Overall Average	-2.32	-2.24	-2.69	-2.66	-1.63	-1.83	-1.87	-2.1	-2.56	-0.76

Fig. 4. The converging value (in the power of 10) of $\beta_{i,j}$ which maps a feature f_j to a behavior/attention metric b_i while performing Lasso analysis. A larger converging value indicates a higher likelihood that the feature is favourable in regression analysis, and therefore we select that value for TDBM.

cars, etc. In addition, the features described in Table I are measured using different units. To improve numerical stability during the regression analysis, we scale the data linearly using the 5th and the 95th percentile samples to minimize the effects of extreme values.

B. Feature Selection

In Section III-A and Table I, we cover a wide range of features used in previous studies that can be extracted from trajectories, along with two new metrics that attempt to summarize some of these features to avoid strong correlation between independent variables during regression analysis. In this section, we apply feature selection techniques to find out which features are most relevant to the driving behaviors.

We perform least absolute shrinkage and selection operator (LASSO[45]) analysis on six driving behaviors b_0, b_1, ..., b_5 and four attention metrics, b_6, b_7, b_8, b_9, from the user responses. The objective function for Lasso analysis conducted on b_i is:

$$\min_{\beta'_i, \beta_i} \left[\frac{1}{N} \sum_{j=1}^{N} (b_i - \beta i' - f_j^T \beta_{i,j}) \right], \text{ subject to } \sum_{j=1}^{F} |\beta_{i,j}| \leq \alpha_i, \tag{4}$$

where N is the number of survey responses and F is the number of features.

Lasso analysis performs regularization and feature selection by eliminating weak subsets of features. The parameter α_i determines the level of regularization that Lasso analysis imposes on the features. As we increase α_i, features f_j will be eliminated in a different order. Unlike regular regression analysis on a single dependent variable, our goal is to select two sets of features: one that can produce a strong regression model for all six driving behavior metrics, and one for all four attention metrics. We sample different values of α_i for all responses b_i, and record the values of α_i at which the component $\beta_{i,j}$ (which mapping feature f_j to response b_i) converges to 0. The results are shown in Figure 4, where converging values of $\beta_{i,j}$ are presented in the power of 10.

The directly computed relative speeds of the cars surrounding the targeted car are least favorable for selection for both regressions for behavior-metrics and attention-metrics.

However, our relative speed metric proposed to capture the correlation between surrounding cars and the targeted car, v_{nei} (Equation 3), is more favorable in terms of being selected. Moreover, our lane following metric, s_{center} (Equation 2), tends to be the last one eliminated as a feature in the variable selection stage.

Our goal is to find two $\alpha_{behavior}$ and $\alpha_{attention}$ that shortlist a subset of features for behavior-metric and attention-metric respectively. Note that $\alpha_{behavior} = \alpha_i$, $\forall i \in [0, 5]$, and $\alpha_{attention} = \alpha_i$, $\forall i \in [6, 9]$ for α_i defined in Equation 4. In terms of behavior, we can either pick $\{s_{center}, v_{nei}, s_{front}\}$ or $\{s_{center}, v_{nei}, s_{front}, v_{avg}, j_l\}$. Given that the mapping component between v_{avg} and j_l has high converging values, they can produce a stronger regression model for aggression, and that aggressiveness is one of the common behaviors as studied in prior literature discussed in Section II-A. We therefore select the latter set of features for behavior mapping. For mapping features with attention regions metrics, we select $\{s_{center}, v_{nei}, v_{avg}\}$.

C. Feature-Behavior Mapping

Using $\{s_{center}, v_{nei}, s_{front}, v_{avg}, j_l\}$ and $\{s_{center}, v_{nei}, v_{avg}\}$ as the features, we perform linear regression to obtain the mapping between these selected features and the drivers' behavior. We normalize the data as described in Section IV-A to increase the numerical stability of the regression process. The results we obtained are below. For $B_{behavior} = [b_0, b_1, ..., b_5]^T$, we obtain

$$\mathbf{B}_{behavior} = \begin{pmatrix} 1.63 & 4.04 & -0.46 & -0.82 & 0.88 & -2.58 \\ 1.58 & 3.08 & -0.45 & 0.02 & -0.10 & -1.67 \\ 1.35 & 4.08 & -0.58 & -0.43 & -0.28 & -1.99 \\ -1.51 & -3.17 & 1.06 & 0.51 & -0.51 & 1.39 \\ -2.47 & -2.60 & 1.43 & 0.98 & -0.82 & 1.27 \\ -3.59 & -2.19 & 1.75 & 1.73 & -0.30 & 0.61 \end{pmatrix} \begin{pmatrix} s_{center} \\ v_{nei} \\ s_{front} \\ v_{avg} \\ j_l \\ 1 \end{pmatrix} \tag{5}$$

Moreover, for $B_{attention} = [b_6, b_7, b_8, b_9]^T$,

$$\mathbf{B}_{attention} = \begin{pmatrix} B_{back} \\ B_{front} \\ B_{adj} \\ B_{far} \end{pmatrix} = \begin{pmatrix} 0.54 & 1.60 & 0.11 & -0.8 \\ -0.73 & 1.66 & 0.63 & -0.07 \\ -0.14 & 1.73 & 0.25 & 0.15 \\ 0.25 & 1.47 & 0.17 & -1.43 \end{pmatrix} \begin{pmatrix} s_{center} \\ v_{nei} \\ v_{avg} \\ 1 \end{pmatrix} \tag{6}$$

We further apply leave-one-out cross-validation to the set of samples S: enumerate through all samples $s_i \in S$ and leave s_i as a validation sample, and use the remaining samples $S - s_i$ to produce regression models $M_{i,j}$ for each behavior $b_{i,j}$. Using $M_{i,j}$, we predict the behaviors $b_{i,j}$ of

s_i. The mean prediction errors of $b_{i,j}$ using $M_{i,j}$ are listed in the table below. The mean prediction error in the cross-validation is less than 1 in a 7-point scale for all behaviors and attention metrics predicted, showing that our mappings described in Equation 5 and 6 are not over-fitted.

b_0	b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9
0.75	0.94	0.78	0.7	0.6	0.89	0.2	0.49	0.38	0.23

TABLE III

MEAN ERROR IN A 7-POINT SCALE WHEN APPLYING CROSS VALIDATION OF LINEAR REGRESSION TO MAP FEATURE TO BEHAVIOR AND ATTENTION METRICS SHOWING OUR MAPPING IS NOT OVER-FITTED.

D. Factor Analysis

Previous studies on mapping walking behavior adjectives with features used to simulate crowds [3], have applied factor analysis to find smaller numbers of primary factors that can represent the personalities or behaviors. We can apply Principal Component Analysis (PCA) to the survey response. The percentages of variance of the principal components are 73.42%, 11.97%, 7.78%, 2.96%, 2.30% and 1.58%. The results indicate that the Principal Component 1, which has variance of 73.43%, can model most of the driving behaviors.

We represent each entry of the user study response with the highest rated behavior and transform these entries into the space of the Principal Components as shown in Figure 5. If the user did not fully agree to any behavior for a video (i.e. responses to all questions are below 'Somewhat agree'), we consider that there to be no representative behavior from this entry (i.e. undefined). Also, if a response indicates more than one behavior as the strongest, then we label those behaviors as undefined if those adjectives contradict each other (i.e. one from negative adjectives {Aggressive, Reckless, Threatening} and one from positive adjectives {Careful, Cautious, Timid}). As observed in Figure 5, the distribution of the data on Principal Component 1, the three negative behavior adjectives we used in the user study, represented in warmer colors, are distributed on the negative side, while the three positive behavior adjectives are distributed on the positive side. Furthermore, the entries that suggest the users' responses were 'Strongly agree', represented by solid color plots in Figure 5, have significantly higher magnitudes in terms of value along Principal Component 1. However, for Principal Components 2 and 3, such a relationship is not observed.

Our studies show that there could be one latent variable that is negatively correlated with aggressiveness and positively correlated with carefulness. We further verify these results by analyzing the correlation of the Principal Components with the amount of awareness that the users indicated they would pay to the targeted car. We take the average of the level of attention, $\frac{b_6+b_7+b_8+b_9}{4}$, recorded for each response and plot these averages as the color on the PCA results in Figure 6. Similar results have been observed from this user evaluation, where the drivers worth more attention have a lower value of Principal Component 1, and those who worth less attention tend to have a higher value.
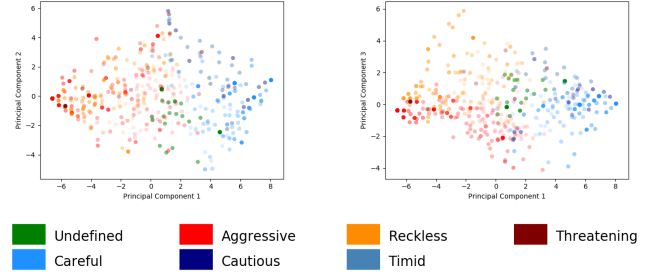

Fig. 5. Principal Component Analysis results for {Principal Components 1(PC1), PC2} (left) and {PC1, PC3} (right). The color of the data point indicates the highest rated driving behavior adjective as shown in the legends, and the alpha value indicates the rating of this behavior (solid for 'Strongly agree', and half-transparent for 'Somewhat agree'). If a user did not agree to any of the behaviors or indicated multiple contradicting behaviors, the data point is marked as undefined in green.

Moreover, there is no clear evidences pointing to correlation between the level of awareness the user rated and Principal Component 2 or 3.
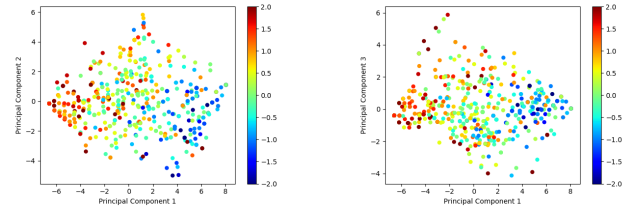

Fig. 6. Principal Component Analysis results for {PC1, PC2} (left) and {PC1, PC3} (right). The color of the data point indicates the average amount of awareness the user rated on a 5 point scale (-2 for not paying any attention at all, and 2 for paying a lot of attention).

Therefore, we consider the Principal Component 1 as a safety score reflecting the amount of attention awareness that a driver or an autonomous navigation system should take into account. TDBM is therefore computed as below:

$$\mathbf{s}_{TDBM} = \begin{pmatrix} -4.78 & -7.89 & 2.24 & 1.69 & -0.83 & 4.69 \end{pmatrix} \begin{pmatrix} s_{center} \\ v_{nei} \\ s_{front} \\ v_{avg} \\ j_l \\ 1 \end{pmatrix} \quad (7)$$

V. NAVIGATION

In this section, we highlight the benefits of identifying driver behaviors and how these ensure safe navigation. We extend an autonomous car navigation algorithm, AutonoVi [7], and show improvements in its performance by using our driver behavior identification algorithm and TDBM. AutonoVi is based on a data-driven vehicle dynamics model and optimization-based maneuver planning, which generates a set of favorable trajectories from among a set of possible candidates, and performs selection among this set of trajectories using optimization. It can handle dynamic lane-changes and different traffic conditions.

The approach used in AutonoVi is summarized below: The algorithm takes a graph of roads from a GIS database, and

applies A* algorithm to compute the shortest global route plan. The route plan consists of a sequence of actions that is composed of {Drive Straight, Turn Left, Turn Right, Merge Left, and Merge Right}. The plan is translated to a static guiding path that consists of a set of way-points, that exhibits C^1 continuity, and that takes Traffic Rules into account (e.g., making a stop at an intersection). AutonoVi then samples the steering angle and velocity in a desirable range of values to compute a set of candidate trajectories, and eliminates the trajectories that lead to possible collisions based on Control Obstacles [46].

Among the set of collision-free trajectories, AutonoVi selects the best trajectory by optimizing a heuristic that penalizes trajectories that lead to: i) deviation from global route; ii) sharp turns, braking and acceleration; iii) unnecessary lane changes; and iv) getting too close to other vehicles and objects (even without a collision).

To avoid getting too close to other neighboring entities, AutonoVi proposed a proximity cost function to differentiate entities only by its class. That is, it considers all vehicles as the same and applies the same penalization factor, $F_{vehicle}$, to them. Further, it applies a higher factor : F_{ped} and F_{cyc} to pedestrians and cyclist respectively. The original proximity cost used in AutonoVi is:

$$c_{prox} = \sum_{n=1}^{N} F_{vehicle}\, e^{-d(n)} \qquad (8)$$

This cost function has two issues: i) it cannot distinguish dangerous drivers to avoid driving too close to them, and ii) it diminishes too rapidly due to its use of an exponential function. We propose a novel proximity cost that can solve these problems:

$$c'_{prox} = \sum_{n=1}^{N} c(n) \qquad (9)$$

$$c(n) = \begin{cases} 0 & \text{if } d \in [d_{t2}, \inf), \\ S_{TDBM} B_{far}\, \frac{d_{t2}-d(n)}{d_{t2}} & \text{if } d \in (d_t, d_{t2}], \\ S_{TDBM}\left[\frac{(d_t-d(n))(B_r-B_{far})}{d_t} + B_{far} \right] & \text{if } d \in (0, d_t]. \end{cases} \qquad (10)$$

where $d(n)$ is the distance between the car navigating with our approach and the neighbor n, d_t is a threshold distance beyond which neighbors are considered as far away, and d_{t2} is a threshold distance beyond which neighbors would no longer have impact on our navigation. S_{TDBM} is derived from Equation 7, B_{far} and B_r are the attention metrics computed using the features extracted from the features using the mapping in equation 6, for $r = \{back, front, adj\}$ if the neighboring car is following, preceding, and next to the navigating car, respectively.

Using this new cost function, we can avoid drivers that are potentially riskier, and select a safer navigation path. Examples of scenarios are illustrated in Figure 7 and the attached video.

(a)

(b)

(c)

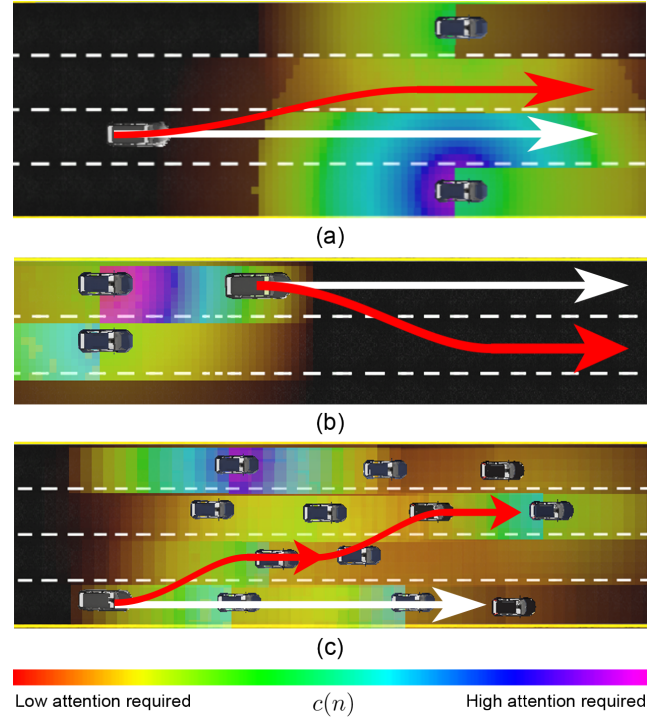Low attention required $c(n)$ High attention required

Fig. 7. Examples of our approach making better navigation decision than AutonoVi. The red route is the one selected by our approach while white is the one selected by AutonoVi. The cost map $c(n)$ is also shown for each neighbor car n indicating the amount of attention needed. In (a), our algorithm chooses to switch lane and keep a distance from the car require more attention. In (b), a car requiring high level of attention tailgates the car running our approach, and we switch to a slower lane to give way. In (c), a heavy traffic ahead causing all four lanes move at a similarly low speed, and our algorithm chooses to the follow the car with the lowest attention required.

VI. CONCLUSION AND FUTURE WORKS

We present a novel data-driven approach to enable safer real-time navigation by identifying human drivers who are potentially hazardous. Our studies and findings are based on a data-driven mapping computation (TDBM). We conclude that although humans use different adjectives when describing driving behavior, there is an underlying latent variable, S_{TDBM} (Equation 7), that reflects the level of attention humans pay to other vehicles' driving behavior. Moreover, we can estimate this variable by a set of novel trajectory features and other existing features.

Current trajectory data tends to be limited due to human labeling or the fact that extra efforts may be needed to extract such annotated data from raw images. With advancement in object detection and other work in computer vision, one can expect more trajectory data would be made available to the autonomous driving research community. Given more data in urban environments (e.g. status of traffic signal, pedestrian/cyclist trajectories, and etc.) , we would like to apply our approach to analyzing and developing different navigation strategies that adapt to these new situations and local driving styles. Preliminary results of extending TDBM to take pedestrians and cyclists into account has been described in a recent work [47].

REFERENCES

[1] G. A. M. Meiring and H. C. Myburgh, "A review of intelligent driving style analysis systems and related artificial intelligence algorithms," *Sensors*, vol. 15, no. 12, pp. 30 653–30 682, 2015.

[2] Z.-X. Feng, J. Liu, Y.-Y. Li, and W.-H. Zhang, "Selected model and sensitivity analysis of aggressive driving behavior," *Zhongguo Gonglu Xuebao(China Journal of Highway and Transport)*, vol. 25, no. 2, pp. 106–112, 2012.

[3] S. J. Guy, S. Kim, M. C. Lin, and D. Manocha, "Simulating heterogeneous crowd behaviors using personality trait theory," in *Proceedings of the ACM SIGGRAPH/Eurographics symposium on computer animation*, 2011, pp. 43–52.

[4] A. Bera, T. Randhavane, and D. Manocha, "Aggressive, tense, or shy? identifying personality traits from crowd videos," in *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, 2017.

[5] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[6] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang *et al.*, "End to end learning for self-driving cars," *arXiv:1604.07316*, 2016.

[7] A. Best, S. Narang, L. Pasqualin, D. Barber, and D. Manocha, "Autonovi: Autonomous vehicle planning with dynamic maneuvers and traffic constraints," *arXiv:1703.08561*, 2017.

[8] A. Aljaafreh, N. Alshabatat, and M. S. N. Al-Din, "Driving style recognition using fuzzy logic," in *Vehicular Electronics and Safety (ICVES), 2012 IEEE International Conference on*, pp. 460–463.

[9] B. Krahé and I. Fenske, "Predicting aggressive driving behavior: The role of macho personality, age, and power of car," *Aggressive Behavior*, vol. 28, no. 1, pp. 21–29, 2002.

[10] K. H. Beck, B. Ali, and S. B. Daughters, "Distress tolerance as a predictor of risky and aggressive driving," *Traffic injury prevention*, vol. 15, no. 4, pp. 349–354, 2014.

[11] J. C. Brill, M. Mouloua, E. Shirkey, and P. Alberti, "Predictive validity of the aggressive driver behavior questionnaire (adbq) in a simulated environment," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 53, no. 18. SAGE Publications Sage CA: Los Angeles, CA, 2009, pp. 1334–1337.

[12] Y. L. Murphey, R. Milton, and L. Kiliaris, "Driver's style classification using jerk analysis," in *Computational Intelligence in Vehicles and Vehicular Systems, 2009. CIVVS'09. IEEE Workshop on*, pp. 23–28.

[13] I. Mohamad, M. A. M. Ali, and M. Ismail, "Abnormal driving detection using real time global positioning system data," in *Space Science and Communication (IconSpace), 2011 IEEE International Conference on*, pp. 1–6.

[14] R. Aufrere, R. Chapuis, and F. Chausse, "A model-driven approach for real-time road recognition," *Machine Vision and Applications*, vol. 13, no. 2, pp. 95–107, 2001.

[15] G. Qi, Y. Du, J. Wu, and M. Xu, "Leveraging longitudinal driving behaviour data with data mining techniques for driving style analysis," *IET intelligent transport systems*, vol. 9, no. 8, pp. 792–801, 2015.

[16] B. Shi, L. Xu, J. Hu, Y. Tang, H. Jiang, W. Meng, and H. Liu, "Evaluating driving styles by normalizing driving behavior based on personalized driver modeling," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 12, pp. 1502–1508, 2015.

[17] W. Wang, J. Xi, A. Chong, and L. Li, "Driving style classification using a semisupervised support vector machine," *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 5, pp. 650–660, 2017.

[18] E. Cheung, T. K. Wong, A. Bera, X. Wang, and D. Manocha, "Lcrowdv: Generating labeled videos for simulation-based crowd behavior learning," in *European Conference on Computer Vision*. Springer, 2016, pp. 709–727.

[19] E. Cheung, A. Wong, A. Bera, and D. Manocha, "Mixedpeds: Pedestrian detection in unannotated videos using synthetically generated human-agents for training," *AAAI*, 2018.

[20] D. Sadigh, K. Driggs-Campbell, A. Puggelli, W. Li, V. Shia, R. Bajcsy, A. L. Sangiovanni-Vincentelli, S. S. Sastry, and S. A. Seshia, "Data-driven probabilistic modeling and verification of human driver behavior," 2014.

[21] "Ionroad app that makes driving safer." [Online]. Available: http://www.ionroad.com/

[22] C.-W. You, N. D. Lane, F. Chen, R. Wang, Z. Chen, T. J. Bao, M. Montes-de Oca, Y. Cheng, M. Lin, L. Torresani *et al.*, "Carsafe app: Alerting drowsy and distracted drivers using dual cameras on smartphones," in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 2013, pp. 13–26.

[23] 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 *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*. IEEE, 2014, pp. 240–245.

[24] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Physical review E*, vol. 62, no. 2, p. 1805, 2000.

[25] A. Kesting, M. Treiber, and D. Helbing, "General lane-changing model mobil for car-following models," *Journal of the Transportation Research Board*, no. 1999, pp. 86–94, 2007.

[26] C. F. Choudhury, "Modeling lane-changing behavior in presence of exclusive lanes," Ph.D. dissertation, Massachusetts Institute of Technology, 2005.

[27] C. Katrakazas, M. Quddus, W.-H. Chen, and L. Deka, "Real-time motion planning methods for autonomous on-road driving: State-of-the-art and future research directions," *Transportation Research Part C: Emerging Technologies*, vol. 60, pp. 416–442, 2015.

[28] M. Saifuzzaman and Z. Zheng, "Incorporating human-factors in car-following models: a review of recent developments and research needs," *Transportation research part C: emerging technologies*, vol. 48, pp. 379–403, 2014.

[29] S. Kolski, D. Ferguson, M. Bellino, and R. Siegwart, "Autonomous driving in structured and unstructured environments," in *Intelligent Vehicles Symposium*. IEEE, 2006, pp. 558–563.

[30] G. M. Hoffmann, C. J. Tomlin, M. Montemerlo, and S. Thrun, "Autonomous automobile trajectory tracking for off-road driving: Controller design, experimental validation and racing," in *American Control Conference ACC'07*, 2007, pp. 2296–2301.

[31] Y. Ma, D. Manocha, and W. Wang, "Autorvo: Local navigation with dynamic constraints in dense heterogeneous traffic," in *arXiv report: 1804.02915*, 2018.

[32] Y. Luo, P. Cai, A. Bera, D. Hsu, W. Lee, and D. Manocha, "Autonomous driving among many pedestrians: Models and algorithms," in *IEEE Robotics Automation Letters (RA-L)*, 2018.

[33] J. Ziegler, P. Bender, M. Schreiber, H. Lategahn, T. Strauss, C. Stiller, T. Dang, U. Franke, N. Appenrodt, C. G. Keller *et al.*, "Making bertha drive—an autonomous journey on a historic route," *IEEE Intelligent Transportation Systems Magazine*, vol. 6, no. 2, pp. 8–20, 2014.

[34] M. Buehler, K. Iagnemma, and S. Singh, *The DARPA urban challenge: autonomous vehicles in city traffic*. springer, 2009, vol. 56.

[35] A. Geiger, M. Lauer, F. Moosmann, B. Ranft, H. Rapp, C. Stiller, and J. Ziegler, "Team annieway's entry to the 2011 grand cooperative driving challenge," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1008–1017, 2012.

[36] R. Kianfar, B. Augusto, A. Ebadighajari, U. Hakeem, J. Nilsson, A. Raza, R. S. Tabar, N. V. Irukulapati, C. Englund, P. Falcone *et al.*, "Design and experimental validation of a cooperative driving system in the grand cooperative driving challenge," *IEEE transactions on intelligent transportation systems*, vol. 13, no. 3, pp. 994–1007, 2012.

[37] C. Englund, L. Chen, J. Ploeg, E. Semsar-Kazerooni, A. Voronov, H. H. Bengtsson, and J. Didoff, "The grand cooperative driving challenge 2016: boosting the introduction of cooperative automated vehicles," *IEEE Wireless Communications*, vol. 23, 2016.

[38] D. Sadigh, S. Sastry, S. A. Seshia, and A. D. Dragan, "Planning for autonomous cars that leverage effects on human actions." in *Robotics: Science and Systems*, 2016.

[39] D. Sadigh, S. S. Sastry, S. A. Seshia, and A. Dragan, "Information gathering actions over human internal state," in *Intelligent Robots and Systems (IROS), 2016 IEEE International Conference on*, pp. 66–73.

[40] S. H. Huang, D. Held, P. Abbeel, and A. D. Dragan, "Enabling robots to communicate their objectives," *arXiv:1702.03465*, 2017.

[41] A. D. D. Dorsa Sadigh, S. Sastry, and S. A. Seshia, "Active preference-based learning of reward functions," in *Robotics: Science and Systems (RSS)*, 2017.

[42] P. B. Harris, J. M. Houston, J. A. Vazquez, J. A. Smither, A. Harms, J. A. Dahlke, and D. A. Sachau, "The prosocial and aggressive driving inventory (padi): A self-report measure of safe and unsafe driving behaviors," *Accident Analysis & Prevention*, vol. 72, no. Supplement C, pp. 1 – 8, 2014.

[43] M. Lan, M. Rofouei, S. Soatto, and M. Sarrafzadeh, "Smartldws: A robust and scalable lane departure warning system for the smartphones," in *Intelligent Transportation Systems, 2009. ITSC'09. 12th International IEEE Conference on*, pp. 1–6.

[44] J. Halkia and J. Colyar, "Interstate 80 freeway dataset," *Federal Highway Administration, U.S. Department of Transportation*, 2006.

[45] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society (Methodological)*, pp. 267–288, 1996.

[46] D. Bareiss and J. van den Berg, "Generalized reciprocal collision avoidance," *The International Journal of Robotics Research*, vol. 34, no. 12, pp. 1501–1514, 2015.

[47] E. Cheung, A. Bera, and D. Manocha, "Efficient and safe vehicle navigation based on driver behavior classification," in *WAD2018, CVPRW*.