



UNIVERSITY OF LEEDS

This is a repository copy of *Gaze-based Intention Anticipation over Driving Manoeuvres in Semi-Autonomous Vehicles*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/151186/>

Version: Accepted Version

Proceedings Paper:

Wu, M, Louw, T orcid.org/0000-0001-6577-6369, Lahijanian, M et al. (4 more authors) (2020) *Gaze-based Intention Anticipation over Driving Manoeuvres in Semi-Autonomous Vehicles*. In: *Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2019)*. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2019), 03-08 Nov 2019, Macau, China. IEEE , pp. 6210-6216. ISBN 978-1-7281-4004-9

<https://doi.org/10.1109/IROS40897.2019.8967779>

© 2019 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. Uploaded in accordance with the publisher's self-archiving policy.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Gaze-based Intention Anticipation over Driving Manoeuvres in Semi-Autonomous Vehicles

Min Wu*, Tyron Louw[†], Morteza Lahijanian*, Wenjie Ruan*, Xiaowei Huang[‡],
Natasha Merat[†], and Marta Kwiatkowska*

*Department of Computer Science, University of Oxford, OX1 3QD, UK

{Min.Wu, Morteza.Lahijanian, Wenjie.Ruan, Marta.Kwiatkowska}@cs.ox.ac.uk

[†]Institute for Transport Studies, University of Leeds, LS2 9JT, UK

{T.L.Louw, N.Merat}@its.leeds.ac.uk

[‡]Department of Computer Science, University of Liverpool, L69 3BX, UK

Xiaowei.Huang@liverpool.ac.uk

Abstract—Anticipating a human collaborator’s intention enables a safe and efficient interaction between a human and an autonomous system. Specifically in the context of semi-autonomous driving, studies have revealed that correct and timely prediction of driver’s intention needs to be an essential part of Advanced Driver Assistance System (ADAS) design. To this end, we propose a framework that exploits drivers’ time-series gaze and fixation patterns to anticipate their real-time intention over possible future actions, enabling a smart and collaborative ADAS that can aid drivers to overcome safety-critical situations. The method models human intention as the latent states of a hidden Markov model and uses probabilistic dynamic time warping distributions to capture the temporal characteristics of the observation patterns of the drivers. The method is evaluated on a data set collected in safety-critical semi-autonomous driving experiments. The results illustrate the efficacy of the framework, which correctly anticipated drivers’ intentions about 3 seconds beforehand with over 90% accuracy.

I. INTRODUCTION

The technology for *fully-autonomous* cars is rapidly improving, but they are still far away from reality. *Semi-autonomous* driving, though, is already here. Cars with Advanced Driver Assistance Systems (ADAS) that provide limited autonomous capabilities are currently available and attracting a lot of attention. Examples include Tesla’s Autopilot and Ford’s Co-Pilot 360. These systems are designed to ensure safety by alerting hazardous traffic conditions or even taking over control to avert impending collisions. Recent accidents, however, have revealed major safety issues with ADAS such as late warning and wrong intervention. These issues are mainly caused by lack of accounting for the human driver’s mental state, specifically, intentions in the design of ADAS. In fact, it is crucial to anticipate the driver’s intentions in order to be able to *safely* assist the driver in critical situations. Our goal is to address this important challenge and design an ADAS that can anticipate and take into account drivers’ intentions. In this work,

we focus on intention prediction in critical situations and propose a method of anticipating a driver’s intended action via analysing the driver’s observation of the surrounding environment, and specifically, eye gaze.

Recent studies [1]–[3] show the importance of human intention prediction in the context of semi-autonomous driving and ADAS design. They explain that it is necessary to detect the driver intentions as early as possible to ensure that information, warnings, and especially system interventions by ADAS do not come into conflict with the driver intentions. Otherwise, conflicting situations, where for instance the intervention of the ADAS can interfere with the driver’s intention of operation, can arise and jeopardise the safety of the driver and the surrounding vehicles. Hence, correct and timely prediction of driver’s intention needs to be an essential part of ADAS design.

The concept of human *intention* can be defined as a commitment to the execution of a particular action [4]. Whilst intention recognition can be achieved by utilising a person’s physical status and/or the system’s measurements, e.g., steering data after a driver has already started to manoeuvre [5]–[7], intention *anticipation* is more challenging as it is achieved before the actual movement. Recent works [1], [2] showed that by relying on multiple sources including inside-vehicle features, e.g., facial points and head motion, together with outside-vehicle features, e.g., vehicle dynamics, road conditions, street maps, it is possible to compute the probability of different future driving manoeuvres with high accuracy. In safety critical situations, however, all these sources of data may not be available. In such cases, a method that relies on an easily accessible feature is preferred.

Gaze has been identified as a revelation of human intention by indicating the direction of attention and future actions [8], [9]. In human-robot collaboration, it has been shown that human gaze can be utilised to interpret human’s intention [10]–[14]. For example, in a collaborative task [15], gaze features are used to predict the participants’ intended requests. Similarly, in shared autonomy [16], user’s gaze is used to estimate the goals of the user. Gaze information is also utilised in driving scenarios to understand the driver’s

*Kwiatkowska, Lahijanian, and Ruan are supported by EPSRC Programme Grant on Mobile Autonomy (EP/M019918/1). Huang is supported by the DSTL project “Test Coverage Metrics for Artificial Intelligence”. Wu is supported by the CSC-PAG Oxford Scholarship.

distraction [17], [18].

Our goal is to design an ADAS that can predict driver intentions and provide safety assistance accordingly in critical situations. As the first step towards this goal, we focus on human intention anticipation solely based on gaze in safety-critical driving situations since it is a reliable source in such cases. In other words, we are interested in utilising real-time gaze observations to anticipate driver’s intention indicated by subsequent actions in an autonomous driving scenario. This is an important yet challenging problem. On one hand, gaze cues, which include head pose implicitly [19], can discriminate between adjacent zones such as front wind-screen and speedometer by subtle eye movements [17]. On the other hand, it is difficult to efficiently use gaze because recorded gaze data may potentially contain noise from sensors, and the temporal dependencies of gaze sequence should be considered. More importantly, individual drivers can exhibit different gaze patterns, and therefore, analysis of the similarity between different gaze patterns under certain actions is necessary.

In this work, we propose a probabilistic Dynamic Time Warping - Hidden Markov Model (pDTW-HMM) architecture to anticipate intention over future manoeuvres based on the gaze pattern. We model human intention as the latent states of an HMM and use gaze sequence as the observations of the states of the HMM. We employ recursive Bayesian estimation to iteratively infer real-time intention. Within this framework, we use DTW to capture the temporal characteristics of the gaze pattern and construct a pDTW distribution to reflect the similarity of gaze patterns under distinct manoeuvres. Finally, we combine these two aspects together by importing the pDTW distribution into the measurement likelihood during the update procedure of inferring the latent states.

The main contribution of this work is the *first* framework for driver’s intention anticipation over driving manoeuvres that relies solely on gaze pattern to the best of our knowledge. Another novelty of the work is probabilistic extension of DTW and applying it to the domain of gaze pattern recognition. Finally, the evaluation of the framework on a driving data set with 124 cases from 75 drivers, collected in a safety critical semi-autonomous driving scenario when the drivers were supposed to take certain manoeuvres to avoid collision. We demonstrate that our approach anticipates intention 3.64 seconds before a real action was carried out with 93.5% accuracy.

A. Related Work

Some previous works focused on lane change recognition based on various data sources. For example, Kuge et al. [5] developed a Hidden Markov Model (HMM) using steering behaviour to recognise emergency and normal lane change as well as lane keeping. Kim et al. [3] employed a Support Vector Machine (SVM) predictor fed with on-board sensor measures along with vehicle state and road condition learnt from an Artificial Neural Network (ANN) to detect a driver’s intention of lane change with high accuracy.

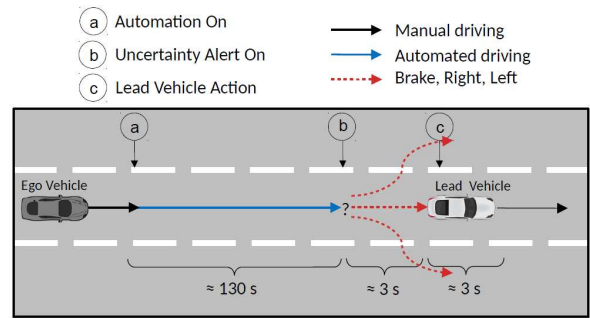


Fig. 1: Schematic representation of the driving scenario.

Later on, researchers tried to predict lane change behaviours slightly beforehand. Salvucci et al. [20] demonstrated that a cognitive model - Adaptive Control of Thought-Rational (ACT-R) - could detect intention of lane change, achieving 90% accuracy within 1 second, using steering-wheel angle, accelerator depression, along with environmental data (lateral position and time headway). Kumar et al. [21] combined SVM and Bayesian filter together, i.e., Relevance Vector Machine (RVM), to predict lane change 1.3 seconds in advance, using lane trajectory from a tracker.

Recently, anticipating future driving manoeuvres a few seconds before has been studied. In particular, Jain et al. [1] proposed an Autoregressive Input-Output HMM (AIO-HMM), which captures context from inside and outside the vehicle, to anticipate manoeuvres 3.5 seconds beforehand with over 80% F1-score. Moreover, they also used Recurrent Neural Networks (RNN) combined with Long Short-Term Memory (LSTM) to anticipate manoeuvres, increasing precision to 90.5% [2].

Further, gaze has been studied to reveal intention in different contexts. For instance, in a collaborative sandwich-making task, Huang et al. [15] developed a SVM based model solely using gaze features to predict the participants’ intended requests of ingredients. In an autonomous driving scenario, Jiang et al. [19] proposed a Dynamic Interest Point Detection (DIPD) methodology, which combines a dynamic random Markov field with an energy function, to infer driver’s points of interest (e.g., shop signs) using gaze tracking data.

II. PROBLEM STATEMENT

In this section, we explain the considered safety-critical driving scenario and formulate the intention anticipation problem.

A. Driving Scenario

We consider a driving scenario given in Fig. 1, where a semi-autonomous vehicle is following a lead vehicle in a highway at 70 mph while in autonomous mode. Suddenly, the semi-autonomous vehicle detects a swift deceleration of 5 m/s^2 of the lead vehicle, at which point (time instant (b)) it sends out an “uncertainty alert” to the driver to take control. The driver has about 3 s to react to the safety-critical situation

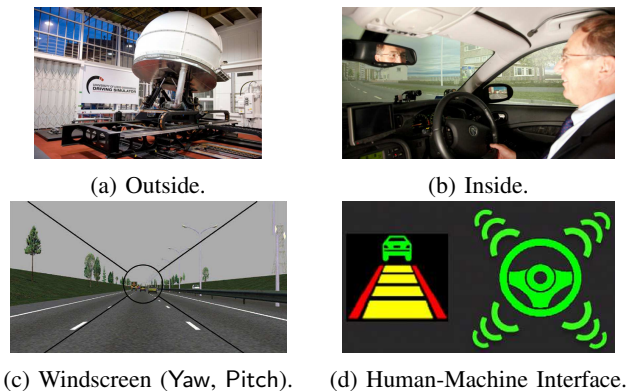


Fig. 2: The Jaguar S-Type Driving Simulator.

to avoid collision, e.g., Brake in time, turn Right or Left to another lane.

We have in fact collected data for this study as part of the EU-funded AdaptIVe project [22]–[24]. The experiments were performed at the University of Leeds Driving Simulator, presented in Fig. 2, which consists of a Jaguar S-Type cab within a 4 m spherical projection dome (Fig. 2(a) and 2(b)), with a 300° field-of-view projection system over two dimensions Yaw (horizontal) and Pitch (vertical) as a windscreen (Fig. 2(c)). Drivers’ eye movements were recorded by a v4.5 Seeing Machines faceLAB eye-tracker at 60 Hz. When in safety-critical condition, the Automation Status symbol (Fig. 2(d)) flashes yellow, acting as an “uncertainty alert”, to invite driver’s intervention to deactivate automation.

B. Problem Formulation

We are interested in anticipating the driver’s intention at each time step by analysing the observation data from time instant (b) until the moment the driver takes a manoeuvre. The definition of intention is in Definition 1, and we formalise the problem in Problem 1.

Definition 1: Given a set of driving manoeuvres \mathcal{M} , a driver’s *intention* is a probability distribution P over \mathcal{M} such that $\sum_{\mathcal{I} \in \mathcal{M}} P(\mathcal{I}) = 1$. We let $\arg \max_{\mathcal{I} \in \mathcal{M}} P(\mathcal{I})$ be the *intended manoeuvre*.

Whereas \mathcal{M} may include many possible driving manoeuvres, for illustration purpose, we focus on three manoeuvres Brake, Right, and Left that suit the scenario.

Problem 1: A driver’s time-series observation, or observation history, is a sequence $\mathbb{O}_T = (o_1, \dots, o_T)$, where $o_t = (\text{Yaw}_t, \text{Pitch}_t)$ for $1 \leq t \leq T$ is an observation point on the Yaw-Pitch plane. Given a prefix $\mathbb{O}_t = (o_1, \dots, o_t)$ of the observation history, a real-time history dependent *intention strategy* δ on time t is a conditional probability $P(\mathcal{I}_t | \mathbb{O}_t)$ such that $\sum_{\mathcal{I}_t \in \mathcal{M}} P(\mathcal{I}_t | \mathbb{O}_t) = 1$. Then, given \mathbb{O}_T , the intention anticipation problem is to find an intention strategy δ to minimise the safety risk.

III. GAZE-BASED INTENTION ANTICIPATION FRAMEWORK

To approach Problem 1, we design a framework that uses HMMs to model human intention, pDTW to capture

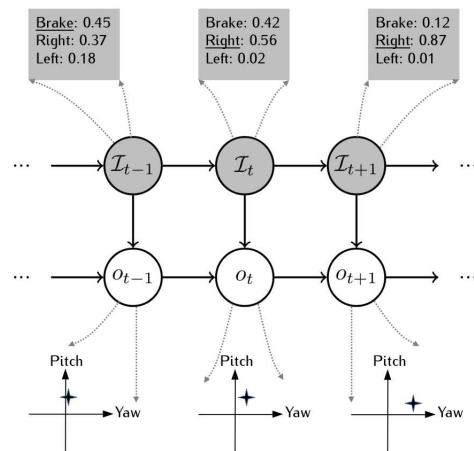


Fig. 3: The HMM graphical model representing real-time history dependent intention, where $\mathcal{I}_t \in \mathcal{M}$ denotes an intended manoeuvre in a latent state, and o_t an observation point in an observed state.

observation pattern, and Bayesian estimation to compute intention strategy.

A. Modelling Intention with HMM

For each driver, an HMM is constructed representing real-time history dependent intention over driving manoeuvres, as exhibited in Fig. 3. At each time step t , a driver’s intention is a probability distribution over manoeuvres \mathcal{M} . We exploit recursive Bayesian estimation [25] to compute $P(\mathcal{I}_t | \mathbb{O}_t)$. It comprises two steps, Prediction and Update, which are explained below.

Prediction: Given a sequence of time-series historical observations $\mathbb{O}_{t-1} = (o_1, \dots, o_{t-1})$, we predict manoeuvre at the next time step \mathcal{I}_t by

$$P(\mathcal{I}_t | \mathbb{O}_{t-1}) = \int P(\mathcal{I}_t | \mathcal{I}_{t-1}) \cdot P(\mathcal{I}_{t-1} | \mathbb{O}_{t-1}) d\mathcal{I}_{t-1}. \quad (1)$$

We assume that, when a driver’s observation is available up to time instant $t-1$, the driver’s intention remains unchanged from $t-1$ to t until a new observation point o_t comes in. That is, when \mathbb{O}_{t-1} is available but o_t is not yet, we have $\mathcal{I}_t = \mathcal{I}_{t-1}$, which implies $P(\mathcal{I}_t | \mathcal{I}_{t-1}) = 1$. Intuitively, since driver’s gaze was recorded at 60 Hz, i.e., every 1/60 s, we assume the driver’s intention did not change until a new gaze point was recorded.

Update: The update of the intention when a new observation point o_t comes, i.e., from \mathbb{O}_{t-1} to \mathbb{O}_t , is

$$P(\mathcal{I}_t | \mathbb{O}_t) = \frac{P(\mathcal{I}_t | \mathbb{O}_{t-1}) \cdot P(o_t | \mathcal{I}_t, \mathbb{O}_{t-1})}{P(o_t | \mathbb{O}_{t-1})}, \quad (2)$$

where $P(\mathcal{I}_t | \mathbb{O}_{t-1})$ is the predicted intention from Equation (1), and $P(o_t | \mathcal{I}_t, \mathbb{O}_{t-1})$ is the measurement likelihood. The latter intuitively means that an observation point is dependent on the current intention and historical observations, shown as the emission probabilities in Fig. 3.

Combining these two steps together, the value of $P(\mathcal{I}_t | \mathbb{O}_t)$ can be computed via Lemma 1.

Lemma 1: Given a driver's time-series observation $\mathbb{O}_T = (o_1, \dots, o_T)$, through modelling intention as an HMM, the driver's real-time history dependent intention strategy δ over a possible manoeuvre $\mathcal{I}_t \in \mathcal{M}$ can be computed by

$$P(\mathcal{I}_t | \mathbb{O}_t) = \frac{P(\mathcal{I}_0) \prod_{i=1}^t P(o_i | \mathcal{I}_i, \mathbb{O}_{i-1})}{\prod_{i=1}^t P(o_i | \mathbb{O}_{i-1})}, \quad (3)$$

where $P(\mathcal{I}_0)$ is the prior distribution. As $\sum_{\mathcal{I}_t \in \mathcal{M}} P(\mathcal{I}_t | \mathbb{O}_t) = 1$, the denominator acts as a normalisation constant thus does not need to be calculated.

Proof: By combining Equations (1) and (2) recursively, when $P(\mathcal{I}_t | \mathbb{O}_t) = 1$, we have $P(\mathcal{I}_t | \mathbb{O}_t)$

$$= \frac{P(\mathcal{I}_{t-1} | \mathbb{O}_{t-1}) \cdot P(o_t | \mathcal{I}_t, \mathbb{O}_{t-1})}{P(o_t | \mathbb{O}_{t-1})} \quad (4)$$

$$= \frac{P(\mathcal{I}_{t-2} | \mathbb{O}_{t-2}) \cdot P(o_{t-1} | \mathcal{I}_{t-1}, \mathbb{O}_{t-2}) \cdot P(o_t | \mathcal{I}_t, \mathbb{O}_{t-1})}{P(o_{t-1} | \mathbb{O}_{t-2}) \cdot P(o_t | \mathbb{O}_{t-1})} \quad (5)$$

$$= \dots \quad (6)$$

$$= \frac{P(\mathcal{I}_0 | \mathbb{O}_0) \prod_{i=1}^t P(o_i | \mathcal{I}_i, \mathbb{O}_{i-1})}{\prod_{i=1}^t P(o_i | \mathbb{O}_{i-1})}, \quad (7)$$

where $P(\mathcal{I}_0 | \mathbb{O}_0) = P(\mathcal{I}_0)$ when there is no observation. ■

Therefore, the problem of intention anticipation based on observation is reduced to the construction of the measurement likelihood $P(o_t | \mathcal{I}_t, \mathbb{O}_{t-1})$, which essentially captures temporal characteristics of observation patterns under distinct driving manoeuvres.

B. Capturing Observation Pattern with pDTW

Dynamic time warping (DTW) [26] measures similarity between two time-dependent sequences via finding an optimal alignment under certain restrictions, and has been applied in speech pattern comparison in automatic speech recognition [27], as well as information retrieval for music and motion [28].

In this work, we extend DTW to probabilistic DTW, or pDTW, to capture driver's observation pattern and fit that into the HMM model to anticipate intention. To be more specific, for a new driver whose intentions are to be predicted, at each time step t , we compute a distance measure $\text{DTW}_{\mathcal{M}}^t$ from a set of experimental drivers whose observation sequences and ultimate manoeuvres have been recorded in Section IV, then extract a probability distribution over the distance measure $p\text{DTW}_{\mathcal{M}}^t$ and let $P(o_t | \mathcal{I}_t, \mathbb{O}_{t-1})$ be the conditional probability of observing o_t under the condition of taking the manoeuvre \mathcal{I}_t .

We first introduce DTW distance below.

Definition 2: Given two time-dependent sequences $X = (x_1, \dots, x_M)$ and $Y = (y_1, \dots, y_N)$ of respective lengths $M, N \in \mathbb{N}^+$, a *warping path* is a sequence $p = (p_1, \dots, p_L)$ such that $p_l = (m_l, n_l) \in [1, M] \times [1, N]$ for $l \in [1, L]$ subject to constraints:

- 1) Boundary condition: $p_1 = (1, 1)$ and $p_L = (M, N)$.
- 2) Continuity: $p_{l+1} - p_l \in \{(1, 1), (1, 0), (0, 1)\}$ for $l \in [1, L - 1]$.

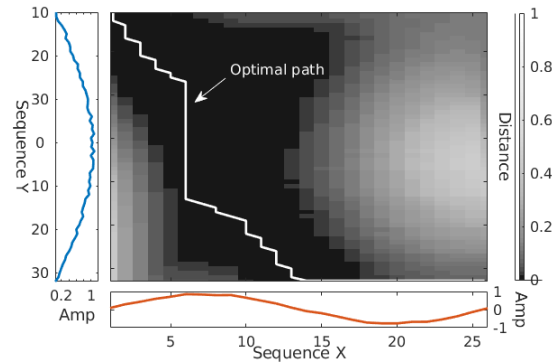


Fig. 4: Optimal alignment of two time-dependent sequences via DTW. Darker area denotes shorter distance in the accumulated distance matrix.

3) Monotonicity: $m_1 \leq \dots \leq m_L$ and $n_1 \leq \dots \leq n_L$.

Let \mathcal{F} be a feature space such that $x_m, y_n \in \mathcal{F}$ for $m \in [1, M]$, $n \in [1, N]$, and $d: \mathcal{F} \times \mathcal{F} \mapsto \mathbb{R}_{\geq 0}$ be the *local distance*, then the *total distance* $d_p(X, Y)$ of a warping path p is $d_p(X, Y) = \sum_{l=1}^L d(x_{m_l}, y_{n_l})$. *DTW distance*, denoted by $\text{DTW}(X, Y)$, is the minimal total distance among all possible warping paths \mathcal{P} . That is,

$$\text{DTW}(X, Y) = \min_{p \in \mathcal{P}} d_p(X, Y). \quad (8)$$

In this work, $\text{DTW}(X, Y)$ computes the minimal Euclidean distance of X and Y , each of which denotes an observation sequence. See graphical illustration of the optimal warping path of two sequences of scalars in Fig. 4.

The construction of a minimal DTW distance measure is in Definition 3.

Definition 3: Given a set of experimental drivers \mathcal{D}_{total} , in which each $D = (\mathbb{O}_{\mathcal{I}}, \mathcal{I})$, $\mathcal{I} \in \mathcal{M}$ denotes that every driver has a recorded observation sequence $\mathbb{O}_{\mathcal{I}}^T$ and a corresponding manoeuvre \mathcal{I} . Let \mathcal{D}_{new} denote a new driver with observations $\mathbb{O}_T = (o_1, \dots, o_T)$, then a *minimal DTW distance measure* w.r.t manoeuvres \mathcal{M} at time step t , denoted by $\text{DTW}_{\mathcal{M}}^t$, is defined as

$$\text{DTW}_{\mathcal{M}}^t = \mathbf{d}_{\mathcal{I} \in \mathcal{M}}^t \quad (9)$$

such that in the vector each

$$d_{\mathcal{I} \in \mathcal{M}}^t = \min\{\text{DTW}(\mathbb{O}_t, \mathbb{O}_{\mathcal{I}}^T) \mid \mathbb{O}_{\mathcal{I}}^T \in \mathcal{D}_{total}\} \quad (10)$$

where $\mathbb{O}_t \preceq \mathbb{O}_T$ is a prefix.

Essentially, the minimal DTW distance $\text{DTW}_{\mathcal{M}}^t$ is a similarity measure that discovers the closest observation patterns between a new observation sequence \mathbb{O}_t and observations in each manoeuvre category. Note that there is a *negative correlation* between the distance value $d_{\mathcal{I}}^t$ and probability $P(\mathcal{I}_t | \mathbb{O}_t)$. Intuitively, if $d_{\mathcal{I}}^t$ increases, then the observation pattern \mathbb{O}_t is less similar to these labelled with manoeuvre \mathcal{I} , i.e., the driver \mathcal{D}_{new} is less likely to take manoeuvre \mathcal{I} , thus $P(\mathcal{I}_t | \mathbb{O}_t)$ decreases.

Now we introduce how to extract a probability distribution over the distance measure, taking the above-mentioned negative correlation into consideration.

Definition 4: Given a new driver D_{new} with observations $\mathbb{O}_T = (o_1, \dots, o_T)$, and this driver's minimal DTW distance measure $\text{DTW}_{\mathcal{M}}^t = \mathbf{d}_{\mathcal{I} \in \mathcal{M}}^t$, $t \in [1, T]$, let $r_{\mathcal{I}}^t$ be the reward of choosing manoeuvre \mathcal{I} , and $c_{\mathcal{I}}^t$ be the cost, then the reward $R^t(\mathcal{M})$ is defined as vector

$$R^t(\mathcal{M}) = \mathbf{r}_{\mathcal{I} \in \mathcal{M}}^t = \mathbf{c}_{-\mathcal{I} \in \mathcal{M}}^t = \sum_{\mathcal{I}' \in \mathcal{M} \setminus \mathcal{I}} \mathbf{c}_{\mathcal{I}'}^t \quad (11)$$

where $c_{\mathcal{I}}^t = \frac{d_{\mathcal{I}}^t}{\sum_{\mathcal{I}' \in \mathcal{M}} d_{\mathcal{I}'}^t}$. Subsequently, by using softmax, the probability distribution over minimal DTW distance measure, denoted by $p\text{DTW}_{\mathcal{M}}^t$, is

$$p\text{DTW}_{\mathcal{M}}^t = \frac{\exp(R^t(\mathcal{I})/\mathcal{T})}{\sum_{\mathcal{I}' \in \mathcal{M}} \exp(R^t(\mathcal{I}')/\mathcal{T})}, \quad (12)$$

where temperature \mathcal{T} is a real constant.

In the above definition, we define reward $r_{\mathcal{I}}^t$ and cost $c_{\mathcal{I}}^t$ to reflect the negative correlation. Intuitively, the reward of carrying out an intended manoeuvre \mathcal{I} is the cost of not taking \mathcal{I} , i.e., $-\mathcal{I}$, which is the sum of the costs of choosing the other manoeuvres. Therefore, if a driver's observation pattern \mathbb{O}_t is closer to these labelled as manoeuvre \mathcal{I} , then the minimal DTW distance $d_{\mathcal{I}}^t$ is smaller, the cost $c_{\mathcal{I}}^t$ is smaller, the reward $r_{\mathcal{I}}^t$ is comparatively greater, and the probability value $p\text{DTW}_{\mathcal{I}}^t$ is greater, i.e., the driver is more likely to carry out this manoeuvre.

C. pDTW-HMM Intention Anticipation

Now we combine the pDTW distribution, which essentially reflects the characteristics of the observation patterns under particular manoeuvres (Section III-B), as the emission distribution in the Update procedure of the previous HMM modelling intention (Section III-A), so that a driver's intention is predicted and updated at each time step whenever a new observation point comes.

The proposed pDTW-HMM framework is presented in Algorithm 1. Here we assume an uninformative uniform prior over the driving manoeuvres.

Algorithm 1 Intention Anticipation

- 1: **Input:** A set of possible driving manoeuvres \mathcal{M} ;
 - 2: A set of experimental drivers D_{total} .
 - 3: **Output:** Intention strategy δ . (Problem 1)
 - 4: **procedure** pDTW-HMM
 - 5: Initialise prior distribution $P(\mathcal{I}_0)$;
 - 6: **for** $t = 1$: a manoeuvre is taken **do**
 - 7: Record observation point o_t ;
 - 8: Compute $p\text{DTW}_{\mathcal{M}}^t$; (Definition 4)
 - 9: Let $P(o_t | \mathcal{I}_t, \mathbb{O}_{t-1}) = p\text{DTW}_{\mathcal{I} \in \mathcal{M}}^t$;
 - 10: Infer and normalise $P(\mathcal{I}_t | \mathbb{O}_t)$; (Lemma 1)
 - 11: Send $\text{dist}(\mathcal{M})$ to ADAS;
 - 12: $t = t + 1$.
 - 13: **end for**
-

Remark. In this work, we let observation denote both a driver's *gaze* and *fixation*, as each of which can be regarded

TABLE I: Participant demographics. Following ethical approval from the University's Research Ethics Committee (Reference Number: LTTRAN-054), 5 groups of 15 drivers (75 in total, 41 male, 34 female) were recruited.

Age (years)	Driving License (years)	Annual Mileage (miles)
36.16 ± 12.38	16.22 ± 12.92	8290.46 ± 6723.08

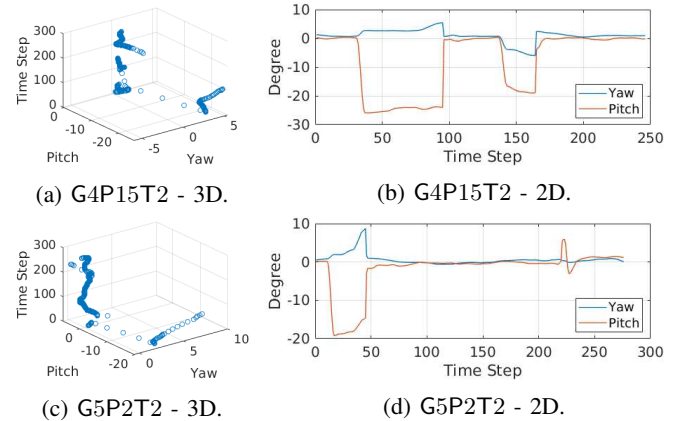


Fig. 5: Illustration of driver's gaze pattern. Left: gaze points in Yaw-Pitch-Time space; Right: separation of Yaw and Pitch degrees on time steps.

as an aspect of observation, while the former normally contains noise and the latter performs as a filtration. Note that the proposed framework works for both as shown in the experimental results.

IV. EXPERIMENTAL RESULTS

This section presents the experimental results of anticipating driver's real-time intention over future manoeuvres based on past observations.

We have 124 valid experimental cases from two trials of the scenario - 61 in Trial 1, and 63 in Trial 2. The demographics of participants are given in Table I. We use $G_xP_yT_z$ to mark Participant y of Group x in Trial z , where $x \in \{1..5\}$, $y \in \{1..15\}$, $z \in \{1..2\}$.

A. Gaze and Fixation

We consider two forms of observations - *gaze* and *fixations* - to anticipate driver's real-time intention over possible manoeuvres. (Min: add eye-tracking algorithm citation) See examples of gaze patterns in Fig. 5.

We define fixation as a driver's gaze maintaining on a fixed area for a certain period of time, e.g., 0.2s. Fig. 6 illustrates fixation extraction from a driver's gaze sequence. Take G2P9T1 as an example, from Fig. 6(a) we observe that 27 fixations were formed from a sequence of 411 gaze points. Although a few gaze points somewhat scattered to the right, all fixations were formed at the centre region of the windscreen, i.e., $[-1^\circ: 1^\circ, 0^\circ: 2^\circ]$, where the lead vehicle was decelerating or probably stopping ahead. This corresponds to the situation where manoeuvre Brake is taken

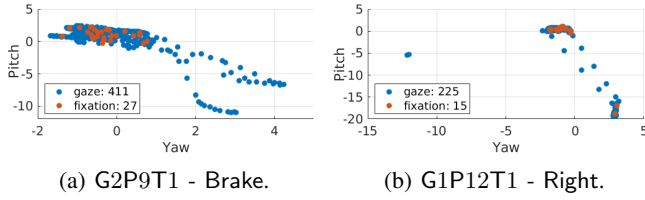


Fig. 6: Extraction of fixations from a sequence of gaze points. Plot on Yaw-Pitch plane for illustration of where the driver was looking at on the windscreen. (frequency $\nu = 60$ Hz, duration $\Delta = 0.2$ s, fixation range $f = 2^\circ$.)

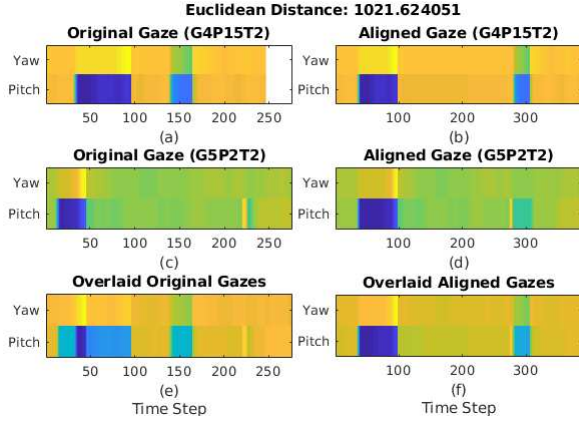


Fig. 7: Comparison of gaze patterns using DTW. Top row: original (a) and aligned (b) gaze sequences of G4P15T2; Middle row: that of G5P2T2; Bottom row: overlaid gaze sequences of both.

as the driver needed to focus on the conditions ahead in order to brake in time to avoid collision.

B. Intention Anticipation over Driving Manoeuvres

Now we present intention anticipation over driving manoeuvres from both gaze and fixation.

Comparison between gaze or fixation patterns is achieved by generating a pDTW distribution (Section III-B). Fig. 7 describes the computation of the Euclidean distance between gaze sequences of two arbitrary drivers. The capability of DTW to capture the gaze pattern is shown via the close match between G2P15T2's Original Gaze (Fig. 7(a)) and 2D plot (Fig. 5(b)), as well as G5P2T2's Original Gaze (Fig. 7(c)) and 2D plot (Fig. 5(d)). Intuitively, once an optimal alignment between the gaze patterns is found, as shown in the Overlaid Aligned Gazes (Fig. 7(f)), the shortest warping distance can be computed. In this case, $DTW(G4P15T2, G5P2T2) = 1021.62$.

A driver's real-time intention strategy generated from gaze sequence and extracted fixations for the same duration is illustrated in Fig. 8 for comparison. We remark the discrepancies between intention anticipation from gaze and fixation. On one hand, the advantage of inferring from gaze points directly is that the strategy can be obtained at each time step almost simultaneously while the gaze point being recorded. Nevertheless, the disadvantage is that the strategy

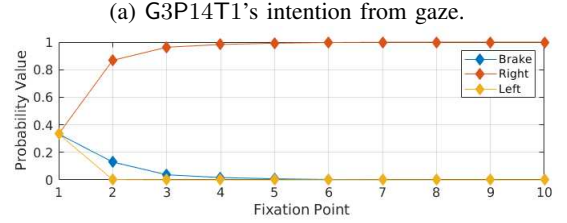
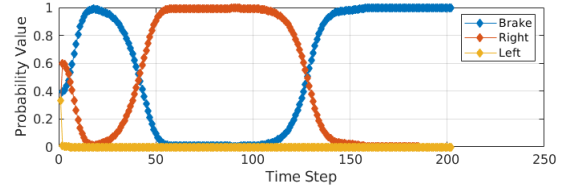


Fig. 8: Real-time intention strategy over three driving manoeuvres Brake, Right, and Left from gaze (a) and fixation (b), respectively. (Leave-one-out cross-validation, temperature $\mathcal{T} = 1/10$.)

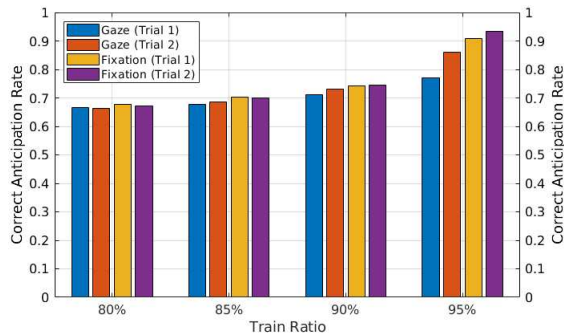
may change drastically at some time instants, and thus exhibits instability. On the other hand, if extracting fixation points before inferring intention, the strategy tends to be more robust, i.e., fewer or no drastic reversals, though in this case the strategy is only available at each fixation point, i.e., every 0.2 s when fixation forms.

C. Accuracy Validation

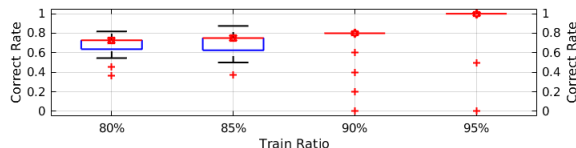
We validate the proposed framework through statistically analysing how accurate the anticipation is, by comparing the predicted manoeuvre to the real manoeuvre that was taken by individual drivers.

For both trials, we separate the total number of drivers randomly into a training set and a test set. Formally, the separation algorithm is as follows. A total set $D_{total}[1, N]$ is classified into three manoeuvre categories $D_{Brake}[1, B]$, $D_{Right}[1, R]$, and $D_{Left}[1, L]$, such that $B + R + L = N$ and $N, B, R, L \in \mathbb{N}^+$. Let $\gamma \in (0, 1)$ be a train ratio, then a training set is $D_{train}[1, \alpha] = \gamma * D_{Brake} \cup \gamma * D_{Right} \cup \gamma * D_{Left}$, where $\alpha = \lceil \gamma B \rceil + \lceil \gamma R \rceil + \lceil \gamma L \rceil$, $*$ denotes random selection, and $\lceil x \rceil$ retrieves the nearest integer greater than or equal to x . A test set is the complement $D_{test}[1, \beta] = D_{total} \setminus D_{train}$ such that $\beta = N - \alpha$.

The general accuracy of intention anticipation from gaze and fixation in both trials is illustrated in Fig. 9. In terms of mean accuracy (Fig. 9(a)), as the train ratio increases from 85% to 95%, the correct anticipation rate increases. Fig. 9(b) describes the box plot of the anticipation accuracy from fixation in Trial 2, corresponding to the purple bars in Fig. 9(a). It shows that, after 500 iterations of the experiment to potentially enlarge the training and test sets, when train ratio is 95%, the correct rate almost reaches 1.0. We believe that the outliers are due to the size of the test set (small in each iteration), and do not compromise the overall result as the mean value is 93.5%.



(a) Mean accuracy from gaze and fixation after 500 iterations.



(b) Box plot of accuracy from fixation in Trial 2, corresponding to the purple bars in Fig. 9(a).

Fig. 9: Accuracy of intention anticipation from gaze and fixation, respectively, in Trial 1 and Trial 2. ($\gamma = 80\% \sim 95\%$, $\mathcal{T} = 1/10$, $\nu = 60$ Hz, $\Delta = 0.2$ s, $f = 2^\circ$.)

The proposed framework pDTW-HMM’s advantage over baseline methods is presented in Table II. We observe that it achieves higher mean accuracy rate (Pr) and longer correct anticipation time before actual manoeuvre (t_b). Specifically, it exhibits that intention anticipation from fixation outperforms that from gaze, e.g., 13.8% higher in Trial 1 and 7.5% higher in Trial 2. Furthermore, it demonstrates that intention anticipation in Trial 2 is more accurate than that in Trial 1, regardless of gaze (9% higher) or fixation (2.7% higher). We also evaluate t_b , which means the time duration when the correctly predicted manoeuvre remains unchanged until a driver starts to take actual action, e.g., $128^{th} \sim 202^{nd}$ time steps in Fig. 8(a), and $2^{nd} \sim 10^{th}$ fixation points in Fig. 8(b). In experiments, t_b is formatted into seconds. Trial 2 shows slightly shorter durations than Trial 1, e.g., 0.18 s shorter from fixation. Combining with the fact that, in Trial 2, none of the drivers (0/63) crashed whilst 15/61 crashes occurred in Trial 1, the reason may be that, after familiarising themselves with the safety-critical scenario and operation of the autonomous vehicle, the drivers were able to be more focused on the environment, thus forming more

TABLE II: Comparison of two baseline methods HMM and DTW with our pDTW-HMM framework, in terms of mean accuracy rate (Pr) and correct anticipation time before actual manoeuvre (t_b). ($\gamma = 95\%$, $\mathcal{T} = 1/10$, $\nu = 60$ Hz, $\Delta = 0.2$ s, $f = 2^\circ$.)

Methodology	Trial 1				Trial 2			
	Gaze		Fixation		Gaze		Fixation	
	Pr	t_b	Pr	t_b	Pr	t_b	Pr	t_b
HMM	73.30%	3.40 ± 1.58 s	72.00%	2.92 ± 1.70 s	56.60%	2.89 ± 1.40 s	47.30%	2.91 ± 1.43 s
DTW	75.00%	2.91 ± 1.71 s	79.50%	2.69 ± 1.37 s	70.10%	2.72 ± 1.47 s	78.00%	2.73 ± 1.41 s
pDTW-HMM	77.00%	3.71 ± 1.85 s	90.80%	3.82 ± 1.27 s	86.00%	3.67 ± 1.32 s	93.50%	3.64 ± 1.09 s

reasonable (i.e., less distracted), not necessarily faster but more precautionous, gaze and fixations patterns. This makes intention easier to anticipate, and eventually leads to successful collision avoidance.

V. CONCLUSION

In this paper, we propose a pDTW-HMM framework, by analysing the gaze and fixation patterns of the driver, especially taking the temporal characteristics of them into consideration. We show in our experiments that it can anticipate a driver’s real-time intention over future manoeuvres around 3 s beforehand with over 90% accuracy. Future work aims to perform strategy synthesis or adaptation to assist drivers in safety-critical situations when resuming control of the vehicle from the autonomous mode. We also intend to study the influence of driver distraction on the predictions.

REFERENCES

- [1] A. Jain, H. S. Koppula, B. Raghavan, S. Soh, and A. Saxena, "Car that knows before you do: Anticipating maneuvers via learning temporal driving models," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 3182–3190.
- [2] A. Jain, A. Singh, H. S. Koppula, S. Soh, and A. Saxena, "Recurrent neural networks for driver activity anticipation via sensory-fusion architecture," in *Robotics and Automation (ICRA), 2016 IEEE International Conference on*. IEEE, 2016, pp. 3118–3125.
- [3] I.-H. Kim, J.-H. Bong, J. Park, and S. Park, "Prediction of drivers intention of lane change by augmenting sensor information using machine learning techniques," *Sensors*, vol. 17, no. 6, p. 1350, 2017.
- [4] B. F. Malle and J. Knobe, "The folk concept of intentionality," *Journal of Experimental Social Psychology*, vol. 33, no. 2, pp. 101–121, 1997.
- [5] N. Kuge, T. Yamamura, O. Shimoyama, and A. Liu, "A driver behavior recognition method based on a driver model framework," SAE Technical Paper, Tech. Rep., 2000.
- [6] L. Jin, H. Hou, and Y. Jiang, "Driver intention recognition based on continuous hidden markov model," in *Transportation, Mechanical, and Electrical Engineering (TMEE), 2011 International Conference on*. IEEE, 2011, pp. 739–742.
- [7] A. Doshi and M. M. Trivedi, "On the roles of eye gaze and head dynamics in predicting driver's intent to change lanes," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 3, pp. 453–462, 2009.
- [8] S. Baron-Cohen, S. Wheelwright, J. Hill, Y. Raste, and I. Plumb, "The reading the mind in the eyes test revised version: a study with normal adults, and adults with asperger syndrome or high-functioning autism," *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, vol. 42, no. 2, pp. 241–251, 2001.
- [9] A. N. Meltzoff and R. Brooks, "Like me as a building block for understanding other minds: Bodily acts, attention, and intention," *Intentions and intentionality: Foundations of social cognition*, vol. 171191, 2001.
- [10] W. Yi and D. Ballard, "Recognizing behavior in hand-eye coordination patterns," *International Journal of Humanoid Robotics*, vol. 6, no. 03, pp. 337–359, 2009.
- [11] Y. Razin and K. M. Feigh, "Learning to predict intent from gaze during robotic hand-eye coordination," in *AAAI*, 2017, pp. 4596–4602.
- [12] H. Ravichandar, A. Kumar, and A. Dani, "Bayesian human intention inference through multiple model filtering with gaze-based priors," in *Information Fusion (FUSION), 2016 19th International Conference on*. IEEE, 2016, pp. 2296–2302.
- [13] C.-M. Huang and B. Mutlu, "Anticipatory robot control for efficient human-robot collaboration," in *Human-Robot Interaction (HRI), 2016 11th ACM/IEEE International Conference on*. IEEE, 2016, pp. 83–90.
- [14] D. Gehrig, P. Krauthausen, L. Rybok, H. Kuehne, U. D. Hanebeck, T. Schultz, and R. Stiefelhagen, "Combined intention, activity, and motion recognition for a humanoid household robot," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*. IEEE, 2011, pp. 4819–4825.
- [15] C.-M. Huang, S. Andrist, A. Saup্পé, and B. Mutlu, "Using gaze patterns to predict task intent in collaboration," *Frontiers in psychology*, vol. 6, 2015.
- [16] H. Admoni and S. Srinivasa, "Predicting user intent through eye gaze for shared autonomy," in *2016 AAAI Fall Symposium Series*, 2016.
- [17] S. Vora, A. Rangesh, and M. M. Trivedi, "On generalizing driver gaze zone estimation using convolutional neural networks," in *Intelligent Vehicles Symposium (IV), 2017 IEEE*. IEEE, 2017, pp. 849–854.
- [18] I.-H. Choi, S. K. Hong, and Y.-G. Kim, "Real-time categorization of driver's gaze zone using the deep learning techniques," in *Big Data and Smart Computing (BigComp), 2016 International Conference on*. IEEE, 2016, pp. 143–148.
- [19] Y.-S. Jiang, G. Warnell, and P. Stone, "Dipd: Gaze-based intention inference in dynamic environments," 2018.
- [20] D. D. Salvucci, "Inferring driver intent: A case study in lane-change detection," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 48, no. 19. SAGE Publications Sage CA: Los Angeles, CA, 2004, pp. 2228–2231.
- [21] P. Kumar, M. Perrollaz, S. Lefevre, and C. Laugier, "Learning-based approach for online lane change intention prediction," in *Intelligent Vehicles Symposium (IV), 2013 IEEE*. IEEE, 2013, pp. 797–802.
- [22] T. Louw and N. Merat, "Are you in the loop? using gaze dispersion to understand driver visual attention during vehicle automation," *Transportation Research Part C: Emerging Technologies*, vol. 76, pp. 35–50, 2017.
- [23] T. Louw, R. Madigan, O. Carsten, and N. Merat, "Were they in the loop during automated driving? links between visual attention and crash potential," *Injury prevention*, vol. 23, no. 4, pp. 281–286, 2017.
- [24] T. Louw, G. Markkula, E. Boer, R. Madigan, O. Carsten, and N. Merat, "Coming back into the loop: Drivers perceptual-motor performance in critical events after automated driving," *Accident Analysis & Prevention*, vol. 108, pp. 9–18, 2017.
- [25] S. Särkkä, *Bayesian filtering and smoothing*. Cambridge University Press, 2013, vol. 3.
- [26] J. Kruskal and M. Liberman, "The symmetric time-warping problem: From continuous to discrete," 1983.
- [27] L. R. Rabiner and B.-H. Juang, *Fundamentals of speech recognition*. PTR Prentice Hall Englewood Cliffs, 1993, vol. 14.
- [28] M. Müller, *Information retrieval for music and motion*. Springer, 2007, vol. 2.