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

Min Wu[1], Tyron Louw[2], Morteza Lahijanian[3], Wenjie Ruan[1], Xiaowei Huang[4],
Natasha Merat[2], and Marta Kwiatkowska[1]

Abstract— Anticipating a human collaborator's intention enables 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 the 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 eye gaze and fixation patterns to anticipate their real-time intention over possible future manoeuvres, 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 of 124 experiments from 75 drivers collected in a safety-critical semi-autonomous driving scenario. The results illustrate the efficacy of the framework by correctly anticipating the 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 (ADASs) 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 ADASs such as late warning and wrong intervention. These issues are mainly caused by the lack of accounting for the human driver's mental state, specifically, intentions [1] in the design of ADASs. In fact, it is crucial to anticipate drivers' intentions in order to be able to *safely* assist them 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 safety-critical situations (Fig. 1), and propose a method of anticipating a driver's intended action

[1] Department of Computer Science, University of Oxford, UK {Min.Wu; Wenjie.Ruan; Marta.Kwiatkowska}@cs.ox.ac.uk
[2] Institute for Transport Studies, University of Leeds, UK {T.L.Louw; N.Merat}@its.leeds.ac.uk
[3] Dept. of Aerospace Engineering Sciences, University of Colorado, Boulder, CO, USA morteza.lahijanian@colorado.edu
[4] Department of Computer Science, University of Liverpool, UK Xiaowei.Huang@liverpool.ac.uk
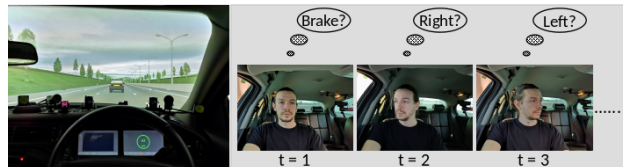
Fig. 1: A driver's intention in a safety-critical scenario.

via analysing the driver's observation of the surrounding environment, and specifically, eye gaze.

Recent studies [2]–[4] demonstrate the importance of human intention prediction in the context of semi-autonomous driving and ADAS design. They explain that it is necessary to detect drivers' intentions as early as possible to ensure that information, warnings, and especially system interventions by ADASs do not come into conflict with drivers' intentions. Otherwise, conflicting situations can arise and jeopardise the safety of the driver and the surrounding vehicles. For instance, where the intervention of the ADAS can interfere with a driver's intention of operation. Hence, correct and timely prediction of drivers' intentions needs to be an essential part of ADAS design.

The concept of a driver's *intention* can be defined as a commitment to the execution of a particular action [1]. While 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 [2], [3] showed that by relying on multiple data 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.

Eye 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 distraction [17], [18]. Nevertheless, eye gaze has never been

used *solely* to predict the driver's intention.

Our goal is to design an ADAS that can predict drivers' 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 the 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 windscreen and speedometer by subtle eye movements [17]. On the other hand, it is difficult to efficiently use gaze because (i) recorded gaze data may potentially contain noise from sensors, (ii) there are temporal dependencies in a gaze sequence, and more importantly, (iii) individual drivers can exhibit different gaze patterns.

In this work, we propose a probabilistic Dynamic Time Warping - Hidden Markov Model (pDTW-HMM) architecture to anticipate intention over future manoeuvres based on drivers' observations, including both recorded raw *gaze* and extracted *fixation* as a filtration. We model human intention as the latent states of an HMM and use gaze or fixation 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 observation pattern and construct a pDTW distribution to reflect the similarity of observation patterns under distinct manoeuvres. Finally, we combine these two aspects 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 or fixation pattern to the best of our knowledge. Another novelty of the work is the probabilistic extension of DTW and applying it to the domain of observation pattern recognition. Finally, the evaluation of the framework is performed on a driving data set with 124 cases from 75 drivers, collected in a safety-critical semi-autonomous driving scenario. We demonstrate that our approach anticipates intention around 3 seconds before a real manoeuvre was carried out with over 90% accuracy.

II. RELATED WORK

Intention prediction is primarily a problem of matching the actions and the goals, for which there are two approaches: generative vs. descriptive [20]–[24]. Within the generative approach (from goals to actions), various architectures have been proposed to encode the causes that can produce the observed actions, such as Demiris et al.'s HAMMER [21], [22] and Wolpert et al.'s MOSAIC [23], whereas within the descriptive approach (from actions to goals), one can derive the goals from patterns of actions, e.g., Csibra et al.'s action-effects associations [20] and Hommel et al.'s theory of event coding [24]. This work, along with the related
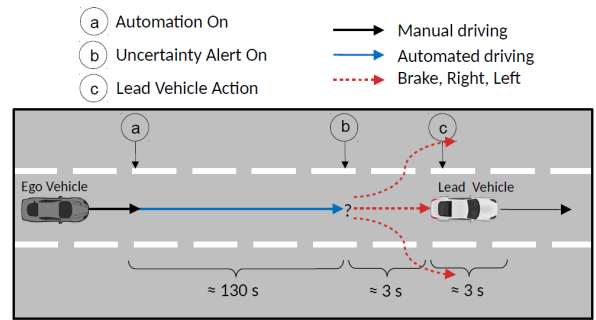

Fig. 2: Schematic representation of the driving scenario.

works mentioned below, falls into this category by implicitly assuming that understanding the intentions of humans is a classification problem, and driver's intentions can be inferred from their observable driving manoeuvres.

In the context of semi-autonomous driving, some previous works focused on lane change recognition based on various data sources [4]–[7]. Later on, researchers endeavoured to anticipate driving manoeuvres slightly beforehand from multi-modal sensory cues [25]–[29]. For instance, Kumar et al. [25] combined Support Vector Machine and Bayesian filter into a Relevance Vector Machine to predict lane change $1.3\,\text{s}$ in advance by using a lane trajectory tracker. In particular, Trivedi et al. [26]–[28] performed lane change prediction by concatenating sensor-rich features from inside and outside of a vehicle. Salvucci et al. [29] proposed a cognitive model, Adaptive Control of Thought-Rational, to detect a lane change by achieving 90% accuracy within $1\,\text{s}$ from using steering-wheel angle, accelerator depression, and environmental data. Moreover, apart from lane change, Jain et al. [2], [3] predicted other driving manoeuvres such as turns. Specifically, they proposed an Autoregressive Input-Output HMM as well as a framework that combines Recurrent Neural Networks with Long Short-Term Memory units to anticipate manoeuvres $3.5\,\text{s}$ beforehand with precision 80% to 90.5%. Nevertheless, in safety-critical scenarios, not all these sensory cues may be available or processed in time. Thus we evaluate an easily accessible feature, i.e., eye gaze.

Meanwhile, gaze has been studied to reveal intention in human-robot collaboration [10]–[14] and semi-autonomous driving [17]–[19]. In particular, in a collaborative sandwich-making task, Huang et al. [15] developed an 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 methodology, which combines a dynamic random Markov field with an energy function, to use gaze to infer driver's points of interest, e.g., shop signs.

III. PROBLEM STATEMENT

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

A. Driving Scenario

We consider the driving scenario depicted in Fig. 2, where a semi-autonomous vehicle is following a lead vehicle in a
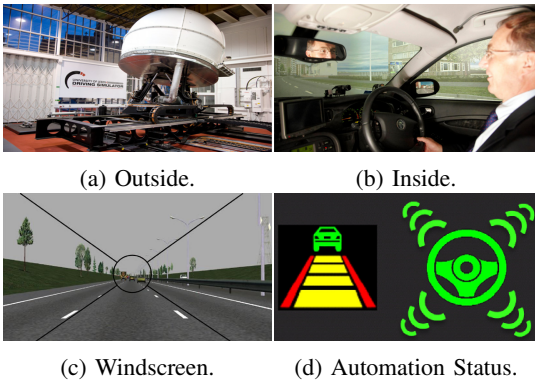
(a) Outside.   (b) Inside.

(c) Windscreen.   (d) Automation Status.

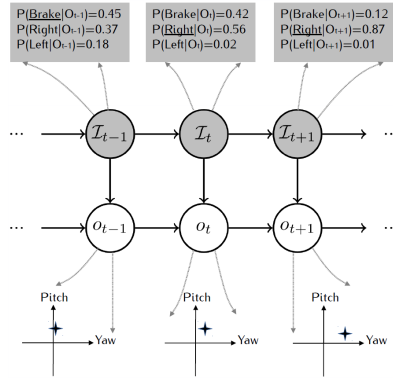Fig. 3: The University of Leeds Driving Simulator.
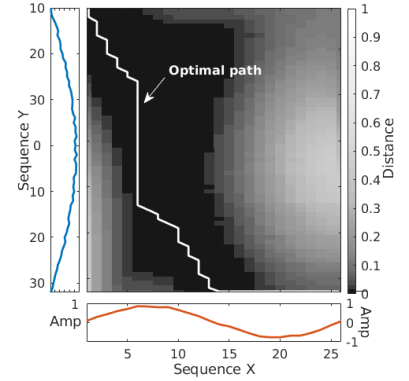

Fig. 4: The graphical HMM model.

Fig. 5: DTW optimal alignment.

highway at $70\,\mathrm{mph}$ while in autonomous mode. Suddenly, the semi-autonomous vehicle detects a swift deceleration of $5\,\mathrm{m/s^2}$ of the lead vehicle, at which point it sends out an uncertainty alert to the driver to take control. The driver has about $3\,\mathrm{s}$ to react to the safety-critical situation and avoid collision by, e.g., braking or turning to the right or left lane.

Data for this study was collected in the high-fidelity, motion-based, University of Leeds Driving Simulator (Fig. 3), as part of the EU-funded AdaptIVe project [30]–[32]. The simulator consists of a Jaguar S-Type cab within a $4\,\mathrm{m}$ spherical projection dome, a $300°$ field-of-view projection system over two dimensions - Yaw (horizontal) and Pitch (vertical) - as a windscreen. Drivers' eye movements were recorded by a v4.5 Seeing Machines faceLAB eyetracker at $60\,\mathrm{Hz}$. It uses infrared cameras to detect both eye and head position to estimate gaze points, thus the luminance of the environment and whether the drivers are wearing glasses should not compromise the accuracy of gaze tracking. When in safety-critical condition, the Automation Status symbol (Fig. 3(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) in Fig. 2 until the driver performs a manoeuvre. To formulate this problem, we first define intention.

Definition 1 (Intention): Given a finite 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 refer to $\arg\max_{\mathcal{I}\in\mathcal{M}} P(\mathcal{I})$ as the *intended manoeuvre*.

Whilst the set $\mathcal{M}$ may include many possible driving manoeuvres, for such safety-critical situations, we primarily focus on three imminent manoeuvres of braking (Brake), turning to the right (Right) or left (Left) lane to avoid collision, i.e., $\mathcal{M} = \{\mathsf{Brake}, \mathsf{Right}, \mathsf{Left}\}$.

A driver's time-series observation, or observation history, is given as a sequence $\mathbb{O}_T = (o_1, \ldots, o_T)$, where $o_t = (\mathsf{Yaw}_t, \mathsf{Pitch}_t)$ for $1 \le t \le T$ is an observation point on the Yaw-Pitch plane, which can be a recorded *gaze* point or an extracted *fixation* point.

Definition 2 (Intention Strategy): Given a prefix $\mathbb{O}_t = (o_1, \ldots, o_t)$ of the observation history $\mathbb{O}_T$, a real-time history-

tory dependent *intention strategy* δ at time t is a conditional probability $P(\mathcal{I}_t \mid \mathbb{O}_t)$ such that $\sum_{\mathcal{I}_t \in \mathcal{M}} P(\mathcal{I}_t \mid \mathbb{O}_t) = 1$.

Note that δ estimates the driver's intention based on the observations up to time t. Therefore, our goal is to find δ with high accuracy seconds before T and send it to ADAS.

Problem 1 (Intention Anticipation): Given a driver's time-series observation $\mathbb{O}_t$, find the intention strategy δ of the driver at each time step $t \in [1, T]$.

IV. GAZE-BASED INTENTION ANTICIPATION FRAMEWORK

To approach Problem 1, we design a framework that uses HMMs to model human intention, pDTW to capture observation pattern, and Bayesian estimation to compute intention strategy.

A. Modelling Intention with HMM

HMMs are widely used to model temporal variations of human intention and activities [2], [33]–[35]. As exhibited in Fig. 4, an HMM is constructed representing real-time history dependent intention over driving manoeuvres $\mathcal{M}$, 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. At each time step t, a driver's intention is a probability distribution over $\mathcal{M}$. We exploit recursive Bayesian estimation [36] to compute $P(\mathcal{I}_t \mid \mathbb{O}_t)$. It comprises two steps: Prediction and Update.

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

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

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 $P(\mathcal{I}_t \mid \mathcal{I}_{t-1}) = 1$. Intuitively, since driver's gaze was recorded at $60\,\mathrm{Hz}$, i.e., every $1/60\,\mathrm{s}$, we assume the driver's intention does not change until a new gaze point is recorded.

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

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

where $P(\mathcal{I}_t \mid \mathbb{O}_{t-1})$ is the predicted intention from Equation (1), and $P(o_t \mid \mathcal{I}_t, \mathbb{O}_{t-1})$ is the measurement likelihood. The latter intuitively means that an observation point o_t is dependent on the current intention $\mathcal{I}_t$ and historical observations $\mathbb{O}_{t-1}$, shown as the emission probabilities in Fig. 4.

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

Lemma 1: Given a driver's time-series observation $\mathbb{O}_T = (o_1, \ldots, o_T)$, 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

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

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

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

B. Capturing Observation Pattern with pDTW

Dynamic time warping (DTW) [37] measures the similarity between two time-dependent sequences via finding an optimal alignment under certain restrictions, and has been applied widely [38]–[40], for example, in automatic speech recognition [39] and information retrieval for music and motion [40].

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. We first introduce DTW distance below, as illustrated in Fig. 5.

Definition 3 (DTW Distance): Given two time-dependent sequences $X = (x_1, \ldots, x_M)$ and $Y = (y_1, \ldots, y_N)$ of respective lengths $M, N \in \mathbb{N}^+$, a *warping path* is a sequence $p = (p_1, \ldots, p_L)$ 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 $l \in [1, L-1]$.
3) Monotonicity: $m_1 \leq \ldots \leq m_L$ and $n_1 \leq \ldots \leq n_L$.

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 $\mathrm{DTW}(X, Y)$, is the *minimal* total distance among all possible warping paths $\mathcal{P}$. That is, $\mathrm{DTW}(X, Y) = \min_{p \in \mathcal{P}} d_p(X, Y)$. In this work, local distance $d(x, y)$ is the Euclidean distance between points x and y, and $\mathrm{DTW}(X, Y)$ computes the minimal Euclidean distance of X and Y, each of which denotes a driver's observation sequence (Fig. 8).

The construction of a minimal DTW distance measure is shown in Definition 4.

Definition 4 (Minimal DTW Distance): Given a set of experimental drivers D_{total}, in which each $\mathsf{D} = (\mathbb{O}_T^{\mathcal{I}}, \mathcal{I})$, $\mathcal{I} \in \mathcal{M}$ denotes that every driver has a recorded observation

sequence $\mathbb{O}_T^{\mathcal{I}}$ and a corresponding manoeuvre $\mathcal{I}$. Let D_{new} denote a new driver with observations $\mathbb{O}_T = (o_1, \ldots, o_T)$, then a *minimal DTW distance measure* w.r.t manoeuvres $\mathcal{M}$ at time step t, denoted by $\mathrm{DTW}_{\mathcal{M}}^t$, is defined as a vector $\mathrm{DTW}_{\mathcal{M}}^t = \mathbf{d}_{\mathcal{M}}^{\mathbf{t}}$ such that for each entry $d_{\mathcal{I} \in \mathcal{M}}^t$ of $\mathbf{d}_{\mathcal{M}}^{\mathbf{t}}$,

$$d_{\mathcal{I}}^t = \min\{\mathrm{DTW}(\mathbb{O}_t, \mathbb{O}_T^{\mathcal{I}}) \mid \mathbb{O}_T^{\mathcal{I}} \in \mathsf{D}_{total}\}, \qquad (4)$$

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

Essentially, the minimal DTW distance $\mathrm{DTW}_{\mathcal{M}}^t$ is a similarity measurement 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 $P(\mathcal{I}_t \mid \mathbb{O}_t)$.

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

Definition 5 (pDTW): Given a new driver D_{new} with observations $\mathbb{O}_T = (o_1, \ldots, o_T)$ and minimal DTW distance measure $\mathrm{DTW}_{\mathcal{M}}^t = \mathbf{d}_{\mathcal{M}}^{\mathbf{t}}$, $t \in [1, T]$, let $r_{\mathcal{I}}^t$ be the reward of choosing manoeuvre $\mathcal{I}$, and $c_{\mathcal{I}}^t$, $c_{\neg\mathcal{I}}^t$ be the cost of choosing and not choosing $\mathcal{I}$, respectively. Then the reward $R^t(\mathcal{M})$ is defined as a vector $R^t(\mathcal{M}) = \mathbf{r}_{\mathcal{I} \in \mathcal{M}}^{\mathbf{t}} = \mathbf{c}_{\neg\mathcal{I} \in \mathcal{M}}^{\mathbf{t}} =$

$$\sum_{\mathcal{I}' \in \mathcal{M} \setminus \mathcal{I}} \mathbf{c}_{\mathcal{I}'}^{\mathbf{t}} \text{ where } c_{\mathcal{I}}^t = \frac{d_{\mathcal{I}}^t}{\sum_{\mathcal{I} \in \mathcal{M}} d_{\mathcal{I}}^t}. \qquad (5)$$

Subsequently, by using softmax, the *probability distribution over minimal DTW distance measure* $p\mathrm{DTW}_{\mathcal{M}}^t$ is

$$p\mathrm{DTW}_{\mathcal{M}}^t = \frac{\exp(R^t(\mathcal{I})/\mathcal{T})}{\sum_{\mathcal{I} \in \mathcal{M}} \exp(R^t(\mathcal{I})/\mathcal{T})}, \qquad (6)$$

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

C. pDTW-HMM Intention Anticipation

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 by pDTW-HMM

Input : A set of possible driving manoeuvres $\mathcal{M}$;
         A set of experimental drivers D_{total}.
Output: Intention strategy δ. (Problem 1)

1 **procedure** PDTW-HMM:
2   Initialise prior distribution $P(\mathcal{I}_0)$;
3   **for** t **in** $1 : T$ **do**
4   Record observation point o_t ;
5   Compute $p\mathrm{DTW}_{\mathcal{M}}^t$ (Definition 5) ;
6   Let $P(o_t \mid \mathcal{I}_t, \mathbb{O}_{t-1}) = p\mathrm{DTW}_{\mathcal{I} \in \mathcal{M}}^t$;
7   Infer and normalise $P(\mathcal{I}_t \mid \mathbb{O}_t)$ (Lemma 1) ;
8   Send δ to ADAS ;
9          $t = t + 1$;

Remark. In this work, we let the notion of "observation $\mathbb{O}_t$" denote both a driver's *gaze* and *fixation* (defined in Section V-A), as each of which can be regarded as an aspect of observation, while the former normally contains noise and the latter performs as a filtration. The proposed framework works for both as shown in the experimental results. Besides, one current limitation that we want to relax in the future is the assumption that $P(\mathcal{I}_t \mid \mathcal{I}_{t-1}) = 1$.

(a) G4P15T2 - 3D.

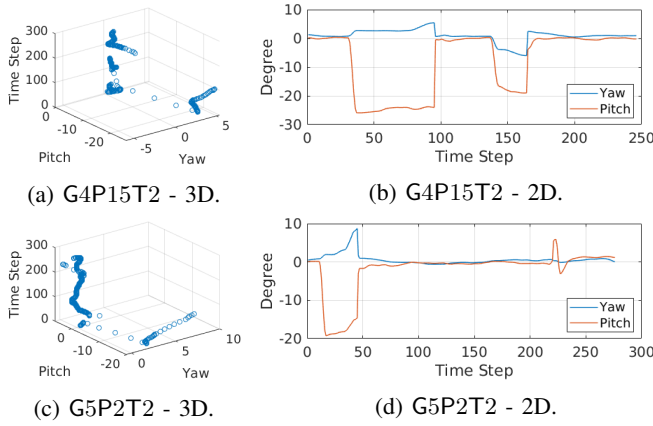(b) G4P15T2 - 2D.

(c) G5P2T2 - 3D.

(d) G5P2T2 - 2D.

Fig. 6: Illustration of driver's *gaze* pattern. Left: gaze points in Yaw-Pitch-Time space; Right: separation of Yaw and Pitch degrees on time steps.

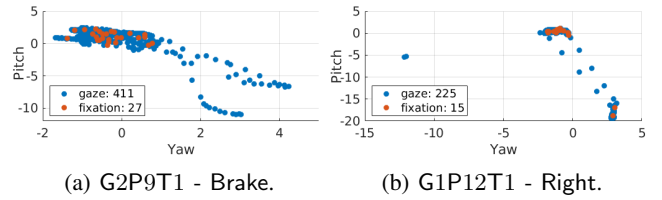

(a) G2P9T1 - Brake.

(b) G1P12T1 - Right.

Fig. 7: 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 $\nu = 60\,\mathrm{Hz}$, duration $\Delta = 0.2\,\mathrm{s}$, fixation range f $= 2°$.)

V. EXPERIMENTAL RESULTS

This section presents the experimental results of anticipating driver's real-time intention over future manoeuvres based on past observations. Overall, 75 drivers participated in our experiments, which produced 124 valid experimental cases from two trials of the scenario (Fig. 2) - 61 in Trial 1, and 63 in Trial 2. We use $\mathsf{G}x\mathsf{P}y\mathsf{T}z$ to mark Participant y of Group x in Trial z, where $x \in \{1, \ldots, 5\}$, $y \in \{1, \ldots, 15\}$, $z \in \{1, 2\}$. The attached video shows a case in which the driver turned to the right lane to overtake the lead vehicle. The experimental setting is a desktop with Intel i5-4690S CPU, 8GB RAM, Fedora 26 (64-bit), and Matlab R2018a.

A. Gaze and Fixation

We consider two forms of observations, *gaze* and *fixations*, to anticipate driver's real-time intention over possible manoeuvres. By gaze, we refer to the raw data collected by the eye-tracker. See examples of gaze patterns in Fig. 6.

We define *fixation* as a driver's gaze maintaining on a fixed area for a certain period of time, e.g., $0.2\,\mathrm{s}$. Fig. 7 illustrates fixation extraction from a driver's gaze sequence. In Fig. 7(a), we observe that 27 fixations were formed from a sequence of 411 gaze points. Although a few gaze points scattered to the right, probably because the driver quickly "evaluated" the right lane and decided not to change lane, all fixations were formed at the centre region of the windscreen, i.e., $[-1° : 1°, 0° : 2°]$, where the lead vehicle was decelerating or perhaps stopping ahead. This corresponds to the actual manoeuvre Brake as the driver needed to focus on the conditions ahead in order to brake in time to avoid a collision.

In this regard, raw gaze points, especially scattered, do not necessarily imply an imminent manoeuvre - on the contrary, a driver may simply glimpse and eliminate the unfeasible manoeuvres. Therefore, we evaluate both *gaze* and the more stable *fixation* points, and in fact, the latter outperforms the former, which will be explained in Table I.

B. Intention Anticipation over Driving Manoeuvres

Comparison between gaze or fixation patterns is achieved by generating a pDTW distribution (Section IV-B). Fig. 8 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. 8(a)) and 2D plot (Fig. 6(b)), as well as G5P2T2's "Original Gaze" (Fig. 8(c)) and 2D plot (Fig. 6(d)). Intuitively, once an optimal alignment between the gaze patterns is found, as shown in the "Overlaid Aligned Gazes" (Fig. 8(f)), the shortest warping distance can be computed. In this case, DTW(G4P15T2, G5P2T2) = 1021.62.

A driver's real-time intention strategy is generated from both gaze sequence and extracted fixations for the *same* duration for better comparison using leave-one-out cross-validation, as illustrated in Fig. 9. 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 is being recorded. Nevertheless, the disadvantage is that the strategy 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 $0.2\,\mathrm{s}$ when fixation forms.

C. Accuracy Validation

We validate the proposed framework through statistically analysing the *accuracy* rate of all the predictions, by comparing the predicted manoeuvre to the *actual* manoeuvre that was taken by an individual driver in an experimental case.

For both trials, we separate the total number N of cases randomly into a training set and a test set. Formally, the separation algorithm is as follows. A total set $\mathsf{D}_{total}[1, N]$ is classified into three manoeuvre categories $\mathsf{D}_{\mathsf{Brake}}[1, B]$, $\mathsf{D}_{\mathsf{Right}}[1, R]$, and $\mathsf{D}_{\mathsf{Left}}[1, L]$, such that $B + R + L = N$ and $N, B, R, L \in \mathbb{N}^+$. Let $\gamma \in (0, 1)$ be a training ratio, then a training set is $\mathsf{D}_{train}[1, \alpha] = \gamma * \mathsf{D}_{\mathsf{Brake}} \cup \gamma * \mathsf{D}_{\mathsf{Right}} \cup \gamma * \mathsf{D}_{\mathsf{Left}}$, 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 $\mathsf{D}_{test}[1, \beta] = \mathsf{D}_{total} \setminus \mathsf{D}_{train}$ such that $\beta = N - \alpha$.

The general accuracy of intention anticipation from gaze and fixation in both trials is illustrated in Fig. 10. In terms of mean accuracy (Fig. 10(a)), as the training ratio γ increases from 85% to 95%, the correct anticipation rate increases.
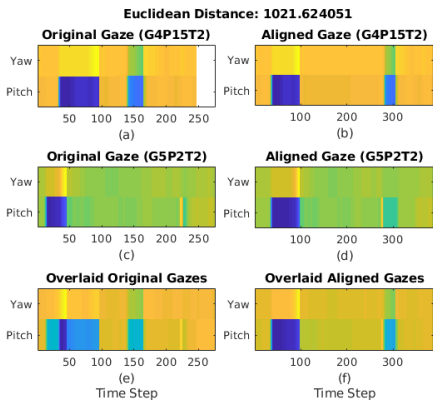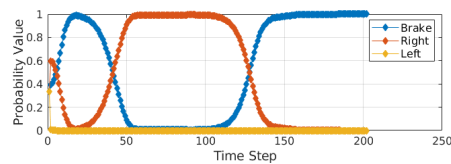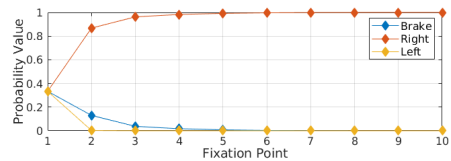
Fig. 8: Comparison of gaze patterns using DTW. Top row: original (a) and aligned (b) gaze sequences of G4P15T2; Middle row: that of G5P2T2; Bottom row: overlaid gaze sequences of both.
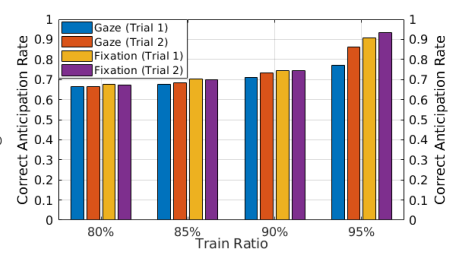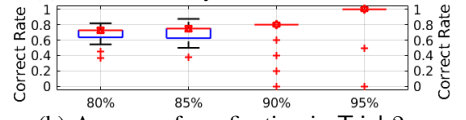

(a) G3P14T1's intention from *gaze*.

(b) G3P14T1's intention from *fixation*.

Fig. 9: Real-time intention strategy over three driving manoeuvres Brake, Right, and Left from *gaze* (a) and *fixation* (b). (Leave-one-out cross-validation, temperature $\mathcal{T} = 1/10$.)

(a) Mean accuracy after 500 iterations.

(b) Accuracy from fixation in Trial 2.

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

Fig. 10(b) describes the box plot of the anticipation accuracy from fixation in Trial 2, corresponding to the purple bars in Fig. 10(a). It shows that, after 500 iterations of random separation of the data set 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%.

The proposed framework pDTW-HMM's advantage over baseline methods, HMM and DTW, is presented in Table I. In this case, HMM essentially updates the intention model utilising the frequency of gaze or fixation points on the windscreen, and discards the temporal dynamics between adjacent points, whereas DTW considers the temporal dependencies of the observation patterns but does not have an intention model. We observe that our framework achieves a higher mean accuracy rate Pr. 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., 128th-202nd time steps in Fig. 9(a), and 2nd-10th fixation points in Fig. 9(b). In experiments, t_b is formatted into seconds. Trial 2 shows slightly shorter duration than Trial 1, e.g., 0.18 s shorter

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

Moreover, with the approximately 3 s of anticipation margin gained from our approach, ADASs can better support driver's decision-making in such safety-critical situations. For instance, if an ADAS detects a driver's tendency to brake in 3 s, however, given the speed of the ego vehicle and the distance to the lead vehicle, after calculation the ADAS realises braking will not prevent collision, then it can advise the driver to turn to the right or left lane to overtake.

VI. CONCLUSION

In this paper, we propose a pDTW-HMM framework, by analysing the gaze and fixation patterns of the driver, especially taking their temporal characteristics into consideration. We show in our experiments that the method can anticipate a driver's real-time intention over future manoeuvres around 3 s beforehand with over 90% accuracy. Future work aims to design a suitable ADAS that can aid drivers in safety-critical situations using the predicted intention through strategy synthesis.

TABLE I: Comparison of our pDTW-HMM framework with two baseline methods HMM and DTW, in terms of mean accuracy rate (Pr) and correct anticipation time before actual manoeuvre (t_b). ($\gamma = 95\%$, $\mathcal{T} = 1/10$.)

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

REFERENCES

[1] B. F. Malle and J. Knobe, "The folk concept of intentionality," *Journal of Experimental Social Psychology*, vol. 33, no. 2, pp. 101–121, 1997.

[2] 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.

[3] A. Jain, A. Singh, H. S. Koppula, S. Soh, and A. Saxena, "Recurrent neural networks for driver activity anticipation via sensory-fusion architecture," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 3118–3125.

[4] 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.

[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 *Proceedings of the International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE)*. 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. chaandar Ravichandar, A. Kumar, and A. Dani, "Bayesian human intention inference through multiple model filtering with gaze-based priors," in *19th International Conference on Information Fusion (FUSION)*. IEEE, 2016, pp. 2296–2302.

[13] C.-M. Huang and B. Mutlu, "Anticipatory robot control for efficient human-robot collaboration," in *The 11th ACM/IEEE international conference on human robot interaction*, 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 *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011, pp. 4819–4825.

[15] C.-M. Huang, S. Andrist, A. Sauppé, 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 *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 *IEEE Intelligent Vehicles Symposium (IV)*, 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 *International Conference on Big Data and Smart Computing (BigComp)*. IEEE, 2016, pp. 143–148.

[19] Y.-S. Jiang, G. Warnell, and P. Stone, "Dipd: Gaze-based intention inference in dynamic environments," in *Workshops at the Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[20] G. Csibra and G. Gergely, "obsessed with goals: Functions and mechanisms of teleological interpretation of actions in humans," *Acta psychologica*, vol. 124, no. 1, pp. 60–78, 2007.

[21] Y. Demiris and G. Hayes, "Imitation as a dual-route process featuring predictive and learning components: A biologically plausible computational model," *Imitation in animals and artifacts*, vol. 327, 2002.

[22] Y. Demiris and M. Johnson, "Distributed, predictive perception of actions: a biologically inspired robotics architecture for imitation and learning," *Connection Science*, vol. 15, no. 4, pp. 231–243, 2003.

[23] D. M. Wolpert, K. Doya, and M. Kawato, "A unifying computational framework for motor control and social interaction," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 358, no. 1431, pp. 593–602, 2003.

[24] B. Hommel, J. Müsseler, G. Aschersleben, and W. Prinz, "The theory of event coding (tec): A framework for perception and action planning," *Behavioral and brain sciences*, vol. 24, no. 5, pp. 849–878, 2001.

[25] P. Kumar, M. Perrollaz, S. Lefevre, and C. Laugier, "Learning-based approach for online lane change intention prediction," in *Intelligent Vehicles Symposium (IV)*. IEEE, 2013, pp. 797–802.

[26] M. M. Trivedi, T. Gandhi, and J. McCall, "Looking-in and looking-out of a vehicle: Computer-vision-based enhanced vehicle safety," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 1, pp. 108–120, 2007.

[27] B. Morris, A. Doshi, and M. Trivedi, "Lane change intent prediction for driver assistance: On-road design and evaluation," in *Intelligent Vehicles Symposium (IV)*. IEEE, 2011, pp. 895–901.

[28] A. Doshi, B. Morris, and M. Trivedi, "On-road prediction of driver's intent with multimodal sensory cues," *IEEE Pervasive Computing*, vol. 10, no. 3, pp. 22–34, 2011.

[29] 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, 2004, pp. 2228–2231.

[30] 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.

[31] 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.

[32] 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.

[33] W. Ruan, L. Yao, Q. Z. Sheng, N. J. Falkner, and X. Li, "Tagtrack: Device-free localization and tracking using passive rfid tags," in *The 11th international conference on mobile and ubiquitous systems: computing, networking and services (MobiQuitous)*, 2014, pp. 80–89.

[34] W. Ruan, Q. Z. Sheng, L. Yao, X. Li, N. J. Falkner, and L. Yang, "Device-free human localization and tracking with uhf passive rfid tags: A data-driven approach," *Journal of Network and Computer Applications*, vol. 104, pp. 78–96, 2018.

[35] W. Ruan, Q. Z. Sheng, L. Yao, T. Gu, M. Ruta, and L. Shangguan, "Device-free indoor localization and tracking through human-object interactions," in *IEEE 17th international symposium on a world of wireless, mobile and multimedia networks (WoWMoM)*. IEEE, 2016, pp. 1–9.

[36] S. Särkkä, *Bayesian filtering and smoothing*. Cambridge University Press, 2013, vol. 3.

[37] J. Kruskal and M. Liberman, "The symmetric time-warping problem: From continuous to discrete," 1983.

[38] W. Ruan, L. Yao, Q. Z. Sheng, N. Falkner, X. Li, and T. Gu, "Tagfall: Towards unobstructive fine-grained fall detection based on uhf passive rfid tags," in *The 12th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services (MobiQuitous)*, 2015, pp. 140–149.

[39] L. R. Rabiner and B.-H. Juang, *Fundamentals of speech recognition*. PTR Prentice Hall Englewood Cliffs, 1993, vol. 14.

[40] M. Müller, *Information retrieval for music and motion*. Springer, 2007, vol. 2.