Modeling the Effect of Driver's Eye Gaze Pattern Under Workload: Gaussian Mixture Approach

Ron M. Hecht (ron.hecht@gm.com), Ariel Telpaz (ariel.telpaz@gm.com), Gila Kamhi (gila.kamhi@gm.com)
Advanced Technical Center Israel, General Motors, 7 Hamada St
Herzelia, 46733 Israel

Omer Tsimhoni (omer.tsimhoni@gm.com)
Warren Technical Center, General Motors, 30565 William Durant Boulevard
Warren, MI 48092, USA

Aharon Bar Hillel (barhille@bgu.ac.il)
Department of Industrial Engineering and Management, Ben-Gurion University of the Negev, P. O. 653
Beer-Sheva 84105, Israel

Naftali Tishby (tishby@cs.huji.ac.il)
Rachel and Selim Benin School of Computer Science and Engineering, The Hebrew University of Jerusalem
Jerusalem 91904, Israel

Abstract

This paper puts forward a Gaussian Mixture Model (GMM) for eye gaze behavior under workload and applies it to the analysis of gaze distributions in an automotive context. Specifically, it extends our work on Information Constrained Control (ICC) (Hecht, Bar-Hillel, Telpaz, Tsimhoni, & Tishby, 2019) (Hecht, Telpaz, Kamhi, Bar-Hillel, & Tisbhy, 2019) (Hecht et al., 2015) (Hecht, Telpaz, Kamhi, Bar-Hillel, & Tishby, 2018) by generating an ICC GMM derivative. We suggest a measure for workload estimation based on the Kullback Leibler divergence (D_{kl}) between tested eye gaze distributions and a reference workload-free distribution. This derivative assumes diagonal Gaussians that are distant from each other. Under these assumptions, we achieve an analytical measure that has significantly fewer parameters than discrete grid-like distributions (Hecht, Bar-Hillel, et al., 2019). Testing our measure on eye gazing data collected during real world driving experiments in a highway environment confirms the effectiveness of this approach.

Keywords: Information Constrained Control; Gaussian Mixture Model; Eye gaze distribution

Introduction

The human visual system has a tendency to shift towards salient regions(Harel, Koch, & Perona, 2006)(Borji & Itti, 2013); however, in the presence of a demanding task this tendency is overridden, and the visual system shifts towards important areas (Lavie & De Fockert, 2005)(Lavie, Hirst, De Fockert, & Viding, 2004)(Lavie, 2010). The interaction between important areas, salient areas, and workload has fascinated researchers for decades, but has never been fully disentangled. One hurdle to a better understanding has to do with modeling the effect of workload on gaze distribution. For example in an automotive environment, Victor et al. (Victor, Harbluk, & Engström, 2005) suggested several simpler measurements for the detection of workload based on gaze patterns (e.g., Standard Deviation of Gaze, Percent Road Center).

The ICC (Tishby & Polani, 2011)(Rubin, Shamir, & Tishby, 2012)(Hecht et al., 2015) constitutes an alternative approach to modeling the effect of workload on the visual system (Hecht, Bar-Hillel, et al., 2019) (Hecht, Telpaz, et al., 2019) (Hecht et al., 2018). This method views one of the goals of the visual system as finding the optimum between two contradictory goals. It aims to find a balance between looking at salient objects and looking at important ones. Workload interacts with this balance and causes a shift in gaze patterns towards important areas. In previous articles, we presented derivatives of the ICC for discrete distributions (Hecht et al., 2018)(Hecht, Bar-Hillel, et al., 2019) and for continuous Gaussian distributions (Hecht, Telpaz, et al., 2019). Unfortunately, gaze distributions are multimodal continuous distributions, and are better modeled by multimodal distributions. Thus here we selected Gaussian Mixture Model (GMM) distributions which are both multimodal and continuous and have a relatively small number of parameters. We generated a GMM derivative to the ICC which we refer to as DIG, which is short for DKL ICC GMM, where D_{kl} is the Kullback Leibler divergence (Cover & Thomas, 2012).

Model

We start the formalization of DIG by defining the action space. In our case, the actions are the direction of sight of the visual system,or more specifically, the intersections between the two-dimensional eye gaze locations with the closest objects in the field of view in an automotive environment. The distribution of gazes over this space is far from uniform. Rather the data are concentrated on several objects that are far away from one another. In our case, the objects are all located inside the vehicle and consist of the mirrors, windshield, dashboard, and instrument cluster. We refer to these two-dimensional points of intersection as the actions and denote them as $\vec{x} \in X$. Figure 2 provides an example of this space.

Information Constrained Control

We start with a short recap of the ICC for the case of a single state (Hecht, Bar-Hillel, et al., 2019). The ICC is defined

formally as the following constrained optimization problem. Three functions are defined over the action space X in the Eq. 1. $R(\vec{x})$ is the reward associated with the execution of action $\vec{x}$ (looking at location $\vec{x}$). $Q(\vec{x})$ is the saliency of action $\vec{x}$. It is defined as the likelihood of execution of action $\vec{x}$ in a situation where no workload exists. $P(\vec{x})$ is the selected distribution over X that the visual system selected to execute.
The equation has two main terms. The main part of the equation presents the optimized term. The optimization is the minimization of the distance between two eye gazing distributions where Q is workload-free (a.k.a. the saliency map) and P, the selected distribution. Intuitively, the goal of this term is to verify that P and Q are as close as possible. The second term represents the constraint, which is reward oriented. This term verifies that a desired level of reward is achieved. Specifically, it is a linear averaged weighted reward. The approach associates reward with workload and associate a high workload with a desire for high reward levels (higher θ), and vice versa.

$$\begin{aligned} \hat{P}(\vec{x}) = \quad & \arg\min_P \quad \int_{\vec{x}\in X} P(\vec{x}) \log \frac{P(\vec{x})}{Q(\vec{x})} d\vec{x} \\ & s.t. \\ & \textstyle\int_{\vec{x}\in X} P(\vec{x}) = 1, \quad \int_{\vec{x}\in X} P(\vec{x}) R(\vec{x}) \geq \theta \end{aligned} \tag{1}$$

According to the model, in a low workload condition, P and Q are expected to be similar , whereas in a high workload condition, it is expected that P and Q are quite different. This suggests that the D_{kl} part of the equation can measure the workload.

Model assumptions

Several assumptions were applied to simplify the model. The first assumption was that all the distributions $P(\vec{x})$ are GMMs where the input is two-dimensional $\vec{x} = (x_1, x_2)$. In addition, we assume that each distribution is composed of M Gaussians.

$$\forall P \in \mathcal{P} \quad P(\vec{x}) = \sum_{i=1}^{M} w_{P,i} g_{P,i} (\vec{x} | \vec{\mu}_{P,i}, \Sigma_{P,i}) \tag{2}$$

where $\mathcal{P}$ is the set of all GMM distributions and i is the Gaussian identity within the distribution.

In the experiment described below, a set of participants drove a vehicle. Each participant drove along the same route under several workload conditions. We modeled the gaze distribution of each participant during each ride using a single GMM distribution. Thus overall, the number of distributions was the number of participants times the number of rides. Our assumption regarding a constant M Gaussians is reasonable in a driving context. For example, a single Gaussian could be associated with looking at the middle of the road, another Gaussian with the left mirror and a third with the instrument cluster. In addition, we assumed that all the Gaussians had a diagonal covariance.

$$\forall P, i \quad \Sigma_{P,i} = \begin{bmatrix} \sigma^2_{P,i,x_1} & & 0 \\ & \ddots & \\ 0 & & \sigma^2_{P,i,x_D} \end{bmatrix} \tag{3}$$

where D is the input dimension (in our case $D = 2$). We made this assumption to simplify the model; however, it is reasonable from the data perspective as well. Furthermore, we assumed “social distancing” of the Gaussians. Within each mixture, Gaussians are very distant from one another.

$$\begin{matrix} \forall P \in \mathcal{P} \quad d \in \{1, \ldots, D\} \\ i, j \in \{1, \ldots, M\} \end{matrix} \quad \frac{\left| \mu_{P,i,x_d} - \mu_{P,j,x_d} \right|}{\max\left(\sigma_{P,j,x_d}, \sigma_{P,i,x_d}\right)} \gg 1 \tag{4}$$

More specifically, it is reasonable to assume that the Gaussian in the middle of the road is relatively distant from the Gaussian located on the left mirror and that both of them are distant from the instrument panel Gaussian. The distance among the means of the Gaussians is visualized by the histogram presented at Fig. 2. This suggests that there is a one to one mapping among the Gaussians in the different mixtures.

$$\forall P_1, P_2 \in \mathcal{P}, i \in \{1, \ldots, M\} \quad \vec{\mu}_{P_1,i} = \vec{\mu}_{P_2,i} = \vec{\mu}_i \tag{5}$$

The first Gaussian of distribution P_1 shares the same mean with the first Gaussian of the distribution P_2. This assumption is reasonable for driving scenarios as well. On one hand, it is reasonable to assume that the instrument cluster Gaussian is situated in the same location. On the other hand, this is merely an approximation and the means in reality can shift a little. In particular, this shift can be observed in Gaussians located on the road. As the Information Constrained Control (ICC) (Hecht, Bar-Hillel, et al., 2019) (Hecht, Telpaz, et al., 2019) (Hecht et al., 2015) (Hecht et al., 2018) (Tishby & Polani, 2011) suggests, there is a baseline distribution Q which models the gaze pattern in a task-free scenario. We assumed this distribution to be a GMM similar to the ones in $\mathcal{P}$.

Distance from baseline distribution

The ICC‘s selected eye gaze distribution is generated by a tradeoff between two goals: achieving a high enough level of reward and maintaining a minimal distance from a baseline distribution. The distance from the baseline distribution Q is defined as:

$$D_{kl}(P, Q) = \int_X P(\vec{x}) \log \frac{P(\vec{x})}{Q(\vec{x})} d\vec{x} \tag{6}$$

Since the Gaussians in the mixture are located very far from one another, we can define areas in which each Gaussian is dominant. We define and denote by R_j the area in which the j Gaussian is dominant. Explicitly, we define dominant to be the area in $\vec{x}$ in which:

$$\begin{aligned}&\forall P \in \mathcal{P} \cup \{Q\}, i, j \in \{1,...,M\}, i \neq j\\&w_{P,j} g_{P,j}(\vec{x}|\vec{\mu}_{P,j}, \Sigma_{P,j}) \gg\\\gg\ &w_{P,i} g_{P,i}(\vec{x}|\vec{\mu}_{P,i}, \Sigma_{P,i}), \frac{\varepsilon}{M-1}\end{aligned} \tag{7}$$

The D_{kl} is approximated by the sum over the following integrals.

$$D_{kl}(P,Q) \approx \sum_{j=1}^{M} \int_{R_j} P(\vec{x}) \log \frac{P(\vec{x})}{Q(\vec{x})} d\vec{x} \tag{8}$$

By plugging the explicit equation of the GMM distribution into Eq. 8 and reordering the summation, the following equation emerges:

$$= \sum_{j=1}^{M} \int_{R_j} \sum_{i=1}^{M} w_{P,i} g_{P,i}(\vec{x}|\vec{\mu}_{P,i}, \Sigma_{P,i}) \log \frac{P(\vec{x})}{Q(\vec{x})} d\vec{x} \tag{9}$$

$$= \sum_{j=1}^{M} \sum_{i=1}^{M} w_{P,i} \int_{R_j} g_{P,i}(\vec{x}|\vec{\mu}_{P,i}, \Sigma_{P,i}) \log \frac{P(\vec{x})}{Q(\vec{x})} d\vec{x} \tag{10}$$

We can use the definition of the region R_j to explicitly write the ratio $\frac{P(\vec{x})}{Q(\vec{x})}$ in that region.

$$\frac{P(\vec{x})}{Q(\vec{x})} = \frac{\sum_{i=1}^{M} w_{P,i} g_{P,i}(\vec{x}|\vec{\mu}_{P,i}, \Sigma_{P,i})}{\sum_{i=1}^{M} w_{Q,i} g_{Q,i}(\vec{x}|\vec{\mu}_{Q,i}, \Sigma_{Q,i})} \tag{11}$$

$$\approx \frac{w_{P,j} g_{P,j}(\vec{x}|\vec{\mu}_{P,j}, \Sigma_{P,j}) + (M-1)\frac{\varepsilon}{M-1}}{w_{Q,j} g_{Q,j}(\vec{x}|\vec{\mu}_{Q,j}, \Sigma_{Q,j}) + (M-1)\frac{\varepsilon}{M-1}} \tag{12}$$

$$\approx \frac{w_{P,j} g_{P,j}(\vec{x}|\vec{\mu}_{P,j}, \Sigma_{P,j})}{w_{Q,j} g_{Q,j}(\vec{x}|\vec{\mu}_{Q,j}, \Sigma_{Q,j})} \tag{13}$$

By combining Eq. 10 and 13, a simplified version of the D_{kl} emerges (for ease of notation the P, j Gaussian is denoted g, without stating explicitly $\mu_{P,j}$ and $\Sigma_{P,j}$. $g_{P,j}(\vec{x}|\vec{\mu}_{P,j}, \Sigma_{P,j})$ is denoted as $g_{P,j}(\vec{x})$)

$$\begin{aligned}&D_{kl}(P,Q) \approx\\\approx\ &\sum_{j=1}^{M} \sum_{i=1}^{M} w_{P,i} \int_{R_j} g_{P,i}(\vec{x}) \log \frac{w_{P,j} g_{P,j}(\vec{x})}{w_{Q,j} g_{Q,j}(\vec{x})} d\vec{x}\\=\ &\sum_{j=1}^{M} \sum_{i=1}^{M} w_{P,i} \int_{R_j} g_{P,i}(\vec{x}) \log \frac{w_{P,j}}{w_{Q,j}} d\vec{x}\\+\ &\sum_{j=1}^{M} \sum_{i=1}^{M} w_{P,i} \int_{R_j} g_{P,i}(\vec{x}) \log \frac{g_{P,j}(\vec{x})}{g_{Q,j}(\vec{x})} d\vec{x}\end{aligned} \tag{14}$$

The D_{kl} is divided into two sets of terms. The First set of terms focuses on the ratio between the log weights of the Gaussians, and the second term is similar to the log likelihood ratio between individual Gaussians. Our assumptions regarding the "social distancing" of the Gaussians gave us an opportunity to split the bigger problem of estimating the D_{kl} between two GMMs into a set of smaller D_{kl}-like problems. Even within the first set of terms, the Gaussians and weights can be decoupled. The weight can be extracted from the integration:

$$\begin{aligned}&\sum_{j=1}^{M} \sum_{i=1}^{M} w_{P,i} \int_{R_j} g_{P,i}(\vec{x}) \log \frac{w_{P,j}}{w_{Q,j}} d\vec{x} =\\=\ &\sum_{j=1}^{M} \sum_{i=1}^{M} w_{P,i} \log \frac{w_{P,j}}{w_{Q,j}} \int_{R_j} g_{P,i}(\vec{x}) d\vec{x}\end{aligned} \tag{15}$$

Our next step is to simplify the last equation by using the definition of R_j. Intuitively, the area R_j was defined to hold most of the probability mass of Gaussian j, and it almost does not hold any probability mass of other Gaussians. More formally, for cases where $i = j$, since the Gaussians are far away from one another, R_j covers most of the probability mass of g_{Pi}.

$$\int_{R_j} g_{P,i}(\vec{x}) = \int_{R_i} g_{P,i}(\vec{x}) \approx 1 \tag{16}$$

For the rest of the cases where $i \neq j$, almost no probability mass is left.

$$\int_{R_j} g_{P,i}(\vec{x}) \approx 0 \tag{17}$$

Eq. 15 can be simplified by splitting the terms in the equation to two groups ($i \neq j, i = j$). The first set of terms where $i = j$ remains (based on Eq. 16), while the other set where $i \neq j$ is nullified (based on Eq. 17). Overall, only M terms have values different from zero (approximation).

$$\begin{aligned}&\sum_{j=1}^{M} \sum_{i=1}^{M} w_{P,i} \log \frac{w_{P,j}}{w_{Q,j}} \int_{R_j} g_{P,i}(\vec{x}) d\vec{x}\\=\ &\sum_{i=1}^{M} w_{P,i} \log \frac{w_{P,i}}{w_{Q,i}} \int_{R_i} g_{P,i}(\vec{x}) d\vec{x}\\+\ &\sum_{j=1}^{M} \sum_{\substack{i=0\\i\neq j}}^{M} w_{P,i} \log \frac{w_{P,j}}{w_{Q,j}} \int_{R_j} g_{P,i}(\vec{x}) d\vec{x}\\\approx\ &\sum_{i=1}^{M} w_{P,i} \log \frac{w_{P,i}}{w_{Q,i}} \cdot 1 + \sum_{j=1}^{M} \sum_{\substack{i=0\\i\neq j}}^{M} w_{P,i} \log \frac{w_{P,j}}{w_{Q,j}} \cdot 0\\=\ &\sum_{i=1}^{M} w_{P,i} \log \frac{w_{P,i}}{w_{Q,i}} = D_{kl}(\vec{w_P}, \vec{w_Q})\end{aligned} \tag{18}$$

The first term of Eq. 14 is the D_{kl} between the two weights vectors.

The second term of Eq. 14 can be approximated in a similar way. We start by dividing the terms into sets ($i = j, i \neq j$). Later, the terms that are associated with $i \neq j$ are nullified (based on Eq. 7). Eventually, only the $i = j$ terms have value different than zero and thus remain.

$$\begin{aligned}
&\sum_{j=1}^{M}\sum_{i=1}^{M} w_{P,i}\int_{R_j} g_{P,i}(\vec{x})\log\frac{g_{P,j}(\vec{x})}{g_{Q,j}(\vec{x})}d\vec{x}\\
=\;&\sum_{j=1}^{M} w_{P,j}\int_{R_j} g_{P,j}(\vec{x})\log\frac{g_{P,j}(\vec{x})}{g_{Q,j}(\vec{x})}d\vec{x}\\
+\;&\sum_{j=1}^{M}\sum_{\substack{i=0\\ i\neq j}}^{M} w_{P,i}\int_{R_j} g_{P,i}(\vec{x})\log\frac{g_{P,j}(\vec{x})}{g_{Q,j}(\vec{x})}d\vec{x}\\
\approx\;&\sum_{j=1}^{M} w_{P,j}\int_{R_j} g_{P,j}(\vec{x})\log\frac{g_{P,j}(\vec{x})}{g_{Q,j}(\vec{x})}d\vec{x}
\end{aligned} \tag{19}$$

The majority of the probability mass of the j^{th} Gaussians are located in the area R_j. This suggests that integration over the entire space is a reasonable approximation to the integration over R_j. This integration over the entire space is by definition the Kullback Leibler divergence (D_{kl}) between two Gaussian distributions.

$$\int_{R_j} g_{P,j}(\vec{x})\log\frac{g_{P,j}(\vec{x})}{g_{Q,j}(\vec{x})}d\vec{x} \approx \int g_{P,j}(\vec{x})\log\frac{g_{P,j}(\vec{x})}{g_{Q,j}(\vec{x})}d\vec{x} = D_{kl}\left(g_{P,j}(\vec{x}),g_{Q,j}(\vec{x})\right) \tag{20}$$

Our next step is plugging Eq. 20 into Eq. 19.

$$\sum_{j=1}^{M} w_{P,j}\int_{R_j} g_{P,j}(\vec{x})\log\frac{g_{P,j}(\vec{x})}{g_{Q,j}(\vec{x})}d\vec{x} \approx \sum_{j=1}^{M} w_{P,j} D_{kl}\left(g_{P,j}(\vec{x}),g_{Q,j}(\vec{x})\right) \tag{21}$$

This term is the weighted D_{kl} between the Gaussians of both distributions.

Recall that both Gaussians have diagonal covariance. For this case, the D_{kl} equals (see supporting material):

$$\begin{aligned}
D_{kl}\left(g_{P,j}(\vec{x}),g_{Q,j}(\vec{x})\right) &= -\frac{D}{2}+\frac{1}{2}\sum_{d=1}^{D}\log\left(\frac{\sigma^2_{Q,j,d}}{\sigma^2_{P,j,d}}\right)\\
+\;&\sum_{d=1}^{D}\frac{\sigma^2_{P,j,d}}{2\sigma^2_{Q,j,d}}+\sum_{d=1}^{D}\frac{\left(\mu_{P,j,d}-\mu_{Q,j,d}\right)^2}{2\sigma^2_{Q,j,d}}
\end{aligned} \tag{22}$$

For Gaussians with diagonal covariance and where $\mu_{P,j}=\mu_{Q,j}$ for all j, the equation becomes:

$$\begin{aligned}
&D_{kl}\left(g_{P,j}(\vec{x}),g_{Q,j}(\vec{x})\right) = \\
=\;& -\frac{D}{2}+\frac{1}{2}\sum_{d=1}^{D}\log\left(\frac{\sigma^2_{Q,j,d}}{\sigma^2_{P,j,d}}\right)+\sum_{d=1}^{D}\frac{\sigma^2_{P,j,d}}{2\sigma^2_{Q,j,d}}
\end{aligned} \tag{23}$$

Overall, the D_{kl} between both distributions is presented in Eq. 24. We refer to this value as the DIG score and it is our measure of workload.

$$\begin{aligned}
&D_{kl}(P,Q)\approx\\
\approx\;& D_{kl}\left(\vec{w_P},\vec{w_Q}\right)+\sum_{j=1}^{M} w_{P,j} D_{kl}\left(g_{P,j}(\vec{x}),g_{Q,j}(\vec{x})\right)\\
=\;&\sum_{i=1}^{M} w_{P,i}\log\frac{w_{P,i}}{w_{Q,i}}\\
+\;&\sum_{j=1}^{M} w_{P,j}\left(-\frac{D}{2}+\frac{1}{2}\sum_{d=1}^{D}\log\left(\frac{\sigma^2_{Q,j,d}}{\sigma^2_{P,j,d}}\right)+\sum_{d=1}^{D}\frac{\sigma^2_{P,j,d}}{2\sigma^2_{Q,j,d}}\right)
\end{aligned} \tag{24}$$

Approximation evaluation

Until now, we have shown the theoretical basis for our approximation. In this subsection, we evaluate it practical quality by comparing our approximation to a baseline on artificial data. The baseline that we selected to estimate the D_{kl} between distributions P and Q is the difference between cross-entropy values (Geyer, Papaioannou, & Straub, 2019). The first step, according to this approach, is to generate a sample set from distribution P. A large enough sample size N is selected to ensure a reasonable coverage of the sample space. In our case of two-dimensional space with four Gaussians that had diagonal covariance matrix, N was selected to be 10,000. We denoted the i sample generated by this process as x_i. The next steps were the estimation of the cross-entropy of the sample with the distribution P ($\sum_{i=1}^{N}\log P(x_i)$) and the distribution Q ($\sum_{i=1}^{N}\log Q(x_i)$). The difference between the two cross-entropy values (Eq. 25) is known as a good estimation to the D_{kl} values between the two distributions.

$$D_{kl}(P,Q)\approx\sum_{i=1}^{N}\log\frac{P(x_i)}{Q(x_i)} \tag{25}$$

where $\{x_i\}_{i=1}^{N}$ were sampled from P.

200 pairs of P,Q distributions were randomly drawn in order to compare the two approaches. For each pair, we approximated the D_{kl} using our approximation (Eq. 24). In addition, for each pair P,Q, we drew 100 times a sample set of 10,000 samples. For each of the 100 sample sets, we estimated the D_{kl} using difference of cross-entropy (Eq. 25). Out of the 100 D_{kl} values, their mean and standard deviation were estimated. Figure 1 shows the results of the comparison. Each pair is represented by its approximated value (Eq. 24) and by the mean and standard deviation of its estimated cross-entropy.

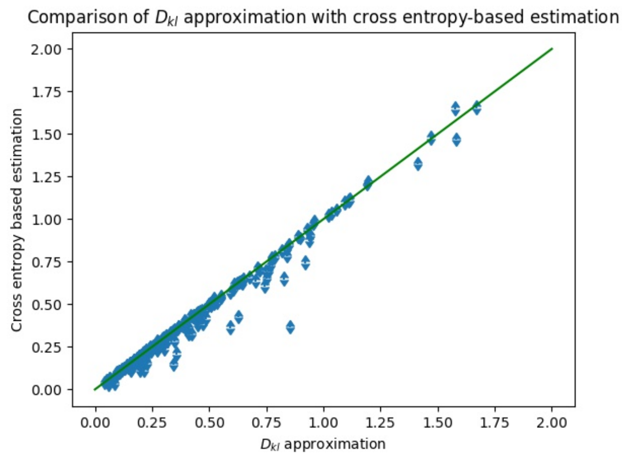

Figure 1: This figure presents the comparison of our D_{kl} approximation with the commonly used difference of cross-entropy estimation of D_{kl}. The axes represent the D_{kl} values according to the approaches. The green line is the optimal condition were both approaches agree. The blue markings represent empirical results on artificial data. Each marking has an error bar of single standard deviation.

The green line represents the ideal situation where there is an agreement between both approaches regarding the D_{kl} values. Most of the times there is a good agreement between the two.

Method

The effectiveness of our model was evaluated on data collected during an on-road experiment with repeated trials. The experiment is described in more detailed in (Tractinsky, 2013). The goal of the experiment was to better understand the way participants learn a new task involving fuel-efficient driving. We focused on a subset of participants that drove in a vehicle that was not changed throughout the experiment. Each participant repeated the same route four times. We associated ease of performing with Drive Identification Number (DIN). In other words, DIN is the chronological number of the iteration /repetition. Ease of driving / performing the task increases with DIN. The participants' first drive is more demanding than the second one and so on. We compared the eye-gaze distributions over repetitions, to detect the ease of the task (Higher DIN were associated with greater ease). We focused on a single segment of the route, that was relatively straight. It consisted of a highway entrance ramp and straight segment of a highway.

Participants

Our subset was composed of twelve participants (six females and six males), ranging in age from 25 to 63 (Mean = 31.6, Median = 29). The participants were required to have a valid driving license for at least two years, and confirm that they drove on a daily basis. 6 participants were using family size vehicle on a regular basis and 5 participants were using a smaller. Only a single participant was using a Sport Utility Vehicle (SUV). All were naïve to the purpose of the study. Participants stated they had normal vision. Due to technical reasons, The height of the participant was limited to 185 cm. Prior to the start of the experiment, the participants gave their informed consent in compliance with the guidelines of the Ben Gurion University - Institutional Review Board. At the end of the experiment, each participant was paid a fee between 150-250 NIS (Today and during the experiment period, 1$ was worth about 3.5 NIS) based on their fuel consumption. An additional 50 NIS was payed for extra time.

Apparatus

The experiment was conducted in an SUV from which the vehicle and eye tracking data was extracted:
Vehicle data - GPS – Location of the vehicle.
Vehicle data - Fuel Consumption Efficiency (FCE) score – Fuel consumption efficiency was extracted directly from the vehicle.
Eye tracking – The Smart-Eye pro eye tracking system (manufactured and developed by Smart Eye AB, Gothenburg, Sweden; http://smarteye.se/) was used to track the participants' eye movements. The system had two IR cameras and two IR LEDs. Data was collected at 60 Hz.
This analysis does not involve the FCE data.

Procedure and design

The experiment started when the participants were educated about the experiment and gave their informed consent to participate in it. Later, a background questionnaire was filled. Finally, after the participants were familiarized with the vehicle, they were asked to drive a 7km ride. The route was composed of an urban and highway setting. The route was repeated four time by each participant. Following each ride, a three minutes break was given to the participants. This rest period was used to provide feedback regarding FCE score. Here, we focused on the highway segment of the route. At the end of the experiment, the participants were paid based on their FCE score.

Independent Variable

As stated in the Method Section, the independent variable was the Drive Identification Number (DIN). This value varied between 1 to 4. The first repetition of the route was denoted by 1 and the last was denoted by 4. Let's recall that the participants are going through a training process. The task is introduced just before the first ride and participants are getting more acquainted in each iteration.

Dependent Variables

The dependent variables themselves were the x and y coordinates of eye gaze samples that were collected over time. Based on those samples a set of statistics were estimated. The statistics were the parameters of a GMM distribution. This

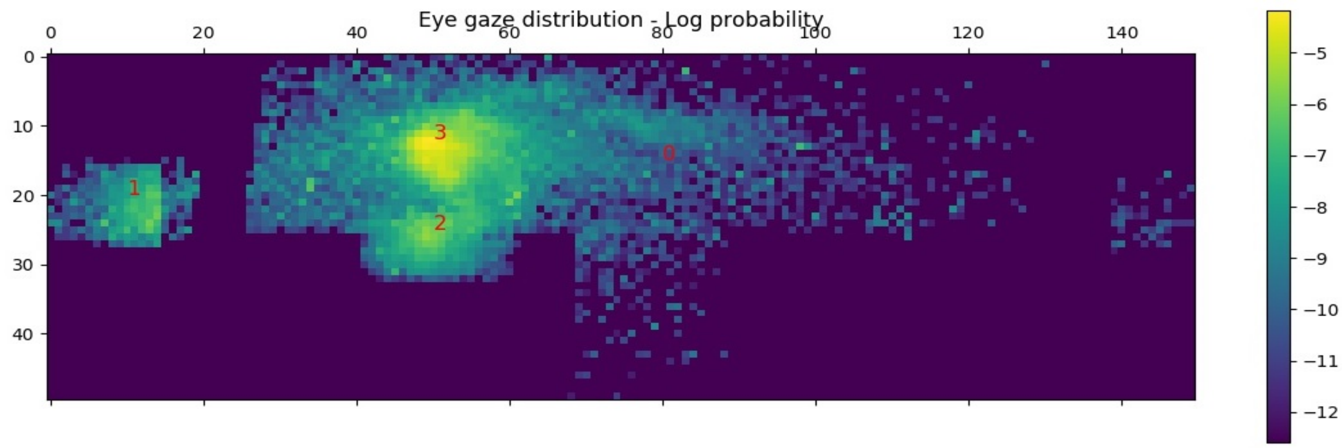

Figure 2: The figure presents the histogram of gaze intersection locations with objects in the field of view. The X and Y axes are the horizontal and vertical coordinates. Light colored areas are associated with areas with high probabilistic density. Three objects are marked by numbers in the figure. Objects 1 and 2 are the left mirror and the instrument cluster respectively. Object 3 is associated with the windshield, and the concentration of probability is linked to the focus on the road ahead.

distribution is the P distribution as presented in former equations throughout the paper. The reference GMM distribution Q was selected to be relatively uniform. It consisted of a four Gaussians with equal weight for each Gaussian and a variance of one. Based on the P and Q GMM distributions the DIG scores (as presented at Eq. 24) were estimated. A DIG score was estimated for each participant and for each participant‘s drive (12 participants X 4 drive).

Results

The overall two dimensional distribution of eye gazes over all participants and rides is presented in Figure 2. The continuous nature of the distribution can easily be observed and specifically its similarity to GMM distribution. The log scale of the color scheme emphasizes that the different components of the distribution are isolated from each other. This isolation runs deep in the DIG assumptions. Our model predicts a monotonic decrease in DIG scores. Higher DIN are expected to have lower DIG scores (The higher the DIN, the greater similarity to the baseline distribution). This was verified by estimating the DIG score for each participant and each drive. Within each participant, we conducted pairwise comparisons for the four drives (1 vs. 2, 1 vs. 3, 1 vs. 4, 2 vs. 3, 2 vs. 4, 3 vs. 4). Within each pair, we tested whether the drive with the higher DIN had lower DIG score. Later, for each participant, we counted the number of times the higher DIN had a lower DIG score. We used a proportion t-test (H_0 was uniform). The average proportion for each participant was 0.6806 relative to the alternative hypothesis of 0.5. This difference was significant ($t = 2.861718$, $df = 11$, $p = 0.015469$). Figure 3 presents the histogram of DIG score differences. The difference is between the score of a ride with lower DIN and one with higher DIN. The comparison was conducted only within participant‘s drives. We expect rides with lower DIN to have higher DIG score, and vice-versa; therefore, we expect the difference between scores to be positive. The histogram was presented for visualization purposes. One can easily observe that shift towards the positive values.

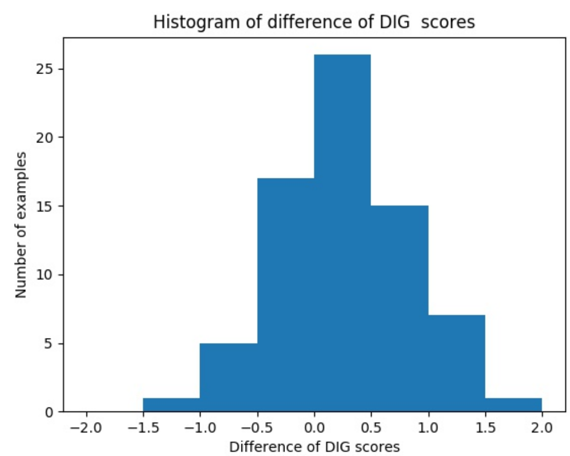

Figure 3: The figure presents the histogram of difference between DIG scores. The experiment were followed by within participant comparisons. Each comparison was conducted between two rides.

Discussion and Conclusion

This paper presented and evaluated a GMM derivative for ICC. This derivative provides an analytical solution that is both intuitive and easy to compute. GMM is a vast family of distributions that are commonly used. This suggests that this derivative might be found useful. It is important to note that although not all of the model's assumptions held all the time, the measure was useful. Unfortunately, due to the small dataset, a comparison across different workload estimation measures could not be performed. This is left for future work.

References

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *35*(1), 185–207.

Cover, T. M., & Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.

Geyer, S., Papaioannou, I., & Straub, D. (2019). Cross entropy-based importance sampling using gaussian densities revisited. *Structural Safety*, *76*, 15–27.

Harel, J., Koch, C., & Perona, P. (2006). Graph-based visual saliency. In *Advances in neural information processing systems* (pp. 545–552).

Hecht, R. M., Bar-Hillel, A., Telpaz, A., Tsimhoni, O., & Tishby, N. (2019). Information constrained control analysis of eye gaze distribution under workload. *IEEE Transactions on Human Machine Systems*.

Hecht, R. M., Bar-Hillel, A., Tiomkin, S., Levi, H., Tsimhoni, O., & Tishby, N. (2015). Cognitive workload and vocabulary sparseness: Theory and practice. In *Sixteenth annual conference of the international speech communication association.*

Hecht, R. M., Telpaz, A., Kamhi, G., Bar-Hillel, A., & Tisbhy, N. (2019). Information constrained control for visual detection of important areas. In *International conference on acoustics, speech, and signal processing.*

Hecht, R. M., Telpaz, A., Kamhi, G., Bar-Hillel, A., & Tishby, N. (2018). *Disentanglement of top-down and bottom-up processes using information constrained control.* in poster without proceedings at the fifth Israeli Conference on Cognition Research.

Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science*, *19*(3), 143–148.

Lavie, N., & De Fockert, J. (2005). The role of working memory in attentional capture. *Psychonomic bulletin & review*, *12*(4), 669–674.

Lavie, N., Hirst, A., De Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, *133*(3), 339.

Rubin, J., Shamir, O., & Tishby, N. (2012). Trading value and information in mdps. In *Decision making with imperfect decision makers* (pp. 57–74). Springer.

Tishby, N., & Polani, D. (2011). Information theory of decisions and actions. In V. Cutsuridis, A. Hussain, & J. G. Taylor (Eds.), *Perception-action cycle* (p. 601-636). Springer New York.

Tractinsky, N. (2013). *Analysis of the field-study questionnaire* (Tech. Rep.). Ben-Gurion University.

Victor, T. W., Harbluk, J. L., & Engström, J. A. (2005). Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(2), 167–190.