What Can Be Predicted from Six Seconds of Driver Glances?

Lex Fridman[1]   **Heishiro Toyoda**[2]   **Sean Seaman**[3]   **Bobbie Seppelt**[3]
Linda Angell[3]   **Joonbum Lee**[1]   **Bruce Mehler**[1]   **Bryan Reimer**[1]

[1] Massachusetts Institute of Technology
[2] Toyota Collaborative Safety Research Center
[3] Touchstone Evaluations

ABSTRACT

We consider a large dataset of real-world, on-road driving from a 100-car naturalistic study to explore the predictive power of driver glances and, specifically, to answer the following question: what can be predicted about the state of the driver and the state of the driving environment from a 6-second sequence of macro-glances? The context-based nature of such glances allows for application of supervised learning to the problem of vision-based gaze estimation, making it robust, accurate, and reliable in messy, real-world conditions. So, it's valuable to ask whether such macro-glances can be used to infer behavioral, environmental, and demographic variables? We analyze 27 binary classification problems based on these variables. The takeaway is that glance can be used as part of a multi-sensor real-time system to predict radio-tuning, fatigue state, failure to signal, talking, and several environment variables.

ACM Classification Keywords

I.5 Pattern Recognition.

Author Keywords

Gaze patterns; driver state prediction; naturalistic on-road study; hidden Markov models.

INTRODUCTION

As the level of vehicle automation continues to increase, the car is more and more becoming a multi-sensor computational system tasked with understanding (1) the state of the driver [37] and (2) the state of the driving environment [25]. From a computer vision perspective, both of these tasks have over a decade of active research that proposes various methods for robust, real-time processing of inward-facing and outward-facing video to extract actionable knowledge with which the car can assist the driver [33]. Driver gaze classification is one of the more successful recent outcomes of these computer vision efforts promising above 90% gaze region classification accuracy [11] in the wild when simplifying the general gaze estimation problem by considering broad segmentation of gaze base on attention allocation semantics: forward

roadway, left mirror, right mirror, rearview mirror, instrument cluster, center stack, and other regions. The promise of accurate real-time gaze classification is what motivates the question posed in this work: once the system infers gaze region from video, what can we predict about the state of the driver and the state of the environment? Put another way, a gaze classification system can be seen as one of several sensors available in the vehicle, and so it is valuable to investigate what actionable information can be inferred from this sensor in order to design a better interface between human and machine in the driving context.

The driving task and the driving environment places the driver under a wide array of physical and cognitive demands. Intuitively, visual demand can be predicted from gaze [29]. However, glance allocation strategies provide a window through which we can predict the more general mental and physical state of the driver outside of just where they are looking (i.e., activity [23], inattention [7], fatigue [15]). The possibility of inference about aspects of the external driving environment based on macro-glances is an open question for which this paper provides promising results. In this paper, "macro-glances" refer to the discretization of driver gaze (see the "Macro-Glances and Micro-Glances" section).

The generalizability of our exploratory look at what can and cannot be predicted from driver macro-glances relies inextricably on the characteristics of the dataset. We use 4,816 annotated six-second epochs of baseline driving from the 100-Car Naturalistic Driving Study database [6]. This dataset includes approximately 2,000,000 vehicle miles, almost 43,000 hours of data, 241 primary and secondary drivers, 12 to 13 months of data collection for each vehicle, and data from a highly capable instrumentation system including five channels of video and vehicle kinematics. This data contains many extreme cases of driving behavior and performance, including severe fatigue, impairment, judgment error, risk taking, willingness to engage in secondary tasks, aggressive driving, and traffic violations. Therefore, we believe that conclusions derived from this data are applicable to general real-world driving.

The contribution, novelty, and validity of this work can be summarized most briefly as follows:

- **Contribution:** Show that just 6 seconds of coarse driver gaze regions can be used to predict a lot of things about the driver, the car, and the driving environment. This helps (1) provide a greater understanding of the "human" in human-to-vehicle interaction and (2) pave the way for a real-time HMI system in the car based on driver gaze that is robust

to challenging real-world conditions. See the "Framework for Gaze-Based HMI in the Car" section below.

- **Novelty:** Drive gaze has been used to predict attention allocation, but not to predict everything else. We try to do just that for the first time and show when it works and when it doesn't.

- **Validity:** The results are based on a large naturalistic on-road study with little to no constraints on the participants, so the data is representative of the general population and is extensive enough to provide a high degree of generalizability.

RELATED WORK

The 100-Car Naturalistic Driving Study dataset has been extensively used to analyze various aspects of driver behavior in the wild [6]. Much of the focus has been on the crashes and near-crashes in the data, and describing the factors that lead to these crashes [22] especially with regard to the long glances away from the road [20]. We focus instead on the baseline driving epochs which are more representative of the variability of driver behavior and driving environment.

Macro-Glances and Micro-Glances

We define the terms "macro-glances" and "micro-glances" to help specify the distinction between context-dependent and context-independent allocations of gaze:

- **Micro-Glances:** Context-independent gaze allocation achieved by fixational eye movement (i.e., saccades) and changes in head orientation. The target "location" of micro-glances is defined by the exact 3D coordinates of the fixation point. Example: driver looking at a stop sign.

- **Macro-Glances:** Gaze allocation categorized into discrete regions that are defined by the context. The target "location" of macro-glances is one of these pre-defined regions and not the exact 3D coordinates of the fixation point. Example: driver looking at the forward roadway.

In this work, we analyze the sequence of driver macro-glances which contains both spatial and temporal information. In particular, the temporal characteristics of the transition between glance regions is the main feature being utilized. This is in contrast to the traditional method of measuring driver state, such as measuring the total time visual attention is directed away from the forward roadway [2] or to elements specific to the operation of the HMI [12]. While these measures are intuitive and have some demonstrated utility, a shortcoming common to both approaches is that they aggregate behavioral information over a given time span to a single number, disregarding temporal dynamics that has to be considered in making predictions about the driver's physical and mental state. A good example of where temporal information is very important but has not been investigated as much as the aggregate measure is in using blink for fatigue classification [24]. Furthermore, work studying a driver's situational awareness of the driving environment has shown the complexity and context-dependent nature of a driver's gaze dynamics [8, 4].

It is intuitive that high-resolution micro-glances such as blinks and individual saccades could be used to predict the behavioral and environmental variables in this work. For example, glances patterns have been correlated with lane change behavior [28]. The open question is whether short-windowed macro-glances can be used for these classification problems. This is an important question because detection and tracking of driver micro-glances in in-the-wild on-road data is much less accurate than detection of macro-glances. The ability to rely on macro-glances alone for predictive tasks allows for the design of robust, real-time driver assistance systems that modify the behavior of the vehicle based on the detected states.

Prediction from Driver Glances

Driver glances have been used to to predict several aspects of driver state including cognitive load [21], secondary activity [27], and drowsiness [35], as covered in this section. The key novel contribution of our work is that we are using glance patterns to predict aspects of driving that are not obviously related to gaze and thus have not been analyzed in prior literature. These aspects include demographics (e.g., age, gender), behavior (e.g., failure to signal, talking), and environment (intersection proximity, lighting conditions, road type). In other words, this work serves as a new and useful exploration of what broad macro eye-movement reveal about the state of the driver and the state of the driving environment.

Cognitive Load and Secondary Tasks

As reviewed in [3], some initial work has been conducted that may be useful in the development of algorithms for identifying periods of driving during which different types of task loading occur. Much of this work was initially focused on identifying visual demand, or periods of visual or visual-manual task loading (e.g., [34, 18, 16, 27]). Additional work been directed at developing algorithms to identify cognitive load using eye glance behavior and driving performance metrics as inputs. Using data collected in a driving simulator, Zhang et al. explored a decision tree approach in [40] to estimate drivers' cognitive workload. Also working with simulator based data, Liang et al. used similar measures in a support vector machine (SVM) approach using a 40-second window with 95% overlap between windows in [21] to detect cognitive distraction, and obtained 91.6% accuracy in the structured predictions to which the model was applied. In [19] Liang et al. worked with Bayesian network models and found that they could identify cognitive load reliably with an average accuracy of 80.1%. They also found that dynamic Bayesian networks (DBNs) gave a better performance than static Bayesian network models. Further, blink frequency and eye fixation measures were particularly indicative of cognitive task workload in structured experimental data. Building on the previous simulator based work, Liang used a hierarchical layered algorithm in [18], which incorporated both a DBN and a supervised clustering algorithm, to identify feature behaviors when drivers were in different cognitive states. Three groups of performance measures were used at the lowest level of this algorithm: (1) eye movement temporal measures (blink frequency, fixation duration, etc.), (2) eye movement spatial measures (spatial location of gaze in x, y, z),

and (3) driving performance measures (steering error, steering wheel standard deviation, lane position standard deviation) that were summarized across 30-second time windows, with no overlap between windows. Liang interpreted a sequential analysis as indicating that from a risk state identification perspective, it is not necessary to detect cognitive distraction if visual distraction is present as the latter dominates. A recent work [39] compared alternate SVM based classification approaches in a simulation context with experimentally defined periods of visual-manual, cognitive, and combined distraction. A "two-stage" classifier first considered visual-manual distraction and then detecting dual or cognitive distraction states was evaluated against a "direct-mapping" classifier developed to identify all distraction states at the same time. Advantages and limitations to both approaches appeared. Liang's [18] work is also relevant to the current effort in that it considers issues related to applying detection algorithms to naturalistic data.

Drowsiness and Impairment
Detection of driver arousal from blink rates, eye movement, and gaze patterns has received considerable attention in the simulated context and on small on-road datasets over two decades. A 10 year old survey paper [35] on driver fatigue detection covers the features of eye and eyelid movement that have continued to be used in papers that followed it. To the best of our knowledge, these features have not yet been proven to be robust to the highly variable naturalistic driving conditions, perhaps due to the costs and challenges associated with evaluating algorithms that require the collection of large driver-facing video datasets. The drowsiness detectors that have been implemented in many commercial vehicles have relied instead on measures of vehicle dynamics and driving performance [9, 30].

GLANCE MODEL AND PREDICTION APPROACH
Each six-second driving epoch contains the gaze region and a timestamp at the beginning of the epoch. Following this tuple is an arbitrary number of similar tuples marking the macro-glance transitions and their associated timestamps. These "glance transitions" refer to the moments in time when, based on the frame-by-frame annotations, the driver's gaze changed from one region to another. "Glance transitions" are event-based (see discrete event simulation [10]) in that they do not contain any self-transitions and only include changes of state. The duration of a glance is encoded in the difference of the timestamps of adjacent transitions.

For the purpose of modeling both glance transitions and durations as a Hidden Markov Model (HMM), we discretize the sequence of "glance transitions" into 25 state samples (spaced 250 milliseconds apart). By definition, the resulting sequence of states allow for self-transitions. The probability of such self-transitions form a simple model of state duration that was evaluated to be sufficient in this context. Explicit modeling of state duration for HMMs is an active area of research [38] and would be an effective extension to the model used in this work if epochs of longer and non-uniform durations were considered. The sampling rate of 4 Hz for the 6-second epochs was

determined to be the lowest-resolution sampling that had below 1% information loss over the original data. The result is that each six-second epoch of "glance transitions" is reduced to a sequence of 25 macro-glance states and induced state transitions.

For classification, a fixed-length sequence of discrete values can be viewed as a categorical feature vector input to a traditional classifier. We investigated this approach using parameter grid search of Random Forest and SVM classifiers, both of which resulted in worse performance than what is reported in the "Dataset and Results" section. The better performing approach was to model the temporal structure of the sequence using a classic hidden Markov model (HMM). Each of the gaze regions in the sequence are modeled as the discrete observation of the HMM. These observations can take on 8 values: (1) rearview mirror, (2) center stack, (3) eyes closed, (4) interior object, (5) right, (6) forward, (7) instrument cluster, and (8) left. An important point about this approach is that distinct macro-glance duration is not explicitly modeled. The explicit-duration hidden semi-Markov model (HSMM) [14] was evaluated for its ability to model the sticky dynamics of each state. However, this approach did not perform well. We believe that this is due to the limited and uniform length of each training sequence (see the "Conclusion" section for discussion of future work that proposes further investigation of this kind of explicit duration modeling).

As described in the "Dataset and Results" section, each prediction question is modeled a binary classification problem. One HMM is trained per class. The number of hidden states in the HMM is set to 8, which does not correspond to any directly identifiable states in the driving context. Instead, this parameter was programmatically determined to maximize classification performance. One HMM is constructed for each of the two classes in the binary classification problem. The HMM model parameters are learned using the GHMM implementation of the Baum-Welch algorithm [31]. This process uses 80% of the sequences from each of the two classes. As shown in Table 1, the classes are often unbalanced. In order to balance the training set, the minority class is over-sampled using the SMOTE algorithm [5].

The result of the training process are two HMM models. Each model can be use to provide a log-likelihood of an observed sequence. The HMM-based binary classifier then takes a 25-observation sequence, computes the log-likelihood from each of the two HMM models, and returns the class associated with the maximum log-likelihood.

DATASET AND RESULTS
The 100-Car Naturalistic Driving Study dataset includes approximately 2,000,000 vehicle miles, almost 43,000 hours of data, 241 primary and secondary drivers, 12 to 13 months of data collection for each vehicle, and data from a highly capable instrumentation system including five channels of video and vehicle kinematics [6]. Our work uses 4,816 six-second baseline driving epochs randomly selected from this dataset. Each epoch was manually annotated for macro-glances based on the video of the driver's face. This annotation serves as the

Binary Classification Problem	Type	Accuracy (Average)	Accuracy (St. Dev.)	Class 1 Size (Epochs)	Class 2 Size (Epochs)
Weather (Clear vs Raining)	Environment	51.6%	2.7%	4,261	345
Behavior (Speeding)	Behavior/State	52.3%	7.4%	3,497	101
Seatbelt (Yes vs No)	Environment	55.8%	2.4%	4,101	601
Traffic Density (Low vs Medium)	Environment	56.2%	0.4%	2,385	2,255
Traffic Divider (Present vs Not Present)	Environment	56.6%	1.8%	3,102	1,423
Travel Lanes (2 or Less vs 3 or More)	Environment	57.7%	1.2%	2,725	1,975
Alignment (Straight vs Curve)	Environment	57.7%	2.6%	4,186	519
Age (Young vs Mature)	Demographic	58.3%	1.9%	1,657	1,188
Traffic Density (Medium vs High)	Environment	58.6%	4.0%	2,255	176
Behavior (Following Too Closely)	Behavior/State	59.1%	2.3%	3,497	871
Lighting (Day vs Night with Light)	Environment	59.2%	1.7%	3,257	1,059
Gender (Male vs Female)	Demographic	59.7%	0.6%	2,514	1,773
Surface Condition (Wet vs Dry)	Environment	60.3%	1.9%	4,309	452
Age (Young vs Middle)	Demographic	60.7%	0.9%	1,657	1,442
Traffic Density (Low vs High)	Environment	61.2%	3.9%	2,385	176
Locality (Rural vs Interstate)	Environment	61.9%	0.8%	1,362	1,298
Locality (City vs Interstate)	Environment	62.6%	2.6%	1,555	1,298
Lighting (Night with vs without Light)	Environment	63.8%	2.2%	831	456
Age (Middle vs Mature)	Demographic	63.8%	1.5%	1,442	1,188
Locality (City vs Rural)	Environment	63.8%	0.9%	1,555	1,362
Traffic Light/Sign (Present vs Not Present)	Environment	64.0%	1.9%	4,252	377
Lighting (Day vs Night without Light)	Environment	66.6%	1.9%	3,257	456
Near an Intersection (Yes vs No)	Environment	70.9%	3.6%	4,025	764
Distraction (Talking)	Behavior/State	75.4%	2.0%	1,330	575
Behavior (Failed to Signal)	Behavior/State	75.5%	1.5%	3,497	247
Distraction (Fatigue)	Behavior/State	80.4%	3.1%	1,330	181
Distraction (Adjusting Radio)	Behavior/State	88.3%	2.1%	1,330	201

Table 1: This table answers the central question posed by this work: what aspects of the driver and driving environment can be predicted using a short sequence macro-glances? Each row specifies the binary classification problem, the variable type, accuracy mean and standard deviation, and the number of 6-second epochs associated with each glance. The rows are sorted according to average classification accuracy in ascending order.

training and evaluation variables for each of the binary classification tasks in the "Binary Classification Performance" section.

Baseline Epoch Dataset
The 100-car study was the first large-scale naturalistic driving study of its kind [6, 17] and the forerunner of the much larger and subsequent SHRP2 naturalistic study. As such, the 100-car study was intended to develop the instrumentation, methods, and procedures for the SHRP2 and to offer an opportunity to begin to learn about how crashes develop, arise, and culminate based on recording of the pre-crash period (which had not been possible prior to the development of methods used in the naturalistic 100-car study).

From the data that were acquired during the 100-car study, two databases were constructed: (1) an event database, and

(2) a baseline database. The event database was comprised of epochs of driving that ended with a conflict. Conflicts were classified at four levels of severity: crash, near-crash, crash-relevant, and proximity-type conflicts. These event epochs were each 6 seconds long âĂŞ (consisting of 5 seconds prior to a precipitating event and 1 second after). After data-acquisition, human analysts performed detailed extraction and coding of data that had been recorded during the study for each 6 s period (including frame-by-frame analysis of glance behavior). In addition, if a secondary task was underway by a driver during this period, analysts coded it, and information about it.

The baseline database was constructed of 20,000 epochs – also each 6 seconds long. These baseline epochs were ones in which the vehicle maintained a velocity over 5 mph âĂŞ and in which driving occurred without incident (without any conflict occurring). Eye glance analyses were conducted on 5,000 of these baseline epochs. Baseline epochs were selected at random from all recorded data (excluding event data) – and this selection did not make use of any kinematic triggers. The 6-second length of these epochs was chosen to match the length of event epochs. Event variables such as "precipitating factor" and "evasive maneuver" (which were coded for event epochs) were not coded for baseline epochs âĂŞ since no conflict occurred within them. While the baseline epochs are free from safety-critical events (i.e., do not contain crashes, near-crashes, or incidents), these epochs of "just driving" nevertheless are rich records of behaviors that are undertaken by ordinary drivers on real roads during everyday driving. This makes them an excellent source of data for the work reported here.

The number of baseline epochs selected from each vehicle for the baseline database was determined by each vehicle's involvement in crash, near-crash, and incident epochs in the event database. A stratified proportional sample of baseline epochs was constructed such that vehicles which were involved in more conflicts, also contributed more baseline epochs to the baseline database. This was done to create the required basis for a case-control design needed for odds-ratio calculations that were planned for subsequent analyses on the dataset.

From the 100-car study, a specific dataset was prepared and made accessible to the scientific community for analysis. It may be downloaded (along with documentation) from [1]. This database contains only de-identified data (i.e., no video data are available).

Binary Classification Performance
We evaluate the degree to which discriminative signal is present in 6-second bursts of macro-glances for the purpose of predicting the following variables. We provide a brief description of each variable and the number of categorical values considered.

- Driving Environment

    - **Proximity to an Intersection (2 values):** A vehicle is at or close to an intersection.

    - **Lighting (3 values):** Daylight or evening, with the latter case considering with and without light.
    - **Traffic Sign (2 values):** Presence of a traffic light or stop sign.
    - **Locality (3 values):** Rural, interstate, and city.
    - **Traffic Density (3 values):** Low, medium, or high. This level is based entirely on number of vehicles, and the ability of the driver to select the driving speed.
    - **Surface Condition (2 values):** Wet or dry.
    - **Weather (2 values):** Clear or rain.
    - **Alignment (2 values):** Geographic curvature of the road: straight or curved.
    - **Travel Lanes (2 values):** 2.5 is the threshold. The two categories are "≤ 2" and "≥ 3".
    - **Traffic Divider (2 values):** Presence or absence of a median divider.
    - **Seatbelt (2 values):** Wearing or not wearing a seatbelt.

- Driver Demographics

    - **Age (3 values):** Young, middle, or mature. 23.5 is the threshold between young and middle. 40.5 is the threshold between middle and mature. Selected for dataset balance not behavioral profile.
    - **Gender (2 values):** Female or male.

- Driver State and Behavior

    - **Behavior (4 values):** Following too closely, failed to signal, speeding, or none.
    - **Distraction (4 values):** Adjusting radio, fatigue, talking, or not distracted.

These variables took on more values than those listed above, but the values were pruned in two ways. First, values for distraction that were directly related to glance were removed. Obviously, 100% accuracy can be achieved in predicting glance region from glance region, so we are only interested in predicting driver state that does not directly relate to glance. Second, we only considered values that were well-represented in the data. The threshold was 100 epochs. In some cases, the categorical values were combined. For example, age was collapsed into three groups: young, middle, and mature. The partitioning was performed in a way that the numbers of epochs associated with each value was balanced.

For variables that take on more than 2 values, we reduce the problem to a binary classification one between all pairs of values. This allows us to explore the discriminating potential of macro-glances with regard to variables that take on 2, 3, or 4 values. For behavior and distraction variables, we only consider the pairing of values with the baseline state of "none" and "not distracted", respectively.

Table 1 shows the result of applying the HMM-based classification method in the "Glance Model and Prediction Approach" section to the binary classification problems associated with the variables listed above. This table contains

(a) Glance transition matrix for the first binary prediction class: "not distracted".

(b) Glance transition matrix for the first binary prediction class: "adjusting radio".

(c) The difference between the transition matrix for the two classes of "not distracted" and "adjusting radio".
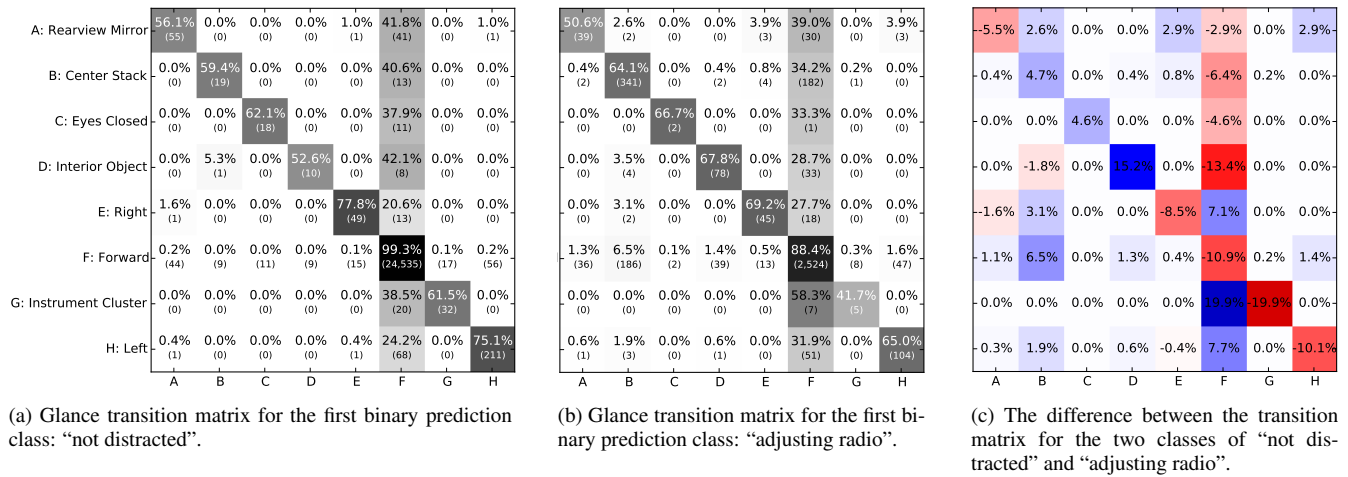
Figure 1: An example of two glance transition matrices (first two subfigures) and their difference (third subfigure) that illustrates the discriminating characteristics of the glance dynamics based on which the two HMM models can make binary predictions. The y-axis is the "transition from" region and the x-axis is the "transition to" region. In the first two subfigures, the percentages designate the probability of transitions and the values in parentheses show the absolute number of times those transitions appear in the dataset. The third subfigure shows the difference in probabilities between the first two subfigures. Blue indicates a positive difference, red indicates a negative difference. The intensity of color fill in each matrix cell, across all three subfigures, indicates magnitude of the values associated with the transition in that cell. The transition matrices and their differences for each of the binary classification problems considered in this paper are provided in the appendix.

the answer to the question posed in the title of this paper. As expected, the problem of predicting anything about the driver or driving environment from just 6 seconds of macro-glances is very difficult, because both the duration of the sequence is too short and the resolution of the captured dynamics is too coarse. Nevertheless, accuracy of above 75% is achieved for the prediction of the following four driver behaviors and states: talking, fatigue, radio-tuning activity, and failure to signal. Also, 70.9% accuracy is achieved in predicting whether or not the vehicle is approaching an intersection. To provide intuition as to why this prediction works at all, Fig. 1 provides a visualization for the dynamics of the radio-tuning activity classification problem. The first two subfigures show the glance transition matrices for the "not distracted" and "adjusting radio" classes. The third subfigure shows the difference in transition probabilities between the first two subfigures. These matrices show that there are significant aggregate differences in macro-glance transitions for the two classes. The results in Table 1 show that these differences can be detected from six-second epochs for several variables critical for the design of intelligent driver assistance systems.

Framework for Gaze-Based HMI in the Car
The focus of this paper is to begin to answer the open question whether temporal patterns of macro-glances in the car contain sufficient discriminative signal to predict the state of the driver, the car, and the world. Our approach naturally leads to a real-time application of its various detectors in a vehicle. Fig. 2 shows an illustrative example of a driver approaching an intersection while operating a Tesla vehicle in Autopilot mode. Fig. 3 proposes how such a sequence of videos would be processing through time. In this proposed system, a sliding window of 6 seconds is used to make a series of binary

predictions. Each of the predictions along with an estimated confidence are exposed through a CAN network API. HMI systems in the car are then able to alter their communication strategy with the driver and the external environment based on the predictions each system listens to for cases when the prediction confidence exceeds a minimum threshold.

In fact, an HMI system may utilize multiple sensor streams, only part of which would be the sensors derived from macro-glances. This is grounded in the design of joint cognitive systems envisioned and developed over the previous 3 decades [36, 13, 26]. For example, for the case in Fig. 2, the prediction may be that the driver's macro-glances are not indicative of proximity to an intersection even though based on the GPS coordinates reported on the CAN network, the car is in fact approaching intersection. This, in combination with the fact that the Tesla is operating under Autopilot and is thus driving itself, can be used by the car's external signaling system to infer that the driver is not paying attention and is unaware of the intersection. This information can then be conveyed to pedestrians and other vehicles, so that they make their movement decisions with a higher degree of caution.

CONCLUSION
This work asks what can and cannot be predicted from short bursts of driver macro-glances. We consider a representative sample of 4,816 annotated six-second epochs of driving from a 100-car naturalistic study. The variables under consideration fall into three categories: driving environment, driver behavior/state, and driver demographic characteristics. We form binary-classification problems from all the well-represented variables available in the dataset and model regularly-sampled macro-glances as a hidden Markov model

Figure 2: Illustrative sequence of snapshots collected in a Tesla vehicle of a driver's macro-glances. A time of $t = 0$ is marked as the point at which the driver can be annotated as being near an intersection.
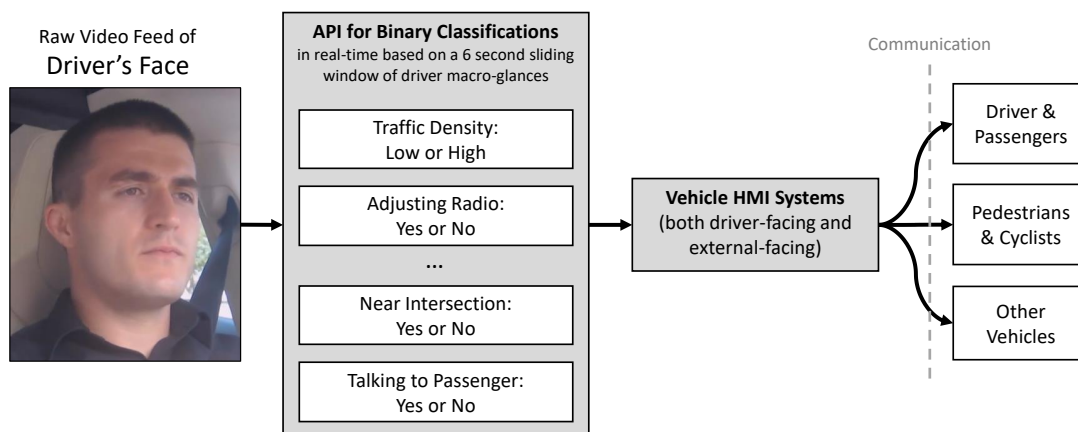

Figure 3: A framework for a system that processes (in real-time) the video of the driver's face and exposes an API to the HMI systems operating in the car that can appropriately alter their communication strategy with the driver, vehicle passengers, and the external driving environment based on the binary sensor signals.

for each class to make the binary prediction. The results show that radio-tuning activity, fatigue state, failure to signal, talking, and proximity to an intersection can be predicted with 70.9% to 88.3% accuracy. Based on these results, the general conclusion of this work is that macro-glances can be part of a multi-sensor system for predicting external environment factors, but on its own is only sufficient to predict a limited but important set of variables related to driver behavior and state. Nevertheless, significant improvements in accuracy may be achievable through further development of the underlying algorithmic approach. To this end, future work will investigate whether other approaches that capture temporal dynamics in the data, such as Hidden Semi-Markov Models (HSMM) [14] or Recurrent Neural Networks (RNN) [32], may perform better than HMMs, in which case macro-glances alone may be used as the basis for environment, behavior, state, and demographic prediction in future real-time driver assistance systems. Furthermore, using macro glance epochs of heterogeneous duration for training and evaluation may result in significant increases in prediction accuracy due to the fact that some environmental or behavioral factors may reveal themselves on different time-scales. For example, detection of talking may only need 1-2 seconds of macro-glances, while the detection of rural versus urban environmental conditions may requires an epoch of 10-20 seconds.

Acknowledgment

REFERENCES

1. 2016. VTTI 100-Car Data. http://forums.vtti.vt.edu/. (2016).

2. National Highway Traffic Safety Administration and others. 2012. Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices. *Washington, DC: National Highway Traffic Safety Administration (NHTSA), Department of Transportation (DOT)* (2012).

3. Linda Angell, Miguel A Perez, and Susan A Soccolich. 2015. Identification of Cognitive Load in Naturalistic Driving. *NSTSCE; 15-UT-037* (2015).

4. Cheryl A Bolstad, Haydee M Cuevas, Jingjing Wang-Costello, Mica R Endsley, and Linda S Angell. 2008. Measurement of situation awareness for automobile technologies of the future. (2008).

5. Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* (2002), 321–357.

6. Thomas A Dingus, Sheila G Klauer, Vicki L Neale, A Petersen, SE Lee, JD Sudweeks, MA Perez, J Hankey, DJ Ramsey, S Gupta, and others. 2006. *The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment*. Technical Report.

7. Tiziana D'Orazio, Marco Leo, Cataldo Guaragnella, and Arcangelo Distante. 2007. A visual approach for driver inattention detection. *Pattern Recognition* 40, 8 (2007), 2341–2355.

8. Mica R Endsley. 1995. Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37, 1 (1995), 32–64.

9. Azim Eskandarian and Ali Mortazavi. 2007. Evaluation of a smart algorithm for commercial vehicle driver drowsiness detection. In *2007 IEEE intelligent vehicles symposium*. IEEE, 553–559.

10. George Fishman. 2013. *Discrete-event simulation: modeling, programming, and analysis*. Springer Science & Business Media.

11. Lex Fridman, Joonbum Lee, Bryan Reimer, and Trent Victor. 2016, In Print. Owl and Lizard: Patterns of Head Pose and Eye Pose in Driver Gaze Classification. *IET Computer Vision* (2016, In Print).

12. Driver Focus-Telematics Working Group and others. 2006. Statement of principles, criteria and verification procedures on driver interactions with advanced in-vehicle information and communication systems. *Alliance of Automotive Manufacturers* (2006).

13. Erik Hollnagel and David D Woods. 1983. Cognitive systems engineering: New wine in new bottles. *International Journal of Man-Machine Studies* 18, 6 (1983), 583–600.

14. Matthew J. Johnson and Alan S. Willsky. 2013. Bayesian Nonparametric Hidden Semi-Markov Models. *Journal of Machine Learning Research* 14 (February 2013), 673–701.

15. Rami N Khushaba, Sarath Kodagoda, Sara Lal, and Gamini Dissanayake. 2011. Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm. *Biomedical Engineering, IEEE Transactions on* 58, 1 (2011), 121–131.

16. Katja Kircher, Christer Ahlstrom, and Albert Kircher. 2009. Comparison of two eye-gaze based real-time driver distraction detection algorithms in a small-scale field operational test. In *Proc. 5th Int. Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*. 16–23.

17. Sheila G Klauer, Thomas A Dingus, Vicki L Neale, Jeremy D Sudweeks, and David J Ramsey. 2006. *The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data*. Technical Report. National Highway Traffic Safety Administration.

18. Yulan Liang. 2009. Detecting driver distraction. (2009).

19. Yulan Liang, John Lee, and Michelle Reyes. 2007. Nonintrusive detection of driver cognitive distraction in real time using Bayesian networks. *Transportation Research Record: Journal of the Transportation Research Board* 2018 (2007), 1–8.

20. Yulan Liang, John D Lee, and Lora Yekhshatyan. 2012. How dangerous is looking away from the road? Algorithms predict crash risk from glance patterns in naturalistic driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 54, 6 (2012), 1104–1116.

21. Yulan Liang, Michelle L Reyes, and John D Lee. 2007. Real-time detection of driver cognitive distraction using support vector machines. *IEEE transactions on intelligent transportation systems* 8, 2 (2007), 340–350.

22. Dominique Lord and Fred Mannering. 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice* 44, 5 (2010), 291–305.

23. Jannette Maciej and Mark Vollrath. 2009. Comparison of manual vs. speech-based interaction with in-vehicle information systems. *Accident Analysis & Prevention* 41, 5 (2009), 924–930.

24. R Martins and JM Carvalho. 2015. Eye blinking as an indicator of fatigue and mental load a systematic review. *Occupational Safety and Hygiene III* (2015), 231.

25. Joel C McCall and Mohan M Trivedi. 2006. Video-based lane estimation and tracking for driver assistance: survey, system, and evaluation. *Intelligent Transportation Systems, IEEE Transactions on* 7, 1 (2006), 20–37.

26. David Bryan Miller and Wendy Ju. 2015. Joint cognition in automated driving: Combining human and machine intelligence to address novel problems. In *2015 AAAI Spring Symposium Series*.

27. Mauricio Munoz, Bryan Reimer, Joonbum Lee, Bruce Mehler, and Lex Fridman. 2016. Distinguishing Patterns in Drivers' Visual Attention Allocation Using Hidden Markov Models. *Transportation Research Part F: Traffic Psychology and Behaviour* 43 (2016), 90–103. DOI:http://dx.doi.org/10.1016/j.trf.2016.09.015

28. Timo Pech, Philipp Lindner, and Gerd Wanielik. 2014. Head tracking based glance area estimation for driver behaviour modelling during lane change execution. In *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*. IEEE, 655–660.

29. Bryan Reimer, Bruce Mehler, Ian Reagan, David Kidd, and Jonathan Dobres. 2016. Multi-modal demands of a smartphone used to place calls and enter addresses during highway driving relative to two embedded systems. *Ergonomics* 59, 12 (2016), 1565–1585.

30. David Sandberg, Torbjörn Akerstedt, Anna Anund, Göran Kecklund, and Mattias Wahde. 2011. Detecting driver sleepiness using optimized nonlinear combinations of sleepiness indicators. *IEEE Transactions on Intelligent Transportation Systems* 12, 1 (2011), 97–108.

31. Alexander Schliep, Benjamin Georgi, Wasinee Rungsarityotin, I Costa, and A Schonhuth. 2004. The general hidden markov model library: Analyzing systems with unobservable states. *Proceedings of the Heinz-Billing-Price* 2004 (2004), 121–135.

32. Jürgen Schmidhuber. 2015. Deep learning in neural networks: An overview. *Neural Networks* 61 (2015), 85–117.

33. Sayanan Sivaraman and Mohan Manubhai Trivedi. 2013. Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis. *Intelligent Transportation Systems, IEEE Transactions on* 14, 4 (2013), 1773–1795.

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

35. Qiong Wang, Jingyu Yang, Mingwu Ren, and Yujie Zheng. 2006. Driver fatigue detection: a survey. In *Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on*, Vol. 2. IEEE, 8587–8591.

36. David D Woods. 1985. Cognitive technologies: The design of joint human-machine cognitive systems. *AI magazine* 6, 4 (1985), 86.

37. Guosheng Yang, Yingzi Lin, and Prabir Bhattacharya. 2010. A driver fatigue recognition model based on information fusion and dynamic Bayesian network. *Information Sciences* 180, 10 (2010), 1942–1954.

38. Shun-Zheng Yu. 2010. Hidden semi-Markov models. *Artificial Intelligence* 174, 2 (2010), 215–243.

39. Yu Zhang and David Kaber. 2016. Evaluation of strategies for integrated classification of visual-manual and cognitive distractions in driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society* (2016), 0018720816647607.

40. Yilu Zhang, Yuri Owechko, and Jing Zhang. 2004. Driver cognitive workload estimation: A data-driven perspective. In *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference on*. IEEE, 642–647.