What Does Visual Gaze Attend to during Driving?

Mohsen Shirpour, Steven S. Beauchemin and Michael A. Bauer
Department of Computer Science, The University of Western Ontario, London, ON, N6A-5B7, Canada

Keywords: Vanishing Points, Point of Gaze (PoG), Eye Tracking, Gaussian Process Regression (GPR).

Abstract: This study aims to analyze driver cephalo-ocular behaviour features and road vanishing points with respect to vehicle speed in urban and suburban areas using data obtained from an instrumented vehicle's eye tracker. This study utilizes two models for driver gaze estimation. The first model estimates the 3D point of the driver's gaze in absolute coordinates obtained through the combined use of a forward stereo vision system and an eye-gaze tracker system. The second approach uses a stochastic model, known as Gaussian Process Regression (GPR), that estimates the most probable gaze direction given head pose. We evaluated models on real data gathered in an urban and suburban environment with the RoadLAB experimental vehicle.

1 INTRODUCTION

The human visual system collects about 90% of the information that is needed to adequately perform driving tasks (Sivak, 1996). Driver gaze has been studied for many years in driving simulators and real driving environments. It has been demonstrated that driver gaze direction in relation to the surrounding driving environment is predictive of driver maneuvers (Khairdoost et al., 2020). In addition to these results, our aim is to elucidate the rules that govern driver gaze with respect to the characteristics of vehicular dynamics. In particular, this contribution reports on our investigation of the relationship that exists between gaze behaviour, vanishing points, and vehicle speed.

1.1 Literature Survey

Driver visual attention plays a prominent role in intelligent Advanced Driver Assistance Systems (i-ADAS). Some driver monitoring systems utilize the driver's head pose and eyes to evaluate the driver's gaze-direction and zone (Jha and Busso, 2018; Shirpour et al., 2020). We recently presented a stochastic model that derives gaze direction from head pose data provided by a contactless gaze tracking system (Shirpour et al., 2020). This model computes a probabilistic visual attention map that estimates the probability of finding the actual gaze over the stereo system's imaging plane, with a Gaussian Process Regression (GPR) technique. Subsequently, we proposed a deep learning model to predict driver eye fixation according to driver's visual attention (Shirpour. et al., 2021). In addition, other contributions use the direction of gaze to detect 2D image gaze regions (Shirpour et al., ; Zabihi et al., 2014). Others have defined a framework that uses the 3D Point of Gaze (PoG) and Line of Gaze (LoG) in absolute coordinates for similar purposes (Kowsari et al., 2014).

In other works, the driver's attentional visual area was modelled as intersection of the elliptical region formed by the cone emanating from the eye position with the LoG as its symmetrical axis along its length, with the imaging plane of the forward stereoscopic vision system installed in the experimental vehicle, as depicted in Figure 1. Using this mechanism, several authors were able to estimate the driver's most probable next maneuver some time before it occurred (Khairdoost et al., 2020; Zabihi et al., 2017). Their evaluation showed a strong relationship between driver gaze behaviour and maneuvers.

In general, a driver concentrates on parts of the driving scene that contain some objective and subjective elements. Objective elements are obtained with bottom-up approaches that consider features extracted from the driving environment such as traffic-related objects. On the other hand, subjective elements are obtained with top-down approaches and are attributed to a driver's internal factors, such as experience or intention (Deng et al., 2016). Top-down strategies provide insight into what a driver's gaze could be fixated on while driving.

Shirpour, M., Beauchemin, S. and Bauer, M.
What Does Visual Gaze Attend to during Driving?.
DOI: 10.5220/0010440204650470
In *Proceedings of the 7th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2021)*, pages 465-470
ISBN: 978-989-758-513-5

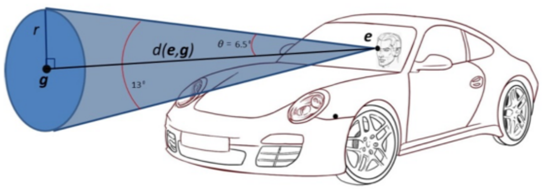


Figure 1: The attentional area is defined as intersection of the elliptical region formed by the cone cone emanating from the eye position with the LoG as its symmetrical axis along its length, and the imaging plane of the forward stereoscopic vision system.

1.2 Human Vision System

The human visual field affords a remarkably broad view of the world, in the range of 90° to the left and right, and more than 60° above and below the gaze (Wolfe et al., 2017). Information within 2° of the gaze is processed in foveal vision. More broadly, parafoveal vision covers up to 6° of visual angle (Engbert et al., 2002). This implies that the existing information in the parafovea is combined with that from the fovea. The information from the fovea is clearer when compared with the information present in the parafovea (Kennedy, 2000). Together, the foveal and parafoveal areas are known as the central visual field, where objects are clearly and sharply seen and used to perform most activities (Wolfe et al., 2017).

1.3 Experimental Vehicle

Our research vehicle is equipped with instruments that capture driver-initiated vehicular actuation and relate the 3D driver gaze direction on the imaging plane of the forward stereoscopic vision system. The vehicle was used to gather data sequences from 16 different test drivers on a pre-determined 28.5km route within the city of London, Ontario, Canada. 3TB of driving sequences were recorded. The data contains significant driving information, including forward stereo imaging and depth, 3D PoG and head pose, and vehicular dynamics obtained with the OBDII CANBus interface. Image and data frames are collected at a rate of 30Hz. The vehicular instrumentation consists of a non-contact infrared remote gaze and head pose tracker, with two cameras mounted on the vehicle dashboard, operating at 60Hz. This instrument provides head movement and pose, eye position, and gaze direction within its own coordinate system. A forward stereoscopic vision system is located on the vehicle's roof to capture frontal view information such as dense stereo depth maps at 30 Hz (See Figure 2). Details concerning this instrumentation are available in (Beauchemin et al., 2011). The sum of our data was recorded with the RoadLAB software system, as shown in Figure 3.

2 METHODOLOGY

This Section presents two models for describing driver gaze visual attention in the forward stereo imaging system. Section 2.1 addresses the calibration procedure applied to provide the Point of Gaze (PoG) onto the imaging plane of the forward stereo system. We introduce a Gaussian Process Regression (GPR) that estimates the probability of gaze direction according to driver head pose in Section 2.2. Section 2.3 describes the technique we employ to locate vanishing points from the stereoscopic imagery.

2.1 Projection of PoGs Onto Stereo System

The calibration process brings the eye tracker data into the coordinate system of the forward stereoscopic vision system. We used a cross-calibration technique developed in our laboratory to transform the 3D driver gaze expressed in the eye tracker reference frame to that of forward stereoscopic vision system (Kowsari et al., 2014). This calibration process is defined as follows:

- *Salient Points Extraction:* A sufficient number of salient points are extracted from the stereoscopic imagery (around 20 points provide sufficient data)
- *Depth Estimation:* The driver's eye fixates on preselected salient points for a short period (about 2 seconds). The depth estimate of the salient point, the gaze vector, and the position of the eye centre are recorded.
- *Estimation of Rotation and Translation Matrices:* The process estimates the rigid body transformation between the reference frame of the stereoscopic system and the remote eye tracker. The elements composing this transformation are known as extrinsic calibration parameters.
- *Gaze Projection Onto the Imaging Plane of Stereoscopic System:* The LoG, expressed in eye tracker coordinates, is projected onto the imaging plane of the stereo system using the extrinsic calibration parameters. The PoG is determined as the location where the LoG intersects with a valid depth estimate within the reference frame of the stereo vision system.

Figure 2: **(left):** Stereo vision system located on the vehicle's roof; **(centre):** infrared gaze tracker; **(right:)** FaceLAB system interface.

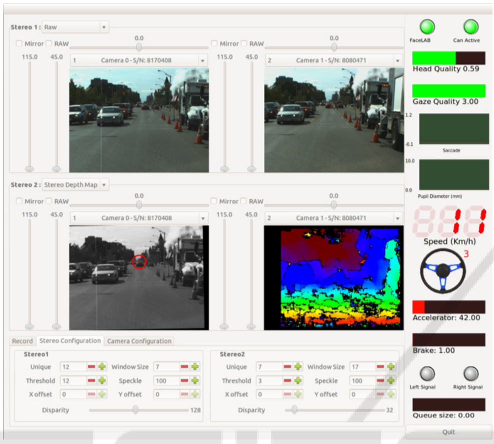


Figure 3: RoadLab software systems: The on-board system displays frame sequences with depth maps, dynamic vehicle features, and eye-tracker data.

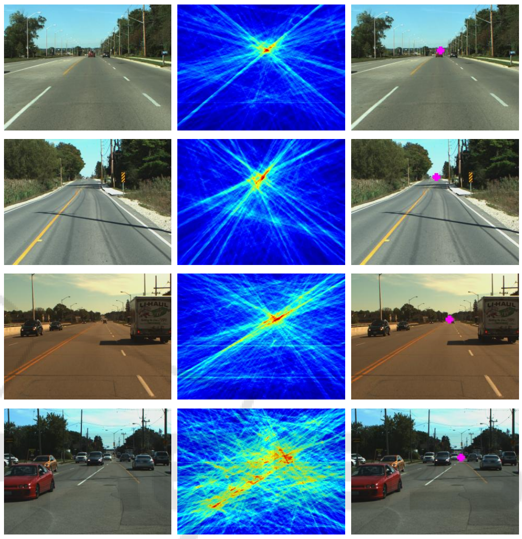


Figure 4: Examples of vanishing points (from left to right:) input frames, voting map, and detected vanishing points.

2.2 Gaussian Process Regression

Technically, direct use of gaze is complicated by the fact that eyes may exhibit rapid saccadic movements resulting in difficulties for assessing the correct image area corresponding to a driver's visual attention. Our laboratory proposed another model to alleviate this problem by approximating the 3D gaze from the 3D head pose, as the head does not experience saccadic movements.

In our recent research, instead of directly estimating the gaze, which depends on the driver's visual cognitive tasks, we introduced a stochastic model for representing driver visual attention. This model inherits the advantage of the Gaussian Process Regression (GPR) technique to estimate the probability of the driver's gaze direction according to head pose over the imaging plane of the stereo system. It establishes a confidence area within which the driver gaze is most likely contained. We have shown that drivers concentrate most of their attention on the 95% confidence interval region estimated from the head pose. We refer the reader to (Shirpour et al., 2020) for details on the GPR technique.

2.3 Vanishing Points

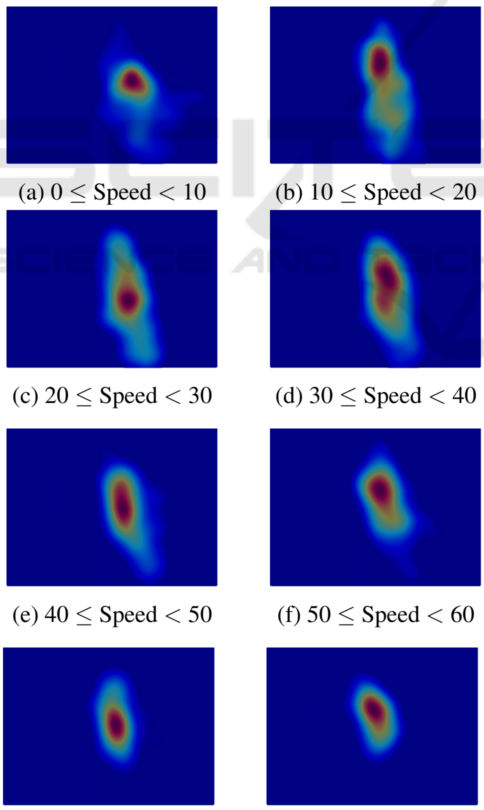
A vanishing point is the location on the image plane where two-dimensional perspective projections of mutually parallel lines in three-dimensional space appear to converge. The vanishing point plays an essential role in the prediction of driver eye fixations. In general, vanishing points are considered as guidance for predicting driver intent, as drivers mostly gaze at traffic objects near those points.

Available methods to detect the vanishing point are mainly edge, region, or texture-based models. Edge-based models are adequate when edge boundaries and lane markings are available within the driving scene. Region-based methods divide the driving view into path and non-path according to low-level features (colour, intensity, etc). These two types of models are suitable for structured roads. They experience difficulty with scenery involving unstructured or complex features.

Because the RoadLab dataset includes both structured and unstructured imaging elements, we adopted

Table 1: Data Description.

Seq#	0≤Speed <10	10≤Speed <20	20≤Speed <30	30≤Speed <40	40≤Speed <50	50≤Speed <60	60≤Speed <70	Speed≥70
Seq. 2	11530	2693	3181	4426	4475	3930	4371	2350
Seq. 8	8515	2556	2959	3297	3594	3679	2157	2543
Seq. 9	7756	2544	3263	4197	3131	3148	3169	2166
Seq. 10	7199	1538	2068	3912	4665	4200	3042	1211
Seq. 11	8008	1714	2425	3373	3417	3330	2954	887
Seq. 13	11545	1956	2098	2248	2447	2711	3528	2605
Seq. 14	4495	1123	1311	1986	2285	2442	1204	1448
Seq. 16	9056	2085	2440	3046	2874	3321	1241	1628



(a) $0 \leq$ Speed < 10 (b) $10 \leq$ Speed < 20 (c) $20 \leq$ Speed < 30 (d) $30 \leq$ Speed < 40 (e) $40 \leq$ Speed < 50 (f) $50 \leq$ Speed < 60 (g) $60 \leq$ Speed < 70 (h) Speed ≥ 70

Figure 5: Driver attention versus vanishing point with respect to speed. **a) to h):** As the speed increases, the driver gaze converges to the vanishing point.

a texture-based model proposed by (Moghadam et al., 2011). Their model is based on Gabor filters to estimate the local orientation of pixels. Figure 4 shows a sample of RoadLab frames with detected vanishing points.

3 ANALYSIS OF DRIVER ATTENTION

In this Section, we describe the preprocessing we applied to the RoadLAB dataset and provide an analysis of the results that were obtained.

3.1 Data Preparation

Our experimental vehicle relies on sensors and cameras to track its driver's behavioural features. The RoadLab software provided a confidence measure on the quality of its estimations of head pose and gaze. The head pose confidence measure ranged from 0 to 2, while the gaze quality metric ranged from 0 to 3. We considered the head pose and gaze as reliable elements when these metrics had a minimum value of 1 or higher for the head pose, and 2 or higher for the gaze. The PoGs that passed the quality metric thresholds were projected onto the forward stereo system for the 5 preceding consecutive frames. Table 1 provides the number of frames selected from test drivers according to vehicular speed.

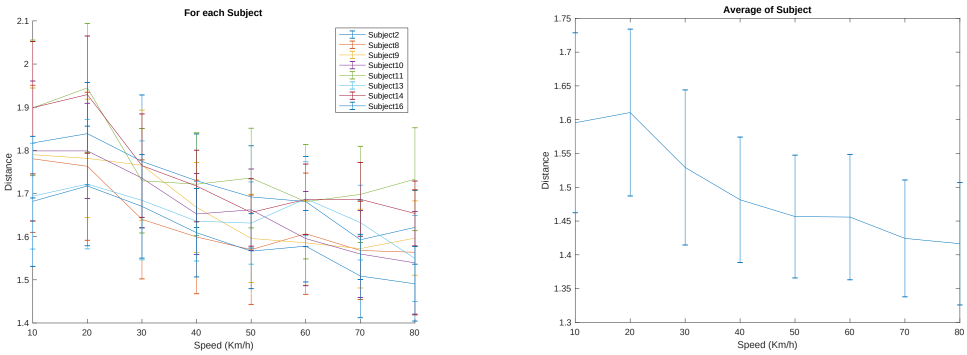


Figure 6: **Model A (Left):** Average and variance of distance from driver gaze fixation to vanishing point versus vehicle speed for each driver. **(right):** Average of all drivers.

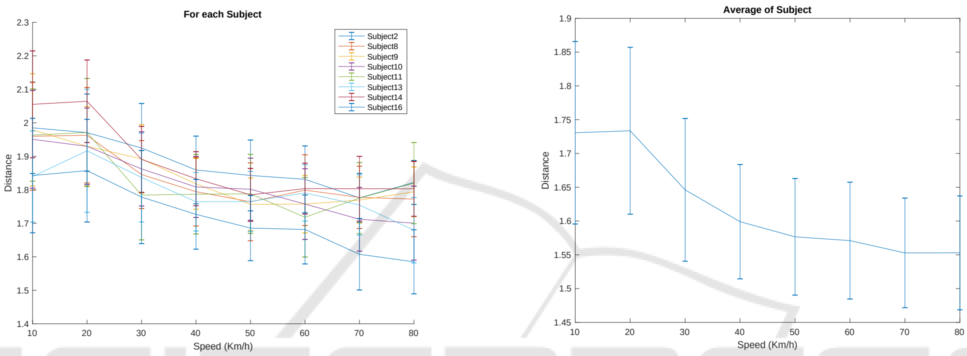


Figure 7: **Model B (Left):** Average and variance of distance from driver gaze fixation to vanishing point versus vehicle speed for each driver. **(right):** Average of all drivers.

3.2 Speed and Visual Attention Analysis

Our results show that drivers generally tend to concentrate their gaze on vanishing points created by the motion of the vehicle. Figure 5 illustrates the fact that the frequency of driver gaze fixations near the vanishing point is considerably higher than that of fixations on other image regions. This indicates that driver attention is more likely to fixate on traffic objects near the vanishing point. Also, Figure 5 illustrates how the gaze position changes at different vehicle speeds (for one particular driving sequence). When the vehicle speed smoothly increases from below 10 km/h to over 70 km/h, the gaze position rapidly converges to the vanishing point.

We estimated driver visual attention with two different models for gaze direction: model A which estimates the probability of driver gaze direction according to head pose, and model B which directly uses the 3D driver gaze in absolute coordinates. We measured the logarithmic distance of gazes from vanishing points and calculated the averages and variances of these distances for a range of vehicle speeds. As observed in Figures 6 and 7 the average distance of gaze fixations and vanishing points decreases significantly with an increase in vehicle speed. These results show that the drivers were more focused on vanishing points at high the vehicle speeds. The variance of gaze fixations at high vehicular speeds is significantly lower than that observed at lower speeds.

The human visual system is limited in the quantity of information it is able to process per time unit, and compensates by decreasing its visual field when the mass of elements to process in the spatial or temporal context increases. In driving circumstances, this generally occurs at high speeds, as the amount of available information per unit of time increases proportionally.

4 CONCLUSIONS

We analyzed driver gaze behaviour in relation to vanishing points with respect to increasing vehicular speeds with the RoadLab dataset obtained from an instrumented vehicle. This research investigated two models for driver gaze estimation. The first model estimated 3D point of gaze in absolute coordinate, while the second model used a probabilistic process to estimate the probability of driver gaze direction based on the head pose. For both models, the results clearly indicate that vanishing points attract driver gaze with increasing force at high vehicle speeds.

REFERENCES

Beauchemin, S., Bauer, M., Kowsari, T., and Cho, J. (2011). Portable and scalable vision-based vehicular instrumentation for the analysis of driver intentionality. *IEEE Transactions on Instrumentation and Measurement*, 61(2):391–401.

Deng, T., Yang, K., Li, Y., and Yan, H. (2016). Where does the driver look? top-down-based saliency detection in a traffic driving environment. *IEEE Transactions on Intelligent Transportation Systems*, 17(7):2051–2062.

Engbert, R., Longtin, A., and Kliegl, R. (2002). A dynamical model of saccade generation in reading based on spatially distributed lexical processing. *Vision Research*, 42(5):621–636.

Jha, S. and Busso, C. (2018). Probabilistic estimation of the gaze region of the driver using dense classification. In *International Conference on Intelligent Transportation Systems*, pages 697–702. IEEE.

Kennedy, A. (2000). Parafoveal processing in word recognition. *The Quarterly Journal of Experimental Psychology A*, 53(2):429–455.

Khairdoost, N., Shirpour, M., Bauer, M., and Beauchemin, S. (2020). Real-time maneuver prediction using lstm. *IEEE Transactions on Intelligent Vehicles*, 5(4):714–724.

Kowsari, T., Beauchemin, S., Bauer, M., Laurendeau, D., and Teasdale, N. (2014). Multi-depth cross-calibration of remote eye gaze trackers and stereoscopic scene systems. In *Intelligent Vehicles Symposium*, pages 1245–1250. IEEE.

Moghadam, P., Starzyk, J. A., and Wijesoma, W. (2011). Fast vanishing-point detection in unstructured environments. *IEEE Transactions on Image Processing*, 21(1):425–430.

Shirpour, M., Beauchemin, S., and Bauer, M. (2020). A probabilistic model for visual driver gaze approximation from head pose estimation. In *Connected and Automated Vehicles Symposium*, pages 1–6. IEEE.

Shirpour., M., Beauchemin., S. S., and Bauer., M. A. (2021). Driver's eye fixation prediction by deep neural network. In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4 VISAPP: VISAPP,*, pages 67–75. INSTICC, SciTePress.

Shirpour, M., Khairdoost, N., Bauer, M. A., and Beauchemin, S. S. Traffic object detection and recognition: A survey and an approach based-on the attentional visual field of driver. *IEEE Transactions on Intelligent Vehicles (in press)*.

Sivak, M. (1996). The information that drivers use: is it indeed 90% visual? *Perception*, 25(9):1081–1089.

Wolfe, B., Dobres, J., Rosenholtz, R., and Reimer, B. (2017). More than the useful field: Considering peripheral vision in driving. *Applied ergonomics*, 65:316–325.

Zabihi, S., Beauchemin, S., and Bauer, M. (2017). Real-time driving manoeuvre prediction using io-hmm and driver cephalo-ocular behaviour. In *Intelligent Vehicles Symposium*, pages 875–880. IEEE.

Zabihi, S., Beauchemin, S., De Medeiros, E., and Bauer, M. (2014). Frame-rate vehicle detection within the attentional visual area of drivers. In *Intelligent Vehicles Symposium*, pages 146–150. IEEE.