A Differential 2D Gaussian Ellipse-Based Eye Movement Analysis

Jiahao Zhang [1], Yang Liu [1], Xu Wen [2], Dengbo He [1]*

[1] *Thrust of Intelligent Transportation, The Hong Kong University of Science and Technology (Guangzhou), Guangdong, China*

[2] *Faculty of Liberal Arts and Social Sciences, The Education University of Hong Kong, Hong Kong, China*

As data-driven analysis methods powered by artificial intelligence have matured, research on visual attention prediction has advanced markedly. However, gaze-point data from eye-tracking devices are often characterized by high noise levels, limiting accuracy in representing real driver behavior. Thus, we proposed the Differential 2D Gaussian Ellipse (D2DGE) representation, which captures gaze distribution within a time window to reduce noise from devices or unconscious glances. To validate D2DGE, a Generative Adversarial Imitation Learning model was trained on both raw gaze data and the D2DGE data, and we assessed the similarity between the raw and generated data using the Kullback-Leibler divergence. Then, we compared the five parameters of D2DGE representation between novice and experienced drivers. Results show that the D2DGE data can better approximate the raw data and contains richer information than raw gaze data. The findings indicate that the D2DGE can be a promising alternative to describe gaze distribution during driving.

INTRODUCTIOG

Drivers' visual behavior is highly associated with driving safety (He, 2020). The visual channel is estimated to take over 90% of the information in driving tasks (Sivak, 1996). To quantitatively analyze the impact of visual cues on driving safety, the eye tracker has been widely adopted. By analyzing eye-tracking data, researchers are able to reveal drivers' attention allocation strategies during driving (Werneke, 2012), based on which, adaptive driver assistance systems, such as adaptive human-machine interfaces (HMI), can be developed. Thus, a number of research has been conducted to model and predict drivers' visual attention allocation strategies (Baee, 2021; Fu, 2023).

Gaze points are commonly utilized in eye-tracking research to estimate drivers' visual attention. Still, they represent only a fraction of overall visual perception (Ahlström, 2021), and can be noisy due to the fast eye movements in order to scan the environment, and the less-than-ideal tracking accuracies. Such characteristics can negatively impact the performance of data-driven visual attention predictive models (Gómez-Poveda, 2016). Specifically, most of the data-driven visual behavior predictions were based on discrete focal gaze points, which may suffer from a decrease in algorithm robustness and prediction accuracy.

As such, this paper proposes a driver visual model called the Differential 2D Gaussian Ellipse (D2DGE) to describe drivers' visual attention allocation with the distribution of multiple focal gaze points considered. D2DGE balances gaze noise across all directions without altering the spatial distribution of the original gaze data. Consequently, it outperforms raw gaze data in data-driven visual attention prediction algorithms. Statistical analysis was performed to compare the gazes from the eye-tracker (i.e., raw gaze) and D2DGE data to evaluate the invariance of D2DGE relative to raw gaze data in their two-dimensional spatial distribution. Further, given the vital role of visual attention prediction in human-machine interface (HMI) design and autonomous driving, we further generated visual attention based on the raw gaze data and D2DGE data using Generative Adversarial Imitation Learning (GAIL) (Ho, J., & Ermon, S., 2016) to check if D2DGE performs better in data-driven visual attention analysis than gaze data.

APPROACH

D2DGE data

As shown in Figure 1, inspired by the research on focal and peripheral vision (Larson, 2009), the D2DGE eye-tracking data representation was proposed as the k times of the sigma (k-sigma) range of the 2D Gaussian distribution that the gaze points follow within a sliding window, where the distance between the participant and the presented visual stimulus determines the magnitude of k. This enables a new eye-tracking data representation based on eye-tracker-recorded gaze points while eliminating noise in the recorded data. For an eye-tracking sequence with a window size of n and a sliding step size of m, a 2D Gaussian distribution is calculated for all gaze points within the window. The k-sigma range of the Gaussian distribution is then extracted to obtain an ellipse, representing the visual attention area of the driver during that time window. This ellipse is described by five parameters: $[\mu_x, \mu_y, s, r_{major}, \theta]$, where μ_x and μ_y represent the mean coordinates of the Gaussian distribution, s is the area of the ellipse calculated from the k-sigma range, r_{major} is the length of the ellipse's major axis, and θ represents the rotation angle of the ellipse. After the calculation for one window, the window is shifted by the step size of m, and then the next ellipse is calculated, thereby obtaining a continuous sequence of D2DGE eye-tracking data from the original gaze points data.

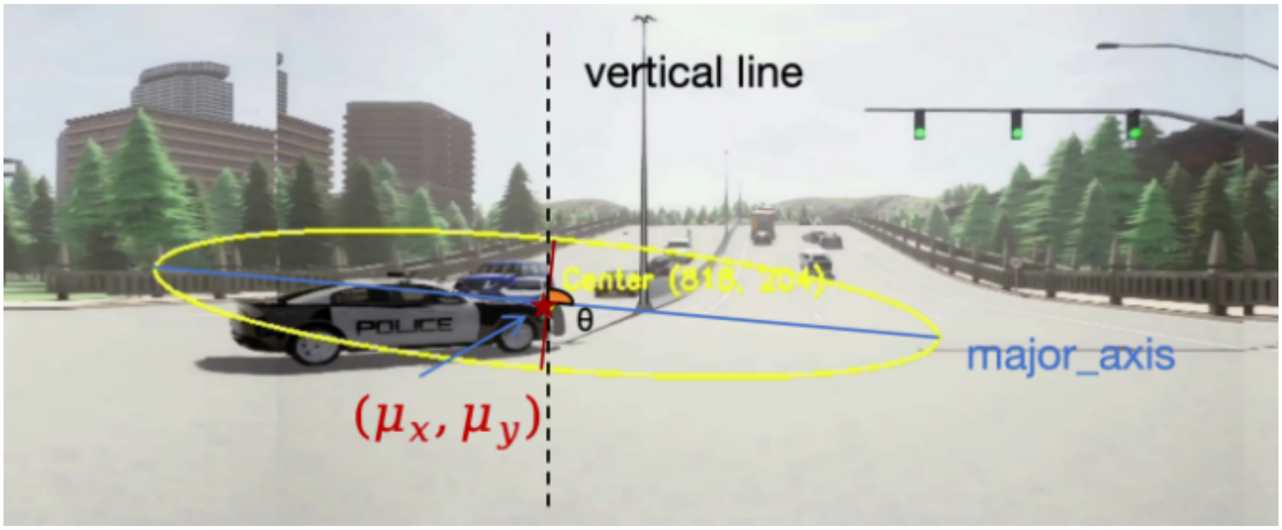

Figure 1 The proposed D2DGE representation

Driving Simulation Data for Validation

To validate the effectiveness of the D2DGE representation, a driving simulator experiment was conducted to record the visual behaviors of experienced and novice drivers during specific driving scenarios in a SAE L3 vehicle (Society of Automotive Engineers). A fixed-base simulator with three 1920x1080 screens was used and the eye-tracking data was captured at 60Hz using a Smart Eye desktop tracker. In total, original image sequences from 87 scenes were extracted based on the Deepaccident dataset (Wang, 2024). Then, the image sequences were enhanced (Wang, 2021), converted to video clips, and interpolated (Wu, 2024), resulting in 5760x1080 resolution videos at 20 FPS. Out of the 87 scenes, 42 involved accidents, while 45 did not. Twelve participants (6 males, 6 females) were recruited, with 3 experienced drivers (years of licensure $\geqslant$ 5 and mileage in the last year over 20,000 km) and 3 novice drivers (years of licensure < 2 and mileage in the last year less than 5,000 km) in each group. Participants were informed of L3 driving automation and instructed to take over the vehicle when hazards were perceived. The experiment lasted 18 minutes for each participant. Based on this experimental design, the parameters for D2DGE were set as follows, i.e., windows size n = 10 frames, sliding window m = 5 frames, and times of sigma range k = 2.

Analyses

To determine whether D2DGE data confer an advantage in data-driven visual attention prediction, a Generative Adversarial Imitation Learning (GAIL) model was trained independently on both D2DGE and raw gaze datasets, and the divergence between the generated outputs and the original gaze distributions was quantified using the Kullback–Leibler (KL) divergence.

GAIL is an imitation learning algorithm in which generative adversarial networks are combined with reinforcement learning, enabling decision-making policies that approximate expert-level performance to be learned without manual specification of reward functions. The KL divergence (Hershey, 2007) is an asymmetric measure of information that quantifies the information loss or discrepancy of one probability distribution relative to a reference distribution.

$$KL(P \parallel Q) = \sum_i P(i) \log\left(\frac{P(i)}{Q(i)}\right) \tag{1}$$

Typically, distributions P and Q are compared via their KL divergence, with smaller values indicating greater similarity, whereas larger values indicating greater divergence.

Specifically, in our work, the two sets of data (i.e., raw gaze and D2DGE data) were used as inputs to train a GAIL algorithm, with Proximal Policy Optimization

(PPO) (Schulman, 2017) as the policy generator. The generated data were then compared with their corresponding original data based on the KL divergence metric. It should be noted that, to make a fair comparison, the central point $[\mu_x, \mu_y]$ of the generated samples based on D2DGE data was determined by calculating the mean coordinates of all gaze points within a specific window, which ensures that the spatial distribution characteristics of the D2DGE central points remain consistent with those of the raw gaze points. Consequently, the transformation from raw gaze points to D2DGE representation preserves the fundamental spatial distribution properties of the raw gaze data.

Further, to explore if the new eye-tracking data representation contains more information compared to the raw data, one-way ANOVA tests were conducted, with the mean of the five parameters of D2DGE, and the mean coordinates of the raw gazes from each trial as dependent variables and driver experience (novice vs. experienced) as the independent variable.

The model was built using the Proc Mixed in SAS on demand, which included a random intercept for participant and a compound symmetry (CS) residual structure, estimated using REML (Corbeil, 1976).

RESULTS

First, as shown in Table 1, the KL divergence between the generated data and the original data showed that, for both novice and experienced drivers, the D2DGE-generated data exhibited smaller differences between the original and generated data as compared to the raw-gaze-generated data for both novice and experienced drivers.

Table 1. KL divergence between the original data and the generated data.

	Data Type	**Driver Type**	**KL Divergence**
Original data	Gaze point	Experienced	0.588
	D2DGE		0.351
VS	Gaze point	Novice	0.709
Generated data	D2DGE		0.389

Table 2. The influence of driving experience on gaze locations.

Outcome	Estimate	SE	DoF	t-value	*p*-value
x	46.9217	48.9552	859	0.96	.3381
y	95.0166	72.2254	859	1.32	.1887

Note: In this table and the following tables, "" denotes p < .05; DoF is the degrees of freedom, which is the approximate value adjusted by the Kenward–Roger method; SE is the standard deviation.*

Table 3. The influence of driving experience on D2DGE Parameters.

Outcome	Estimate	SE	DoF	t-value	*p*-value
μ_x	15.4	44.0	280	0.35	.7264
μ_y	65.6	63.8	280	1.03	.3050
s	-21933	9392.9	280	-2.34	.0202*
r_{major}	-106.76	108.56	280	-0.98	.3262
θ	-4.5919	12.7964	280	-0.36	.7200

Then, a linear mixed-effects model was used to examine the influence of driving experience on raw gaze-point coordinates and the five D2DGE parameters for each participant. As shown in Table 2, no significant differences were observed between experienced and novice drivers in either lateral or longitudinal gaze positions ($p > .05$), a result that held for both the raw gaze points and the gaze based on D2DGE representation.

Likewise, we did not observe a significant difference between novice and experienced drivers regarding D2DGE major-axis length r_{major} and orientation angle θ ($p > .05$). However, the D2DGE ellipse area s differed significantly between experienced and novice drivers ($p = .02$), indicating that novice drivers scanned a broader region of the traffic scene than experienced drivers within any specific time window.

DISCUSSIONS

In this study, we proposed a new eye-tracking data format, the D2DGE, to describe the characteristics of drivers' visual scanning behaviors during driving. A driving simulator experiment was conducted to validate the effectiveness of the proposed model based on the distribution of the gaze positions and the data generated by GAIL.

First, as indicated by the KL divergence, the D2DGE can better recover the original distribution of the gaze positions than directly using the raw gaze position data. First, it is possible that due to the technical difficulties, the raw gaze locations are noisy. For example, the commonly used eye-tracking systems, such as SmartEye, Tobii or Dikablis, can only achieve an accuracy of 0.5, 0.6 and 0.5 degrees (Smart Eye, 2025; Tobii, 2025; EST, 2025). Second, some of the fast eye movements might be random and may not be associated with conscious attention allocation. Thus, the models based on the raw data may be overfitted and not be able to learn meaningful visual attention allocation strategies.

Second, a comparison between the novice and experienced drivers regarding the key metrics of D2DGE indicates that richer information can be extracted from the D2DGE data representation. Specifically, though lateral or longitudinal gaze positions, length and orientation angle were not found to be different between novice and experienced drivers, the ellipse area differed between novice and experienced drivers. It should be noted that, the lack of difference in terms of the lateral or longitudinal gaze positions might be due to the design of the experiment – it does not require drivers to drive the vehicle, but just to observe the environment. Thus, the superior performance of experienced drivers in distributing their visual attention to a broader area (Underwood, 2007) was not observed in our study. Instead, the driving scenarios started from a moment when predictable hazards are about to appear. Thus, experienced drivers might be quicker to notice the potential hazardous areas and pay more sustained attention to them compared to novice drivers. As a result, they exhibited an even narrower observational field compared to novice drivers. This result indicates that, compared to raw gaze-point data, D2DGE representation, owing to its greater number of parameters, may capture richer visual attention information and thus offers advantages in statistical analyses and gaze data generation based on artificial intelligence.

However, several limitations remain in this study. First, the parameters used to compute D2DGE, namely, the sliding window size, step length, and the k-sigma range, directly affect the resulting data and its characteristics, and the optimal k-sigma is dependent on the distance between participants and the presented stimulus, which should be carefully calibrated in future research. Second, we used relatively homogeneous driving scenarios. Future research should further validate our results in more diverse scenarios, preferably based on the data from on-road studies. Further, we only considered one algorithm to generate the eye-tracking data. Future research should consider more advanced and robust models to generate eye-tracking data based on both the raw gaze data and the D2DGE data, and explore how additional contextual information can be incorporated so that each generated trajectory can more accurately reflect real-world driver gaze behavior.

In summary, our results showed that the new eye-tracking data representation based on the original gaze data preserved the spatial distribution characteristics of the visual attention allocation in the 2D plane and contains more information describing drivers' eye-tracking behaviors than the raw gaze data. Additionally, the D2DGE representation enabled superior performance for data-driven eye-tracking analysis. Future research is needed to explore how D2DGE parameters are associated with driving performance.

CONFLICT OF INTEREST

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

AKNOWLEDGEMENTS

This work was supported by the Red Bird MPhil Program at the Hong Kong University of Science and Technology (Guangzhou), and the National Natural Science

Foundation of China (No. 52202425), and in part by the Guangzhou Municipal Science and Technology Project (No. 2023A03J0011), and Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things (No. 2023B1212010007).

REFERENCES

He, D. (2020). Understanding and Supporting Anticipatory Driving in Automated Vehicles. University of Toronto (Canada).

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

Werneke, J., & Vollrath, M. (2012). What does the driver look at? The influence of intersection characteristics on attention allocation and driving behavior. Accident Analysis & Prevention, 45, 610-619.

Baee, S., Pakdamanian, E., Kim, I., Feng, L., Ordonez, V., & Barnes, L. (2021). Medirl: Predicting the visual attention of drivers via maximum entropy deep inverse reinforcement learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 13178-13188).

Fu, R., Huang, T., Li, M., Sun, Q., & Chen, Y. (2023). A multimodal deep neural network for prediction of the driver's focus of attention based on anthropomorphic attention mechanism and prior knowledge. Expert Systems with Applications, 214, 119157.

Ho, J., & Ermon, S. (2016). Generative adversarial imitation learning. In Proceedings of the 30th International Conference on Neural Information Processing Systems (pp. 4572-4580).

Ahlström, C., Kircher, K., Nyström, M., & Wolfe, B. (2021). Eye tracking in driver attention research - How gaze data interpretations influence what we learn. Frontiers in Neuroergonomics, 2, 778043.

Gómez-Poveda, J., & Gaudioso, E. (2016). Evaluation of temporal stability of eye tracking algorithms using webcams. Expert Systems with Applications, 64, 69-83.

Wolfe, B., Dobres, J., Rosenholtz, R., & Reimer, B. (2017). More than the useful field: Considering peripheral vision in driving. Applied Ergonomics, 65, 316-325.

Vater, C., Wolfe, B., & Rosenholtz, R. (2022). Peripheral vision in real-world tasks: A systematic review. Psychonomic Bulletin & Review, 29(5), 1531–1557.

Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. Journal of Vision, 9(10), 6.

Wang, T., Kim, S., Wenxuan, J., Xie, E., Ge, C., Chen, J., … Luo, P. (2024, March). DeepAccident: A motion and accident prediction benchmark for V2X autonomous driving. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 38, No. 6, pp. 5599–5606).

Wang, X., Xie, L., Dong, C., & Shan, Y. (2021). Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1905–1914).

Wu, G., Tao, X., Li, C., Wang, W., Liu, X., & Zheng, Q. (2024). Perception-oriented video frame interpolation via asymmetric blending. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2753–2762).

Hershey, J. R., & Olsen, P. A. (2007, April). Approximating the Kullback-Leibler divergence between Gaussian mixture models. In 2007 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (Vol. 4, pp. IV-317). IEEE.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.

Corbeil, R. R., & Searle, S. R. (1976). Restricted maximum likelihood (REML) estimation of variance components in the mixed model. *Technometrics*, *18*(1), 31-38.

Smart Eye. (2025). *Smart Eye Pro: Remote eye tracking for human behavior research* [Product webpage]. Smart Eye AB. Retrieved May 2, 2025, from https://www.smart-eye.se/smart-eye-pro/

Tobii AB. (2025). *Tobii Pro Glasses 3: Specifications* [Product webpage]. Retrieved May 3, 2025, from https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3#specifications

EST, Engineering Systems Technologies GmbH & Co. KG. (2025). *DIKABLIS GLASSES 3: High precision eye tracking technology* [Product webpage]. Retrieved May 3, 2025, from https://www.est-kl.com/zh-hans/dikablis-glasses-3/

Underwood, G. (2007). Visual attention and the transition from novice to advanced driver. Ergonomics, 50(8), 1235-1249.