MDPI

Article

Dual-Cameras-Based Driver's Eye Gaze Tracking System with Non-Linear Gaze Point Refinement

Yafei Wang, Xueyan Ding, Guoliang Yuan * and Xianping Fu

School of Information Science and Technology, Dalian Maritime University, Dalian 116026, China; wangyafei@dlmu.edu.cn (Y.W.); dingxueyan@dlmu.edu.cn (X.D.); fxp@dlmu.edu.cn (X.F.)
* Correspondence: yuan@dlmu.edu.cn

Abstract: The human eye gaze plays a vital role in monitoring people's attention, and various efforts have been made to improve in-vehicle driver gaze tracking systems. Most of them build the specific gaze estimation model by pre-annotated data training in an offline way. These systems usually tend to have poor generalization performance during the online gaze prediction, which is caused by the estimation bias between the training domain and the deployment domain, making the predicted gaze points shift from their correct location. To solve this problem, a novel driver's eye gaze tracking method with non-linear gaze point refinement is proposed in a monitoring system using two cameras, which eliminates the estimation bias and implicitly fine-tunes the gaze points. Supported by the two-stage gaze point clustering algorithm, the non-linear gaze point refinement method can gradually extract the representative gaze points of the forward and mirror gaze zone and establish the non-linear gaze point re-mapping relationship. In addition, the Unscented Kalman filter is utilized to track the driver's continuous status features. Experimental results show that the non-linear gaze point refinement method outperforms several previous gaze calibration and gaze mapping methods, and improves the gaze estimation accuracy even on the cross-subject evaluation. The system can be used for predicting the driver's attention.

Keywords: driving environment; gaze tracking; non-linear refinement

1. Introduction

Monitoring a driver's attention is one of the most vital components of advanced driver assistance systems (ADASs) which monitor the driver's status and prevents traffic accidents from happening when the driver is distracted [1–4]. According to the demand for safe driving, drivers should maintain sufficient awareness in order to take a series of actions during the whole driving process. Distraction is always facilitated by driving over long durations under monotonous conditions or otherwise being "lost in thought", which could decrease the driver's awareness and impair their reaction speed with regard to regaining control of the vehicle. A better understanding of driver behaviors through driver attention analysis could promote ADAS to generate optimal control strategies corresponding to the current driving situation.

Eye gaze is often closely related to driving behaviors and is treated as a vital cue for detecting the driver's visual attention. Many remote gaze-tracking-based techniques, which are based on head and eye pose features extracted from driver videos, have popularly been used in driving scenarios due to their high user friendliness. Gaze mapping is the essential step to build the mapping relationship between the driver's eye gaze features and the corresponding gaze location for gaze prediction inside or outside the vehicle. Manual gaze mapping methods necessarily require driver compliance [5,6] and human cooperation before each session, which adds significant inconvenience for driver eye gaze tracking systems. Several automatic mapping methods have approached driver gaze zone estimation utilizing gaze behavior or gaze frequency on specific driving data [7–9], which

are heavily dependent on the parameter estimation accuracy and fragile to the outliers. This makes the driver gaze estimation system session-exclusive or person-exclusive, and requires re-calibration or re-mapping in the setup procedure before each session, which seriously affects the applicability of the pre-trained gaze model. However, there is little progress that works on such intrinsic problems in the driver's gaze estimation system.

This paper presents a domain knowledge-based solution to mapping and tracking the driver's predicted eye gaze points to their real location in the naturalistic dual-cameras-based monitoring system. The driver's natural mirror-checking actions are used for finding representative gaze points in the context, which are later generated by progressive clustering in an automatic manner. The representative gaze points can be treated as known gaze points for gaze point refinement. The non-linear gaze point refinement treats these representative gaze points as known gaze points, builds simple and straightforward re-mapping for gaze estimation, eliminates the estimation bias related to the session, and makes the pre-trained gaze model more generalized. The main contributions are as follows:

- A dual-cameras-based driver eye gaze tracking system using non-linear gaze point refinement is presented for deploying a pre-trained supervised gaze model in the unconstrained environment. This method makes an initial attempt to reduce the estimation bias in separate model training. It increases the flexibility of system setup and does not require any human intervention.
- An effective gaze point non-linear global refinement with two-stage clustering is presented to extract the typical gaze points by maximizing fixation possibilities. This method aligns the initial unknown gaze points to specific calibration points by topology preservation. It is person-independent and can be directly utilized as post-processing for many pre-trained gaze models.
- Experimental results of real driving scenarios demonstrate that the proposed method reduces the gaze estimation error of the pre-trained model and even has a better performance on cross-subject evaluations. It can be used as a simple-but-effective baseline method for driver gaze calibration or gaze mapping.

The rest of this paper is organized as follows. In Section 2, the driver eye gaze prediction and gaze calibration works are introduced. Section 3 describes the details of the proposed driver eye gaze tracking system. In Section 4, several experimental results and error analysis are given. Section 5 presents a brief conclusion of the proposed system.

2. Related Works

The human head and eye dynamics are fundamental to revealing the drivers' gaze points that represent their current visual attention. Therefore, it has been widely used to detect the visual distraction and understand driver behaviors by exploiting the driver's head and eye orientation [10]. In early works, several methods and devices were approached for accurate gaze information in driving environments, such as head-mounted eye trackers. These trackers are intrusive and costly, and might change driving habits and behavior. With the advances of the remote driver eye gaze tracking techniques, non-intrusive systems based on computer vision have been well applied due to their user-friendliness [11].

2.1. Driver's Eye Gaze Estimation

2.1.1. Feature-Based Systems and Appearance-Based Systems

In vision-based driver behavior monitoring systems, it is acceptable for the driver's head orientation regarding as coarse gaze direction [12]. Through a gaze zone estimator, the probability of the driver gaze direction could be obtained. These estimators classify the gaze points into partitioned gaze zones that are inside or outside the vehicle, such as side mirrors, rear-view mirror and windshields. Most of the driver gaze estimation works build the estimator with facial appearance [13,14], head movement [15–18], or eye movement [19,20] in a monocular camera system.

Wang et al. [21] estimated the driver's head pose in the depth image by a point-cloud alignment-based approach, and computed the eye direction in the RGB image by

appearance-based gaze estimation approach. In this work, the author determined the gaze zone results by the weighted sum of the head and eye direction. Jha et al. [22] used the head position and orientation as the input features and built the estimator via Gaussian process regression (GPR). Lundgren et al. [23] applied both head features and eye features to train the estimator via the GPR method.

2.1.2. Deep Learning-Based Systems

With the vigorous development of deep learning technology, various attempts have started being made to achieve eye gaze estimation in the driving environment. All these deep learning-based methods consist of a facial landmark detection or feature extraction module connected to the model input. The network is used to fuse the high-dimensional rich features. Yu et al. [24] fused appearance-based features and geometric information in the convolutional neural networks (CNNs) for efficient driver gaze prediction. Lyu et al. [25] trained the CNN model for gaze classification. This work carried out a fusion network that combined two classification networks. Lollett et al. [26] combined face detection, facial landmarks detection, eye image post-processing in one system, and classified the extracted feature vector into the driver's gaze zone. They tested the model's robustness in several challenging driving scenarios, such as non-uniform illumination and face occlusions.

For the dual-cameras-based system, the gaze points are predicted on the dynamic scene, not just the interior components of the vehicle. The face camera is often fixed at the windshield or center console to capture the driver's video in the field of view. The scene camera is installed behind the driver to mimic the driver's view and shows the gaze points in a naturalistic driving environment. Most current driver gaze estimation research studies have evaluated the predicted accuracy of gaze location regression or gaze zone classification inside or outside the vehicle. These studies are based on the domain knowledge in the training dataset and build the gaze model beforehand. It is vulnerable and inapplicable to the uncertain and complex on-road driving environment. Therefore, the driver's gaze tracking system needs a more flexible setup for the subjects about looking by head movements and eye movements.

2.2. Driver's Eye Gaze Calibration

Gaze calibration is a significant step in the gaze tracking system, which is employed to build the relationship between the human eye and the gaze location. The gaze calibration of driving scenes is usually realized by using typical characteristics of the scene, such as common components inside the vehicle, or the driving behavior characteristics of the driver. Some methods bring the calibration features of the indoor scene as additional prior information.

Fu et al. [7] introduced a calibration method for head pose estimation that regards the common gaze zones as the calibration points, such as the mirror gaze zone and the instrument board gaze zone. This method was realized by a self-learning algorithm, and its parameters can be periodically updated. Through the solving of the relevant coefficients, it refined the predicted head pose to the real head pose. Yamashiro et al. [8] used less gaze zone related to the driving behavior as calibration points under specific lane-changing conditions such as the rear-view mirror gaze zone, one-side mirror gaze zone, and the forward gaze zone. This method assumes that the data clustering centers should be the calibration points due to the possibility. This established the mapping relationship through the transformation of the three calibration points. Yoon et al. [27] introduced a stereo-camera-based calibration method using one calibration point. This method takes full advantage of the driver's prior information collected in the desktop monitor. Yang et al. [5] utilized a non-linear polynomial method to build the mapping relationship linking the features to the gaze points. Yuan et al. [9] used six gaze zones related to driver gaze glance as calibration points. This extracted and clustered more representative gaze features to build the regression by GPR. Due et al. [6] established the mapping between the driver view and road view by I-DGAZE model, but the status of the drivers was still varying and

needed to be fine-tuned by driver-specific gaze calibration. Their results showed that the gaze mapping or gaze calibration can greatly reduce the error of the trained CNN model.

It is worth noting that the gaze classification accuracy varies significantly between drivers and that the person-specific driver gaze estimation system obviously makes sense. To the best of our knowledge, few studies have aimed to implement the gaze mapping system with person-specific gaze point refinement for a naturalistic driving environment.

3. Proposed Method

The proposed method contains three modules: driver status tracking, pre-trained gaze model, and non-linear gaze point refinement. For the dual cameras system, the driver status tracking module infers the status features from the face camera, and the pre-trained gaze model module maps the features to the gaze points on the scene camera.

In the driver status tracking module, all the relevant driver status is defined and initialized. The status contains both the signals of the head movement and the eyes movement. Since the observed measurements have random disturbances to a certain extent, the Kalman filter is adopted to track and update the measurements. In the pre-trained gaze model module, the corresponding gaze mapping from the driver status features to the gaze points is built during the offline training. The model is then used to give the initial gaze points. At this point, the initial driver gaze tracking system has been built.

In the non-linear gaze point refinement module, the initial gaze points is corrected to the real locations. Firstly, the candidate gaze points are automatically extracted by two-stage gaze point clustering. The forward gaze zone center is measured by density peak clustering. The candidates of the three mirror gaze zones are selected from the maximum distance points on the time sample fragments. All possible gaze zone centers are found by Gaussian mixture cluster, and the parameters are updated by the expectation-maximization algorithm in each gaze zone. Then, based on the known mapping relationships of the candidate gaze points, a simple non-linear method is adopted to further adjust the gaze points on the image plane.

The whole framework is running automatically without any manual interpretation. Within it, the non-linear gaze point refinement module periodically refines and updates the gaze mapping parameters, making the framework itself an automatic system. Figure 1 illustrates the proposed system. This section describes each module of the system.
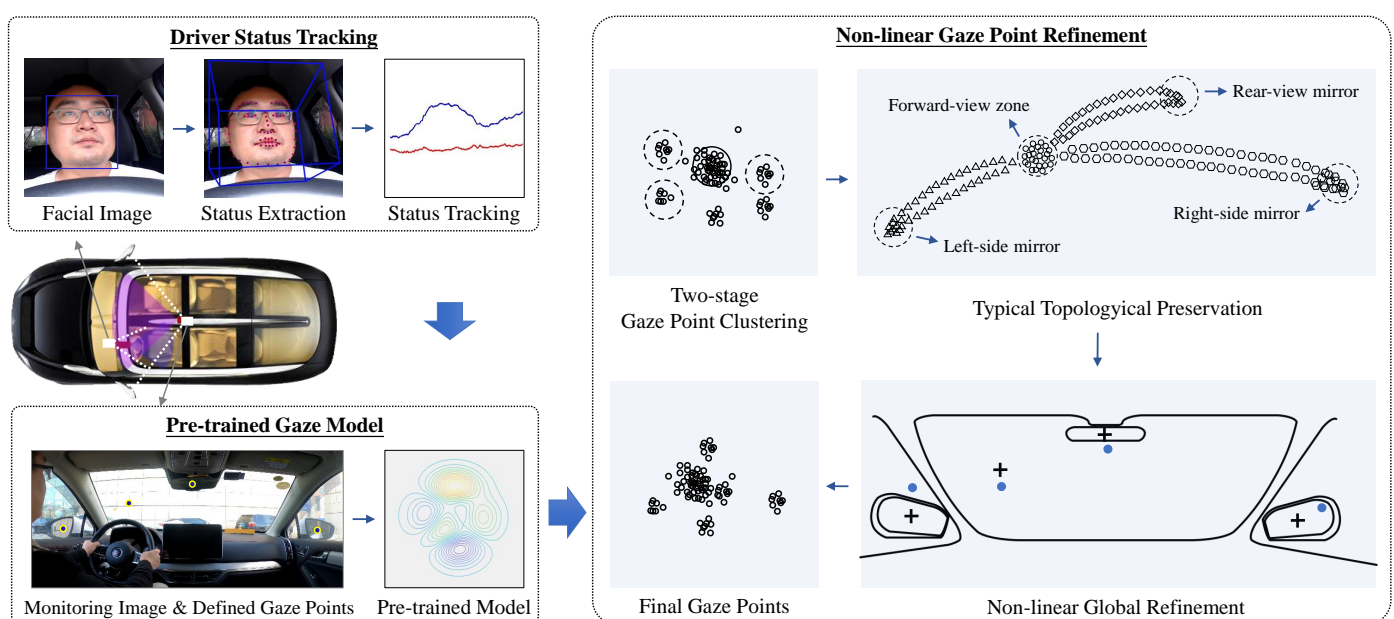

Figure 1. Overview of the proposed system.

3.1. Driver Status Tracking

This paper uses a single camera to achieve the signal acquisition of the driver. Since head position, head rotation, and eye rotation all contribute to driver gaze estimation, this paper considers all these signals in the driver status model. This is consistent with the driver's eye gaze glance in real driving scenarios, as drivers usually move their head and eyes to focus on the target. To perform the prediction step, the Unscented Kalman filter is employed.

In this paper, the driver state $\mathbf{x}_k$ that contains information about the driver's head position, head rotation, and gaze direction is introduced as the indicator of the driver's visual attention. The properties in the state features make a continent way to determine the driver's eye gaze if they were obtained. In practice, the state $\mathbf{x}_k$ is divided into three components: head state; left eye state; and right eye state. $\mathbf{x}_k = \left[\mathbf{h}_k^T, \mathbf{e}_{k,l}^T, \mathbf{e}_{k,r}^T \right]^T$, where:

$$
\begin{aligned}
\mathbf{h}_k &= \left[x_k, y_k, z_k, h_k^\alpha, h_k^\beta, h_k^\gamma, \dot{x}_k, \dot{y}_k, \dot{z}_k, \dot{h}_k^\alpha, \dot{h}_k^\beta, \dot{h}_k^\gamma \right]^T, \\
\mathbf{e}_{k,l} &= \left[e_{k,l}^\alpha, e_{k,l}^\beta, e_{k,l}^\gamma, \dot{e}_{k,l}^\alpha, \dot{e}_{k,l}^\beta, \dot{e}_{k,l}^\gamma \right]^T, \\
\mathbf{e}_{k,r} &= \left[e_{k,r}^\alpha, e_{k,r}^\beta, e_{k,r}^\gamma, \dot{e}_{k,r}^\alpha, \dot{e}_{k,r}^\beta, \dot{e}_{k,r}^\gamma \right]^T.
\end{aligned}
\tag{1}
$$

The head position and head rotation are denoted by $[x_k, y_k, z_k]^T$ and $[h_k^\alpha, h_k^\beta, h_k^\gamma]^T$. Furthermore, their corresponding change rates are denoted by $[\dot{x}_k, \dot{y}_k, \dot{z}_k]^T$ and $[\dot{h}_k^\alpha, \dot{h}_k^\beta, \dot{h}_k^\gamma]^T$. The eye gaze direction of both the left eye and the right eye is independent from the head rotation and is denoted by $[e_{k,l}^\alpha, e_{k,l}^\beta, e_{k,l}^\gamma]^T$ and $[e_{k,r}^\alpha, e_{k,r}^\beta, e_{k,r}^\gamma]^T$. The change rates of the two eye rotations are given as $[\dot{e}_{k,l}^\alpha, \dot{e}_{k,l}^\beta, \dot{e}_{k,l}^\gamma]^T$ and $[\dot{e}_{k,r}^\alpha, \dot{e}_{k,r}^\beta, \dot{e}_{k,r}^\gamma]^T$. The eye model is the same for the left eye and the right eye, and the gaze directions are the relative eye-in-head rotations. Therefore, the eye movement is independent from the head movement, and this property is useful for the gaze estimation process.

In this paper, the corresponding measurement vectors are given as $\mathbf{z}_k^g = [\mathbf{z}_k^h, \mathbf{z}_{k,l}^e, \mathbf{z}_{k,r}^e]$, where,

$$
\begin{aligned}
\mathbf{z}_k^h &= [x_k, y_k, z_k, h_k^\alpha, h_k^\beta, h_k^\gamma], \\
\mathbf{z}_{k,l}^e &= [e_{k,l}^\alpha, e_{k,l}^\beta, e_{k,l}^\gamma], \\
\mathbf{z}_{k,r}^e &= [e_{k,r}^\alpha, e_{k,r}^\beta, e_{k,r}^\gamma].
\end{aligned}
\tag{2}
$$

3.1.1. Process Model

This paper uses the following head model and eye model as the process models for the information processing. In practice, the driver's head and eyes correspond to the camera coordinate system, respectively. Hence, the driver's state can be expressed using the head status and eye status.

3.1.2. Head Model

Inspired by Ref. [23], a constant velocity model is utilized in the head model. The current head pose is determined by the previous head pose, head change rate, and the model noise. Therefore, the head model is defined as

$$
\mathbf{h}_k = \mathbf{A}_h \mathbf{h}_{k-1} + \mathbf{B}_h \mathbf{v}_{k-1},
\tag{3}
$$

where $\mathbf{v}_{k-1} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}^h \right)$ is Gaussian model noise. $\mathbf{A}$ and $\mathbf{B}$ are stated as

$$
\mathbf{A}_h = \begin{bmatrix} \mathbf{I} & T_s \mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \mathbf{B}_h = \begin{bmatrix} \frac{T_s^2}{2} \mathbf{I} \\ T_s \mathbf{I} \end{bmatrix},
\tag{4}
$$

where T_s denotes the sampling time.

3.1.3. Eye Model

The eye movement can generally be categorized into three types: fixation; saccades; and smooth pursuit. The eye model describes eye motion that either fixates on the target, quickly changes between fixation points, or smoothly follows the moving target. The correlation between the head movement and the eye movement is considered in the eye model. The previous method used the gaze signal as the combination of the head pose and gaze pose; here, in this paper, the eye rotation provided by the eye model is just the gaze pose. The model describes the eye rotation as the weighted sum of the current eye rotation and the eye rotation generated by the head model:

$$\mathbf{e}_k = \lambda_e \mathbf{e}_k^1 + (1 - \lambda_e)\mathbf{e}_k^2, \tag{5}$$

where $0 < \lambda_e < 1$ is the distribution determined by the head model:

$$\mathbf{e}_k^1 = \mathbf{A}_e^1 \mathbf{e}_{k-1}^1 + \mathbf{B}_e^1 \mathbf{v}_{k-1}, \tag{6}$$

where $\mathbf{v}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}^e)$ is the Gaussian model noise and:

$$\mathbf{A}_e^1 = \left[\begin{array}{c} T_s\mathbf{I} \\ \mathbf{I} \end{array} \right], \mathbf{B}_e^1 = \left[\begin{array}{c} \frac{T_s^2}{2\mathbf{I}} \end{array} \right], \tag{7}$$

$$\mathbf{e}_k^2 = \mathbf{A}_e^2 \mathbf{e}_{k-1}^2 + \mathbf{B}_e^2 \mathbf{v}_{k-1}, \tag{8}$$

where,

$$\mathbf{A}_e^2 = \left[\begin{array}{c} \mathbf{0} \\ \mathbf{I} \end{array} \right], \mathbf{B}_e^2 = \left[\begin{array}{c} \frac{T_s^2}{2\mathbf{I}} \end{array} \right]. \tag{9}$$

3.1.4. Measurement Model

This part introduces the measurement model as the function of the driver state. As the process model, the measurement model consists of the head model and eye model.

Head Model: The head position and head rotation is measured in this paper, and the state vector in the measurement model is defined as

$$\mathbf{z}_k^h = \mathbf{H}\mathbf{h}_k + \mathbf{w}_k = [\mathbf{I} \quad \mathbf{0}]\mathbf{h}_k + \mathbf{w}_k \tag{10}$$

where $\mathbf{w}_k \sim \mathcal{N}\left(\mathbf{0}, \mathbf{R}^h\right)$ is the measurement noise.

Eye Model: The eye model uses the three-dimensional gaze direction as the state variable. The measurement model is stated as

$$\mathbf{z}_k^e = \mathbf{e}_k + \mathbf{w}_k \tag{11}$$

where $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}^e)$. When the observations are available in the measurement state, its measurement vector $\mathbf{z}_k^g$ can be obtained.

3.2. Pre-Trained Gaze Model

In this paper, the traditional driver gaze estimation algorithm is used to train the gaze model when getting enough status samples in the dataset. Generally, the linear or non-linear model regresses the driver status features to specific gaze points on scene image. The pre-trained gaze model should not be over-fitted and has a balanced performance on the labelled training dataset and testing dataset. Mathematically, the gaze model can be written as

$$\begin{aligned} g_x &= f_x\left(\mathbf{z}_1^g, \mathbf{z}_2^g, \ldots, \mathbf{z}_k^g\right) \\ g_y &= f_y\left(\mathbf{z}_1^g, \mathbf{z}_2^g, \ldots, \mathbf{z}_k^g\right) \end{aligned} \tag{12}$$

where $\mathbf{z}_k^g$ denotes the k-th obtained status features in the dataset. g_x and g_y represent the gaze points in horizontal and vertical directions, respectively. $f_x(\cdot)$ and $f_y(\cdot)$ are the given linear or non-linear mappings of the gaze points. In this work, the gaze points are modeled in two independent directions.

It is worth noting that the robustness of the pre-trained gaze model obviously has a significant impact on the next step. If the prediction results of the pre-trained model are particularly poor, it is necessary to add much driver status features to accommodate the data diversity and train a new gaze model. Our goal is to refine the output of the model and make it easy to apply. At this point, it is also practical to build an online model as Ref. [9]. The refinement method works on the online model.

3.3. Non-Linear Gaze Point Refinement

In this section, the main processes of non-linear gaze point refinement are described.

3.3.1. Two-Stage Gaze Point Clustering

Mirror-checking behaviors can be regarded as typical eye glance allocation that benefits from both head and eye movement. The forward-view gaze zone is the major region of the driver's visual attention, which means that the glancing of the mirror gaze zone goes back and forth from the forward-view gaze zone.Figure 2 gives an example of the mirror-checking behavior. It will be facilitated to the initial gaze points by supervised refinement. Previous works having studied the mirror gaze zones' detection or classification include [28–31]. In this work, a two-stage gaze point clustering method was adopted to seek the gaze points of representative mirror gaze zones in continuous naturalistic driving data.

(**a**)

(**b**)
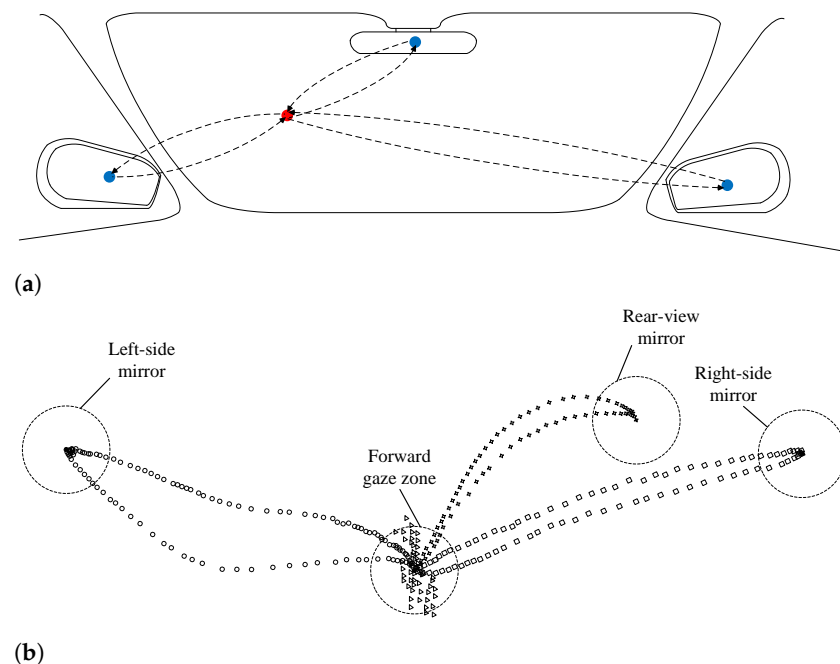
Figure 2. Typical eye glance allocation of mirror-checking behavior. (**a**) Schematic view of eye glance allocation. (**b**) Real-world examples in the gaze points space. Intuitively, the glance allocation for each mirror-checking behavior is discrete. The triangles denote the gaze points when the driver is facing the frontal area in the vehicle, and the circles, rectangles, and four-pointed stars denote the features' trajectory when the driver turns to the left-side mirror, right-side mirror, and rear-view mirror, respectively. Here, the furthest feature points are used as the representative candidates for the driving behavior.

3.3.2. Gaze Points Clustering

The driver maintains their visual attention on the road by moving their head and eye during the on-road driving, and uses the mirror gaze zones to assist their observation of the surrounding environment. Thus, most of the gaze points should be fixated on the forward-view gaze zone. This paper sets the center of the forward-view gaze zone as the global density center of the initial gaze points which is consistent with the actual situation. Considering the peak density measurement, the center has the characteristics of the largest value in the local neighbor density and the smallest value in the point distance. Hence, the local neighbor density and point distance can be computed by

$$
\begin{aligned}
\rho_k &= \sum_{k'} sgn(d_{kk'} - d_c) \\
\delta_k &= \min_{k' \in \Omega - k} d_{kk'}
\end{aligned}
\tag{13}
$$

where $sgn(\cdot)$ denotes the signum function, its value equals 0, when x is negative; otherwise, its value equals 1. $d_{kk'}$ means the distance between the k-th gaze point and k'-th gaze point in the given space. d_c represents the cutoff distance. Mathematically, the center of the forward-view gaze zone e_k^g is the gaze point with $\max(\rho_k)$ and $\min(\delta_k)$. Its cluster density radius is given as $\tau = \frac{\lambda}{|\Omega|} \sum_{k' \in \Omega - \bar{k}} d_{\bar{k}k'}$, where $|\Omega|$ denotes the data size of the initial gaze points. λ means the statistical probability of the forward-view gaze zone.

3.3.3. Mirror Gaze Points Clustering

The mirror gaze zones also have a higher local neighbor density, thence previous works directly clustered the forward-view gaze zone and mirror gaze zones in the same procedure. This matches the relatively dense distribution of the gaze points during long-term driving. However, the complexity and uncertainty make the exceptions inevitable. It is important to increase the effectiveness of the clustering results and improve the speed of the algorithm. Therefore, this paper filters out the candidate gaze points of the mirror gaze zone on the segmented data out of the forward-view gaze zone, and further use the Gaussian mixture clustering method to eliminate the noise.

The original gaze points data are converted to binary status data with directionality $\mathcal{B}$. These status data indicate whether the gaze point is still in the forward-view gaze zone. When the data are out, its value is positive; otherwise, its value is negative. The data beyond a certain distance can be considered as those on mirror-checking action. Here, the binary status is defined as $B_k = sgn(d_{\bar{k}k'} - \tau)$.

Mirror-checking behavior is a continuous action with processive time samples. The candidate gaze point for the single action can be the representative gaze point among the time samples. In this paper, the time samples are segmented into an importance fragment $\mathcal{M}$. Hence, the candidate gaze points e_k^* is computed as

$$
g_k^* = \arg\max_{g_k \in \mathcal{M}} d_{g_{\bar{k}} g_{k'}}
\tag{14}
$$

In this paper, the mirror gaze zones are assumed as the region with high local density and modeled by a Gaussian mixture model as single two-dimensional distribution for each mirror gaze zones. The gaze points of each gaze zone are i.i.d. with an underlying density of $p(\mathbf{x})$. The finite mixture model $p(\mathbf{x})$ is computed as

$$
p(\mathbf{x}) = \sum_{m=1}^{M} \alpha_m \mathcal{N}(g_m \mid \mu_m, \Sigma_m)
\tag{15}
$$

where μ_m and Σ_m are the parameters defined over the density or distribution. α_k are the mixture weights, $\sum_{m=1}^{M} \alpha_m = 1$. Here, M is the components number as well as the number

of the mirror gaze zone. In this scenario, the maximum likelihood is defined and estimated as follows:

$$\mathcal{L}_{ml}(\theta_{ml}) = \prod_{m=1}^{M} p(g_m; \theta_{ml}) \tag{16}$$

where $\theta_{ml} = \{\mu_1, \ldots, \mu_M, \Sigma_1, \ldots, \Sigma_M\}$. This cluster is solved by the expectation-maximization (EM) algorithm to proceed with a posteriori estimation of clustering parameters.

3.3.4. Typical Topology Preservation

After obtaining the clustered gaze points, the main task is to assign the unsupervised gaze points with correct gaze zone labels and anchor positions. This paper assigns the clusters by typical topology preservation. As shown in Figure 2, the eye glance allocation of the mirror-checking behavior is similar in the gaze points space. At the same time, the motion of the initial gaze points holds the related position. Based on this observation, this paper assumes that the representative points of the cluster center has an approximate position that consists of the annotations on the image plane and the same motions with the head movement. In this work, the gaze point is uncalibrated and output by the pre-trained gaze model. Hence, the cost function is given as

$$\mathcal{L}_{tp}(\theta_{tp}) = \min \sum_{j} \cos^{-1} \frac{(g_j - g_1) \cdot (h_j - h_1)}{|g_j - g_1||h_j - h_1|} \tag{17}$$

where $\theta_{tp} = \{g_1, \ldots, g_J, h_1, \ldots, h_J\}$. g_1 and g_j are the cluster centers of the forward-view gaze zone and the mirror gaze zones, respectively. h_1 and h_j are the head movement variable of the forward-view gaze zone and the mirror gaze zones, respectively. The loss function in the gaze points space can be minimized by keeping the similar topological structure in the physical space. The function formulation is optimized in a particular order as follows. The points of the left-side mirror and right-side mirror have higher priority. Because these points have large distances in the opposite directions, they are convenient to decide. When the cluster centers of the left-side mirror and the right-side mirror are determined, the cluster center of the rear-view mirror can be naturally assigned. At this end, the initial gaze points (the gaze points of mirror gaze zones and forward-view gaze zone) are comprised of assigned labels and can be used to refine the gaze points.

3.3.5. Non-Linear Global Refinement

This paper uses the non-linear global refinement to fine-tune the output of the pre-trained model. The corrected positions of the initial gaze points have undergone an unknown transformation relative to the initial values. Therefore, the refinement method should take care of the accuracy of all gaze zones and achieve the balance of gaze points' adjustment. The refinement method is computed as

$$\begin{aligned} g_x^{refine} &= f_x^{refine}(g_x) \\ g_y^{refine} &= f_y^{refine}(g_y) \end{aligned} \tag{18}$$

where g_x, g_y are the initial gaze points in the horizontal and vertical directions, respectively. $g_x^{refine}, g_y^{refine}$ are the corresponding labeled gaze points. $f_x^{refine}(\cdot)$ and $f_y^{refine}(\cdot)$ are the non-linear functions of the gaze points. In its simplest form, the non-linear function model can be $f^{refine}(\cdot) = \omega f(\cdot) + \nu$, where ω and ν are function parameters.

4. Experiments and Discussions

In this section, we experiment with several pre-trained models to validate the generalization ability of the proposed system. Firstly, the driving data collection under naturalistic conditions is introduced. Then, the pre-trained model and baseline method for comparisons is given. Subsequently, evaluations of the proposed system are performed. Finally, an ab-

lation study is given on within-subject evaluation and cross-subject evaluation, and error analysis is carried out on gaze estimation accuracy.

4.1. Naturalistic Data Collection

To evaluate the proposed system, a dataset for the field study is collected in the naturalistic driving environment. The face camera was fixed to the bottom of the windshield and the scene camera was mounted behind the driver. The images captured from the face camera are 720P high-quality images. The frame rate of the camera video is 30FPS.

The gaze region in front of the driver is partitioned into 36 gaze zones, as shown in Figure 3. All gaze zones are utilized as an indicator of the coarse gaze directions. The partitioned gaze zone covers all possible gaze regions in front of the driver to increase the variance of the head and eye status. The standard data of the driver's gaze in different gaze zone were collected in advance to verify the manual calibration gaze error.
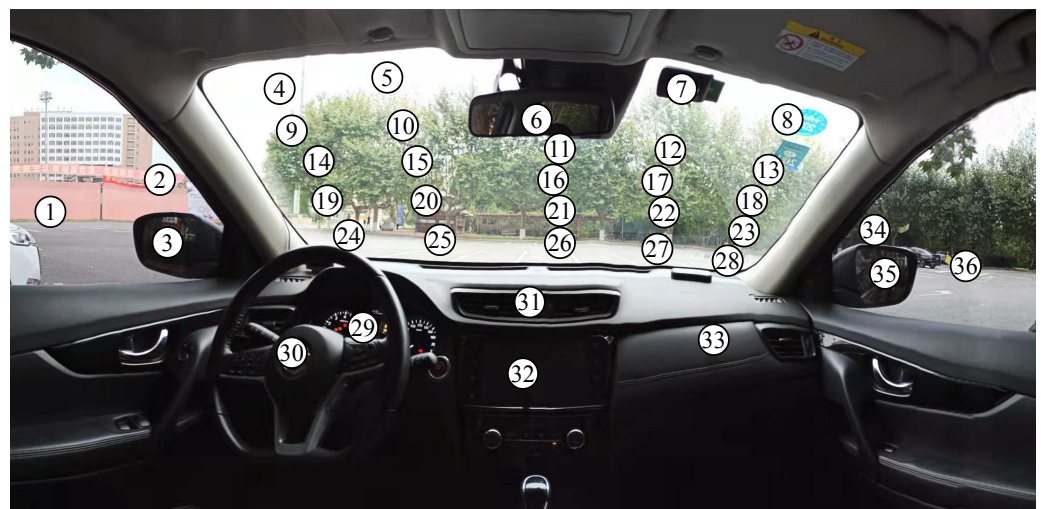

Figure 3. Target gaze zone partition of the on-the-road area. Gaze zone 1–36 are the target regions in the front of the driver.

To facilitate the data collection, yellow markers were pasted into the center of every gaze zone, which can easily be noticed in the vehicle cabin. The drivers who participated were asked to look at the marker one after another during the calibration procedure. In the training and testing data for calibration, the dataset retains five driver subjects' data. For each subject, the dataset has 720 annotated gaze zone data to train the model, and another 720 data to test the model. This means 20 annotated data remain for each gaze zone.

In this dataset, we adopted OpenFace [32] for convenient driver head and eye status extraction. OpenFace tools achieved stable and reliable performance on the head and gaze estimation in the collocated video compared with other existing tools. Advances in this area can benefit the driver gaze analysis and develop more robust solutions [33,34]. Figure 4 shows the face and eye detection results of different gaze zones.

4.2. Pre-Trained Models and Baseline Methods

In addition to the dataset, we evaluated the proposed method with the common pre-trained model, Gaussian process for regression (GPR) [23,35], partial linear square regression (PLSR) [14], non-linear square regression (NLSR). All these models were individually trained on the training dataset for different drivers. The GPR method builds the probabilistic model by taking the prior of the function space as the Gaussian process and carries out the solution according to the Bayesian inference. Here, the parameters of GPR are similar to Ref. [9]. The PLSR method trains the regression model on the principle component by projecting the variables into new space. Here, the components' value of

the PLSR model is 4, which is the same as that of the calibration gaze zones. Inspired by Ref. [5], a single NLSR model is utilized for non-linear estimation to link the inputs and output. For each driver, we trained the GPR model, PLSR model, and NLSR model for pre-trained model evaluation. Here, several calibration methods are used as the baseline method for gaze prediction comparisons, including Gaussian process mapping (GPM), homography transform projection (HTP). The GPM method builds the mapping relationship as performed in Ref. [9], and the HTP method treats the gaze point refinement as a homography transform from the primary image plane to the final image plane.

Figure 4. The driver status extraction results of all the partitioned gaze zone. The first row shows the results of GZ1∼GZ6. The second row shows the results of GZ7∼GZ12. The third row shows the results of GZ13∼GZ18. The fourth row shows the results of GZ19∼GZ24. The fifth row shows the results of GZ25∼GZ30. The last row shows the results of GZ31∼GZ36. GZ: Gaze Zone.

Here, the experiments are conducted on within-subject evaluation, cross-subject evaluation, and normal evaluation. Within-subject evaluation means that testing and training are performed on the same driver subject's data and the total number of experiments is $5(subject) \times 1(experiment/subject) = 5(experiments)$. While the cross-subject evaluation means that training on one driver subject's data and testing on another driver subject's data, the total number of experiments is $5(subject) \times 4(experiment/subject) = 20(experiments)$. For the normal evaluation without specific instructions, the testing is on all driver subject's data and the total number of experiments is $5(subject) \times 5(experiment/subject) = 25(experiments)$.

4.3. Gaze Point Prediction Results

This section compares the proposed method with the baseline methods on the pre-trained models. The dataset consists of the driver status for each gaze zone, with labels of pixel-level gaze points and annotations of the gaze zone index. The locations of the gaze zones are labeled on the image coordinate axis. For the particular driver, the proposed method will gradually revise the gaze points and build the refinement of the gaze model.

The gaze estimation error is calculated in the perspective of the direction angle view and image pixel view. The ground-truth values are provided by the labeled anchors of gaze zone centers. The size of the scene image is 2200×1080, and the relevant angle of the

full field-of-view is $137.5° \times 67.5°$. All the gaze estimation methods are measured by the absolute mean error of both the horizontal and vertical directions of eye gaze.

Table 1 presents the absolute mean error on a different calibration method. The proposed method achieves better results with a more than 1.4 degree error drop on all pre-trained models. Among them, the HTP method has a relatively large estimation error which may establish incorrect mapping in all possible solutions of the image plane. The HTP method can especially not obtain an output on several GPR model evaluations. At this point, the mean error of the HTP method is computed on the positive solution. The proposed method with the PLSR model has the best gaze estimation accuracy and largest accuracy improvement.

Table 1. Absolute mean error (degree) on different calibration methods and pre-trained models. GPM: Gaussian process mapping, HTP: homography transform mapping.

Calibration Method	GPR Model	PLSR Model	NLSR Model	Average
wo/Calibration	7.74	8.97	8.09	8.27
GPM	7.16	8.37	8.20	7.91
HTP	17.98	17.63	16.99	14.45
Ours	6.39	6.13	6.59	6.37

Figure 5 gives the absolute mean error for each gaze zone. The proposed method decreases the estimation error on more than two-thirds of the gaze zone. Among them, the error distribution of each gaze zone is relatively flat on the results of the PLSR model. In this table, the HTP method is ignored due to the unstable results on each gaze zone. Although the error decrease level varies among gaze zones, all gaze zones adjacent to the calibration gaze zones achieve a lower gaze prediction error.
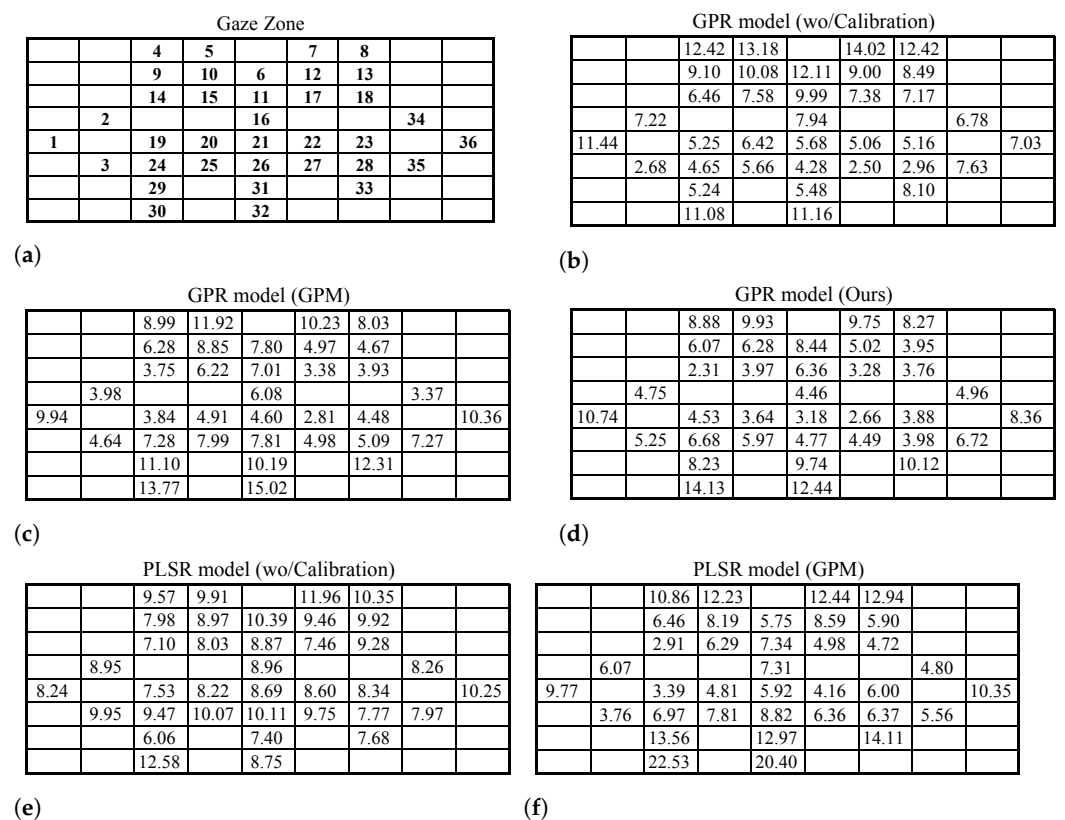
Gaze Zone

		4	5		7	8		
		9	10	6	12	13		
		14	15	11	17	18		
	2			16			34	
1		19	20	21	22	23		36
	3	24	25	26	27	28	35	
		29		31		33		
		30		32				

(**a**)

GPR model (wo/Calibration)

		12.42	13.18		14.02	12.42		
		9.10	10.08	12.11	9.00	8.49		
		6.46	7.58	9.99	7.38	7.17		
	7.22			7.94			6.78	
11.44		5.25	6.42	5.68	5.06	5.16		7.03
	2.68	4.65	5.66	4.28	2.50	2.96	7.63	
		5.24		5.48		8.10		
		11.08		11.16				

(**b**)

GPR model (GPM)

		8.99	11.92		10.23	8.03		
		6.28	8.85	7.80	4.97	4.67		
		3.75	6.22	7.01	3.38	3.93		
	3.98			6.08			3.37	
9.94		3.84	4.91	4.60	2.81	4.48		10.36
	4.64	7.28	7.99	7.81	4.98	5.09	7.27	
		11.10		10.19		12.31		
		13.77		15.02				

(**c**)

GPR model (Ours)

		8.88	9.93		9.75	8.27		
		6.07	6.28	8.44	5.02	3.95		
		2.31	3.97	6.36	3.28	3.76		
	4.75			4.46			4.96	
10.74		4.53	3.64	3.18	2.66	3.88		8.36
	5.25	6.68	5.97	4.77	4.49	3.98	6.72	
		8.23		9.74		10.12		
		14.13		12.44				

(**d**)

PLSR model (wo/Calibration)

		9.57	9.91		11.96	10.35		
		7.98	8.97	10.39	9.46	9.92		
		7.10	8.03	8.87	7.46	9.28		
	8.95			8.96			8.26	
8.24		7.53	8.22	8.69	8.60	8.34		10.25
	9.95	9.47	10.07	10.11	9.75	7.77	7.97	
		6.06		7.40		7.68		
		12.58		8.75				

(**e**)

PLSR model (GPM)

		10.86	12.23		12.44	12.94		
		6.46	8.19	5.75	8.59	5.90		
		2.91	6.29	7.34	4.98	4.72		
	6.07			7.31			4.80	
9.77		3.39	4.81	5.92	4.16	6.00		10.35
	3.76	6.97	7.81	8.82	6.36	6.37	5.56	
		13.56		12.97		14.11		
		22.53		20.40				

(**f**)

Figure 5. *Cont.*

PLSR model (Ours)

		5.95	6.65		7.34	7.43		
		5.01	5.88	5.86	7.57	4.21		
		3.82	5.01	5.72	3.76	4.85		
	6.33			6.74			4.07	
5.42		4.71	6.30	6.96	4.05	5.90		6.11
	7.24	7.37	7.98	8.45	6.23	5.33	5.06	
		4.08		6.71		6.39		
		11.57		8.60				

(**g**)

NLSR model (wo/Calibration)

		5.94	6.63		11.66	16.89		
		5.32	6.42	6.04	10.27	11.36		
		4.53	5.25	8.18	12.16	12.34		
	11.11			8.38			6.54	
14.50		4.60	5.73	7.74	9.87	8.68		7.99
	6.77	6.32	6.56	7.99	8.11	8.59	5.07	
		3.01		5.34		9.85		
		9.29		6.02				

(**h**)

NLSR model (GPM)

		8.99	9.40		13.28	12.72		
		5.66	7.39	3.10	9.83	9.45		
		1.84	4.79	8.24	8.10	8.44		
	6.39			8.02			3.85	
10.20		3.67	6.27	7.27	9.84	6.71		5.63
	2.24	7.61	9.23	9.99	10.10	8.35	2.80	
		10.54		11.11		13.24		
		14.79		16.11				

(**i**)

NLSR model (Ours)

		5.30	5.86		10.93	15.17		
		4.81	5.16	2.31	8.56	10.04		
		3.11	3.53	6.99	8.92	9.11		
	9.26			6.27			5.60	
12.50		3.24	5.10	5.61	8.64	5.56		6.73
	3.12	5.89	6.31	6.43	7.55	7.28	2.90	
		2.91		4.13		9.37		
		8.41		4.74				

(**j**)

Figure 5. Absolute mean error (degree) on different gaze zones: (**a**) gaze zone partition; (**b**) GPR model without calibration; (**c**) GPR model with GPM method; (**d**) GPR model with the proposed method; (**e**) PLSR model without calibration; (**f**) PLSR model with GPM method; (**g**) PLSR model with the proposed method; (**h**) NLSR model without calibration; (**i**) NLSR model with GPM method; and (**j**) NLSR model with the proposed method. Compared with the related pre-trained models without calibration, the pre-trained models with GPM method or the proposed method have reduced the estimation error on gaze zones.

4.4. Ablation Study and Error Analysis

This section studies the effect of various gaze zone calibration data. Table 2 gives the ablation study results calibrated with different gaze zone data. It shows that the calibration method without the right-mirror gaze zone data has the largest estimation error. This suggests that the right-mirror gaze zone is important in the calibration data since it is the only calibrated gaze point at the right side. The proposed method with scattered calibration points outperforms the method with concentrated calibration points and significantly reduces the prediction error.

Table 2. Absolute mean error (degree) of gaze estimation performance on different calibrated gaze zones. wo/means without the calibration data from the given gaze zone. Front: GZ14; Left: GZ3; Middle: GZ6; Right: GZ35; wo/Calibration: without any calibration.

Calibration Method	GPR Model	PLSR Model	NLSR Model	Average
wo/Front	8.71	6.26	6.78	7.25
wo/Left	6.70	5.96	6.62	6.42
wo/Middle	7.00	6.50	6.80	6.77
wo/Right	11.09	9.22	7.21	9.17
Ours	6.39	6.13	6.59	6.37

Tables 3 and 4 show the within-subject performance and cross-subject performance of the proposed method. As can be observed, all methods have poorer performance on the cross-subject evaluation than on within-subject evaluation. The proposed method can reduce the gaze estimation error in both cross-subject and within subject evaluations. There is no significant difference on the results of the proposed method with the GPR model. The mean error on cross-subject evaluation is less than 7 degrees. The gaze error of the proposed method on within-subject evaluation using the PLSR model is approximately five degrees. In the driving scenarios, the gaze tracking system has a somewhat large error tolerance, which is sufficient.

In addition, the error analysis on the horizontal and vertical directions of the image plane is given for further validation. As shown in Figure 6, all pre-trained models benefit from the proposed method, and have a large error reduction after gaze point refinement. In Figure 6, each block represents a five degree angle in the horizontal and vertical directions of the image plane. The difference values with the positive effect of the pre-trained model with the proposed method are denoted by colors. Based on this statistic analysis, the gaze estimation error of the three mirror gaze zone is significantly decreased among consecutive gaze zones, whose maximum accuracy improvement can reach 15 degrees. This is effective for driver gaze refinement.

Table 3. Absolute mean error (degree) of within-subject performance on different calibration methods. GPM: Gaussian process mapping; HTP: homography transform mapping.

Calibration Method	GPR Model	PLSR Model	NLSR Model	Average
wo/Calibration	6.99	5.26	5.67	5.97
GPM	6.77	8.13	7.30	7.40
HTP	7.49	10.79	5.67	7.98
Ours	6.19	5.04	5.67	5.63

Table 4. Absolute mean error (degree) of cross-subject performance on different calibration methods. GPM: Gaussian process mapping, HTP: homography transform mapping.

Calibration Method	GPR Model	PLSR Model	NLSR Model	Average
wo/Calibration	7.93	9.90	8.69	8.84
GPM	7.26	8.43	8.45	8.05
HTP	20.78	19.34	19.81	19.98
Ours	6.44	6.40	6.82	6.55

(**a**)

Figure 6. *Cont.*

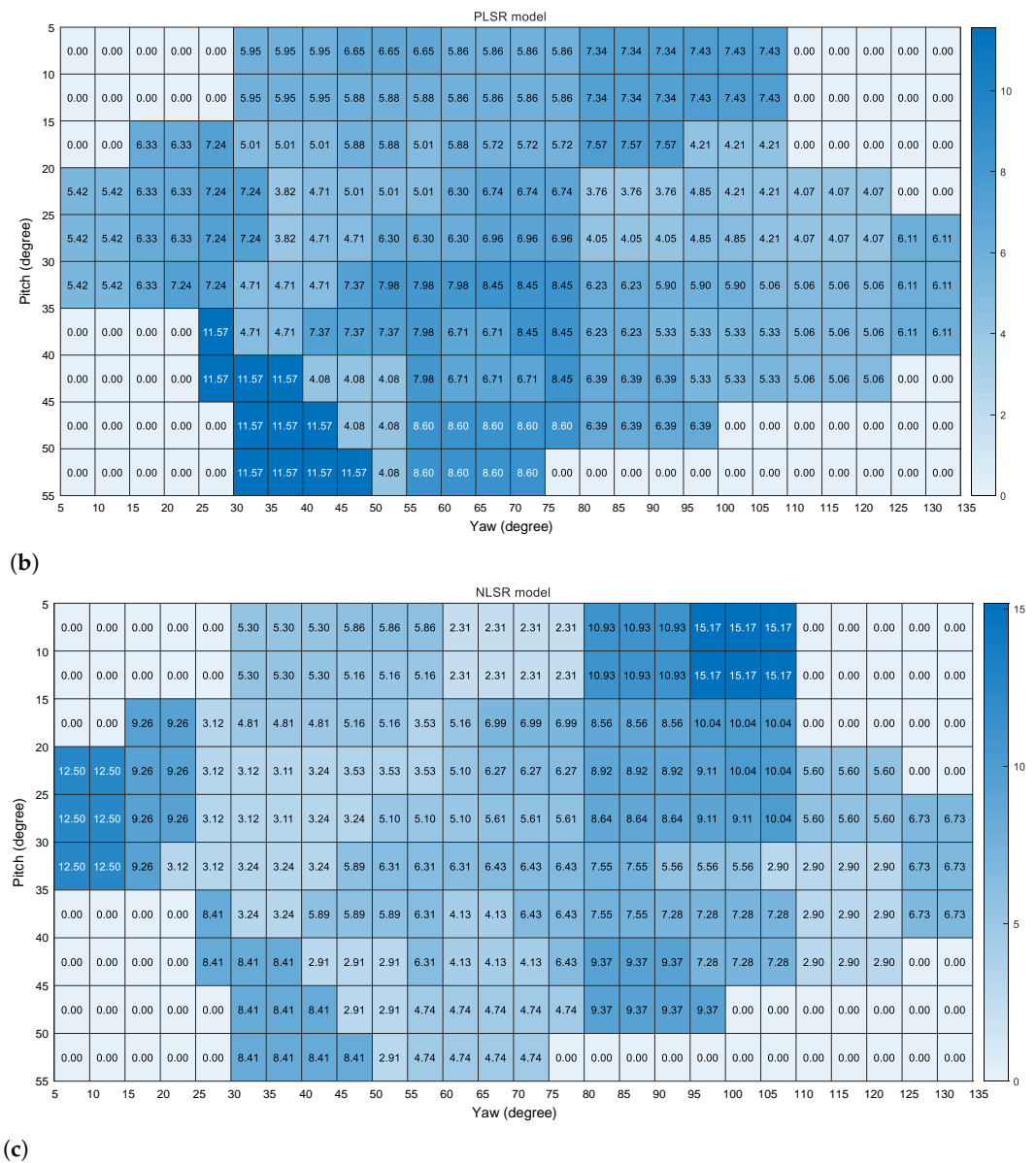PLSR model

(**b**)

NLSR model

(**c**)
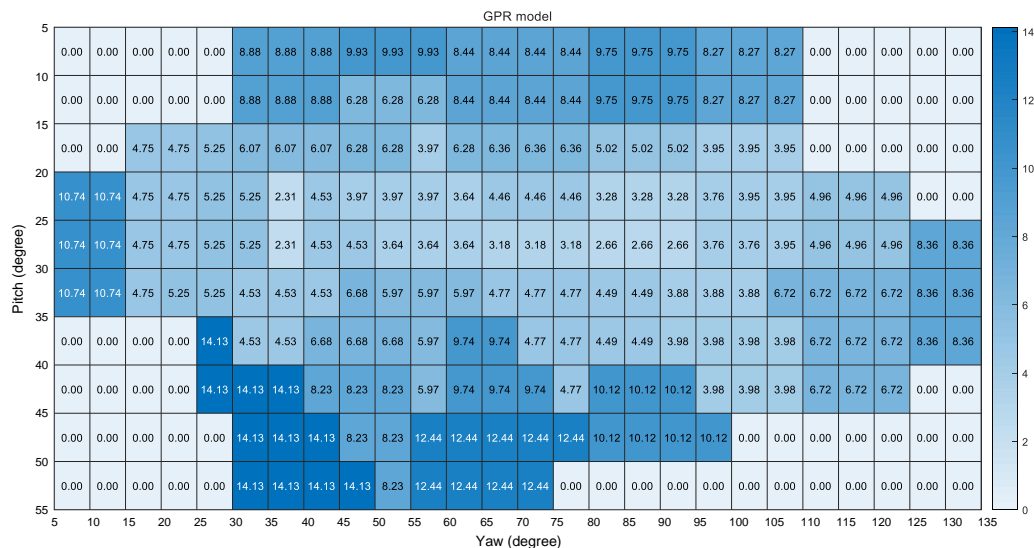
Figure 6. Error reduction of the pre-trained model in the horizontal and vertical directions of the image plane, the proposed method with (**a**) GPR model; (**b**) PLSR model; and (**c**) NLSR model. Each block is a square with angles of 5×5. The color of the block indicates the increase level of the gaze estimation accuracy.

5. Conclusions

This paper implements a gaze tracking and mapping system for monitoring driver eye gaze using a face camera and a scene camera. The proposed system applies a non-linear gaze point refinement mechanism to automatically facilitate the gaze prediction of the pre-trained model. Extensive experiments show the potential usage of several gaze models, and give the effectiveness of the proposed method. Future works should further investigate the impact of gaze estimation variance in long-term experiments, and apply it in the deep learning-based gaze estimation method.

Author Contributions: Conceptualization, Y.W. and G.Y.; methodology, Y.W.; validation, X.D.; formal analysis, X.D. and Y.W.; writing—original draft preparation, Y.W.; writing—review and editing, X.D. and G.Y.; supervision, G.Y. and X.F.; project administration, G.Y.; funding acquisition, X.F. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dong, Y.; Hu, Z.; Uchimura, K.; Murayama, N. Driver inattention monitoring system for intelligent vehicles: A review. *IEEE Trans. Intell. Transp. Syst.* **2010**, *12*, 596–614. [CrossRef]
2. Kaplan, S.; Guvensan, M.A.; Yavuz, A.G.; Karalurt, Y. Driver behavior analysis for safe driving: A survey. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 3017–3032. [CrossRef]
3. Sikander, G.; Anwar, S. Driver fatigue detection systems: A review. *IEEE Trans. Intell. Transp. Syst.* **2018**, *20*, 2339–2352. [CrossRef]
4. Khan, M.Q.; Lee, S. Gaze and eye tracking: Techniques and applications in ADAS. *Sensors* **2019**, *19*, 5540. [CrossRef] [PubMed]
5. Yang, L.; Dong, K.; Dmitruk, A.J.; Brighton, J.; Zhao, Y. A dual-cameras-based driver gaze mapping system with an application on non-driving activities monitoring. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 4318–4327. [CrossRef]
6. Dua, I.; John, T.A.; Gupta, R.; Jawahar, C. DGAZE: Driver Gaze Mapping on Road. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Las Vegas, NV, USA, 24 October–24 January 2020; pp. 5946–5953.
7. Fu, X.; Guan, X.; Peli, E.; Liu, H.; Luo, G. Automatic calibration method for driver's head orientation in natural driving environment. *IEEE Trans. Intell. Transp. Syst.* **2012**, *14*, 303–312. [CrossRef]
8. Yamashiro, K.; Deguchi, D.; Takahashi, T.; Ide, I.; Murase, H.; Higuchi, K.; Naito, T. Automatic calibration of an in-vehicle gaze tracking system using driver's typical gaze behavior. In Proceedings of the 2009 IEEE Intelligent Vehicles Symposium (IV), Xi'an, China, 3–5 June 2009; pp. 998–1003.
9. Yuan, G.; Wang, Y.; Yan, H.; Fu, X. Self-calibrated driver gaze estimation via gaze pattern learning. *Knowl.-Based Syst.* **2022**, *235*, 107630. [CrossRef]
10. Wang, J.; Chai, W.; Venkatachalapathy, A.; Tan, K.L.; Haghighat, A.; Velipasalar, S.; Adu-Gyamfi, Y.; Sharma, A. A Survey on Driver Behavior Analysis from In-Vehicle Cameras. *IEEE Trans. Intell. Transp. Syst.* **2021**, 1–24. [CrossRef]
11. Shehu, I.S.; Wang, Y.; Athuman, A.M.; Fu, X. Remote Eye Gaze Tracking Research: A Comparative Evaluation on Past and Recent Progress. *Electronics* **2021**, *10*, 3165. [CrossRef]
12. Wang, Y.; Zhao, T.; Ding, X.; Bian, J.; Fu, X. Head pose-free eye gaze prediction for driver attention study. In Proceedings of the 2017 IEEE International Conference on Big Data and Smart Computing (BigComp), Jeju, Korea, 13–16 February 2017; pp. 42–46.
13. Wang, Y.; Shen, T.; Yuan, G.; Bian, J.; Fu, X. Appearance-based gaze estimation using deep features and random forest regression. *Knowl.-Based Syst.* **2016**, *110*, 293–301. [CrossRef]
14. Wang, Y.; Zhao, T.; Ding, X.; Peng, J.; Bian, J.; Fu, X. Learning a gaze estimator with neighbor selection from large-scale synthetic eye images. *Knowl.-Based Syst.* **2018**, *139*, 41–49. [CrossRef]
15. Tawari, A.; Trivedi, M.M. Robust and continuous estimation of driver gaze zone by dynamic analysis of multiple face videos. In Proceedings of the 2014 IEEE Intelligent Vehicles Symposium (IV), Dearborn, MI, USA, 8–11 June 2014; pp. 344–349.
16. Jha, S.; Busso, C. Analyzing the relationship between head pose and gaze to model driver visual attention. In Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), Rio de Janeiro, Brazil, 1–4 November 2016; pp. 2157–2162.
17. Tawari, A.; Martin, S.; Trivedi, M.M. Continuous head movement estimator for driver assistance: Issues, algorithms, and on-road evaluations. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 818–830. [CrossRef]
18. Jha, S.; Busso, C. Probabilistic estimation of the driver's gaze from head orientation and position. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; pp. 1–6.
19. Tawari, A.; Chen, K.H.; Trivedi, M.M. Where is the driver looking: Analysis of head, eye and iris for robust gaze zone estimation. In Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), Qingdao, China, 8–11 October 2014; pp. 988–994.
20. Vicente, F.; Huang, Z.; Xiong, X.; De la Torre, F.; Zhang, W.; Levi, D. Driver gaze tracking and eyes off the road detection system. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 2014–2027. [CrossRef]
21. Wang, Y.; Yuan, G.; Mi, Z.; Peng, J.; Ding, X.; Liang, Z.; Fu, X. Continuous driver's gaze zone estimation using rgb-d camera. *Sensors* **2019**, *19*, 1287. [CrossRef] [PubMed]

22. Jha, S.; Busso, C. Estimation of Driver's Gaze Region from Head Position and Orientation Using Probabilistic Confidence Regions. *arXiv* **2020**, arXiv:2012.12754.

23. Lundgren, M.; Hammarstrand, L.; McKelvey, T. Driver-gaze zone estimation using Bayesian filtering and Gaussian processes. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 2739–2750. [CrossRef]

24. Yu, Z.; Huang, X.; Zhang, X.; Shen, H.; Li, Q.; Deng, W.; Tang, J.; Yang, Y.; Ye, J. A Multi-Modal Approach for Driver Gaze Prediction to Remove Identity Bias. In Proceedings of the 2020 International Conference on Multimodal Interaction, Online, 25–29 October 2020; pp. 768–776.

25. Lyu, K.; Wang, M.; Meng, L. Extract the Gaze Multi-dimensional Information Analysis Driver Behavior. In Proceedings of the 2020 International Conference on Multimodal Interaction, Virtual Event, The Netherlands, 25–29 October 2020; pp. 790–797.

26. Lollett, C.; Kamezaki, M.; Sugano, S. Towards a Driver's Gaze Zone Classifier using a Single Camera Robust to Temporal and Permanent Face Occlusions. In Proceedings of the 2021 IEEE Intelligent Vehicles Symposium (IV), Nagoya, Japan, 11–17 July 2021; pp. 578–585.

27. Yoon, H.S.; Hong, H.G.; Lee, D.E.; Park, K.R. Driver's eye-based gaze tracking system by one-point calibration. *Multimed. Tools Appl.* **2019**, *78*, 7155–7179. [CrossRef]

28. Li, N.; Busso, C. Detecting drivers' mirror-checking actions and its application to maneuver and secondary task recognition. *IEEE Trans. Intell. Transp. Syst.* **2015**, *17*, 980–992. [CrossRef]

29. Xing, Y.; Tang, J.; Liu, H.; Lv, C.; Cao, D.; Velenis, E.; Wang, F.Y. End-to-end driving activities and secondary tasks recognition using deep convolutional neural network and transfer learning. In Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV), Changshu, China, 26–30 June 2018; pp. 1626–1631.

30. Xing, Y.; Lv, C.; Wang, H.; Cao, D.; Velenis, E.; Wang, F.Y. Driver activity recognition for intelligent vehicles: A deep learning approach. *IEEE Trans. Veh. Technol.* **2019**, *68*, 5379–5390. [CrossRef]

31. Shahverdy, M.; Fathy, M.; Berangi, R.; Sabokrou, M. Driver behavior detection and classification using deep convolutional neural networks. *Expert Syst. Appl.* **2020**, *149*, 113240. [CrossRef]

32. Baltrusaitis, T.; Zadeh, A.; Lim, Y.C.; Morency, L.P. Openface 2.0: Facial behavior analysis toolkit. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG), Xi'an, China, 15–19 May 2018; pp. 59–66.

33. Jha, S.; Busso, C. Challenges in head pose estimation of drivers in naturalistic recordings using existing tools. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; pp. 1–6.

34. Araluce, J.; Bergasa, L.M.; Ocaña, M.; López-Guillén, E.; Revenga, P.A.; Arango, J.F.; Pérez, O. Gaze Focalization System for Driving Applications Using OpenFace 2.0 Toolkit with NARMAX Algorithm in Accidental Scenarios. *Sensors* **2021**, *21*, 6262. [CrossRef]

35. Shirpour, M.; Beauchemin, S.S.; Bauer, M.A. A probabilistic model for visual driver gaze approximation from head pose estimation. In Proceedings of the 2020 IEEE 3rd Connected and Automated Vehicles Symposium (CAVS), Victoria, BC, Canada, 18 November–16 December 2020; pp. 1–6.