A Novel Driving Behavior Learning and Visualization Method With Natural Gaze Prediction

GUOLIANG YUAN[iD], YAFEI WANG[iD], JINJIA PENG, AND XIANPING FU[iD]

Information Science and Technology College, Dalian Maritime University, Dalian 116026, China

Corresponding author: Yafei Wang (wangyafei@dlmu.edu.cn)

ABSTRACT Driving behavior analysis is vital for the advanced driving assistance system, aiming to improve driving behavior and decrease traffic accidents. Most existing driving behavior learning methods focus on either vehicle sensor information or driver's attention information, and provide a classification result on the current time data samples. The visualization of driving behavior on time series data samples could give an understanding and review of the driver's continuous actions. However, there has been little progress in combining the multi-modal vehicle and driver information on driving behavior learning and visualization. A multi-information driving behavior learning and visualization method with natural gaze prediction is proposed in this paper, which automatically integrates driver's gaze direction estimated from face camera, and various vehicle sensor data collected from on-board diagnostics (OBD) system. To accurately estimate the eye gaze under large head movement, a novel head pose-free eye gaze prediction method without calibration is proposed based on global and local scale sparse encoding, which treats the direction mapping as small gaze region classification. To understand driving behavior more intuitively, the latent features that represent different driving behaviors are extracted by FastICA from the fused time series data, and mapped into RGB color space for distinguished visualization. Experimental results demonstrate the effectiveness of the proposed method, and show that the proposed method performs better than the compared methods.

INDEX TERMS Driving behavior, gaze prediction, visualization.

I. INTRODUCTION

According to the technical survey, unsafe driving behavior is the main factor in today's traffic accidents [1], [2]. Identifying abnormal driving behavior and providing a meaningful evaluation of the driving risk level could help drivers revise their driving behavior and improve the driving safety. Based on the identification and evaluation results, systems like advanced driving assistance systems (ADAS) alert and guide the driver to avoid dangerous driving in time, such as fatigued driving, drunk driving and distracted driving [3]. The evaluation report is also beneficial for reasonable traffic signs assignment and user-friendly design of the in-vehicle layout [4]. However, drivers have different driving characteristics due to driving skills, driving preferences. Even for the same driver, the

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Awais Javed[iD].

driving behaviors may differ from situation to situation according to his/her physical states and emotions. To understand the driving behavior more simply, intuitive visualization of driver behavior is essential. The visualization of driving behavior could help drivers review their driving behaviors effectively to find incorrect driving behaviors by themselves and then get into good driving habits. Furthermore, it is also a critical tool for professionals in driving schools, transportation enterprises, and accident investigation departments to investigate and analyze driving behaviors efficiently [5], [6].

In recent years, a large number of driving behavior learning methods were proposed [7]–[9]. Head movements, body movements, facial features, and eye states are usually used to identify whether a driver's current state is in danger-driving, such as fatigue, drunk, drowsy, and distracted driving. Rongben *et al.* [10] distinguished the driver's mouth to normal, yawning, or talking state respectively by mouth

region's geometric features. Abtahi *et al.* [11] detected a yawning state in varying illumination conditions through the driver's facial features. Yan *et al.* [12] recognized the driver's hand movements to predict the driving behavior is safe or not. Vicente *et al.* [13] detected the facial landmark points by the geometric configurations of face shape with supervised descent method using the scale-invariant feature transform descriptor. After the facial landmark detection process, the corresponding head pose value and landmark features can be extracted. The facial landmark detection method has a fast convergence speed and a certain degree of robustness to light changes. However, this method cannot always find the eye landmark features if the drivers wear eyeglasses, making the eye gaze prediction perform poorly under the shifting light condition. Many head movement-based driver state analysis methods have been proposed to identify fatigue or distracted driving, and the detailed survey can be found in Ref. [14].

Meanwhile, various efforts have been made aiming at exploiting appearance representation to estimate the eye state. The appearance-based gaze prediction method has achieved better performance than the model-based method. Breitenstein *et al.* [15] utilized the depth appearance of the human nose as the appearance descriptor of the head pose, and trained the regression model by the reference depth appearance dataset. Their errors was calculated by the difference between the candidate value and the input value. Fridman *et al.* [16] exploited to segment the pupil blob by an adaptive threshold of the histogram, which did not work well in the non-uniform varying lighting conditions. Wang *et al.* [17] presented a simple-but-effective appearance-based gaze estimation framework with neighbor selection to fully use the large-scale synthetic eye images. Lu *et al.* [18] divided the eye images for three lines, which are utilized to calculate the value of each part to form a low dimensional representation. Wang *et al.* [19] introduced a deep feature extracted by convolutional neural networks to estimate the gaze regions. Lu and Chen [20] exploited to solve the problem of gaze estimation by training a dictionary with various bases of small patches of eye images. Although several appearance-based gaze estimation methods have achieved success in predicting gaze angle, these methods were only verified in controlled laboratory conditions. It might lead to the accuracy drops significantly due to the large head motions in the driving environment. Therefore, this paper focuses on the head pose-free eye gaze prediction, and applies the appearance-based gaze estimation under large head rotations.

Except for the driver state data, vehicle operating parameters and environmental data are also the significant factors of driving behavior detection, such as vehicle speed, steer angle, acceleration, brake, weather, road conditions. Lv *et al.* [21] explored the impact of influencing factors of driving behavior by regression function on multi-source data, such as vehicle speed, acceleration, weather, and road conditions. Wang and Wang [22] evaluated the driving behavior by clustering and classifying behavior parameters based on massive vehicle control data. Their results demonstrate that high-risk

driving behavior is related to vehicle speed. Zhang *et al.* [23] annotated six types of driving behavior using multivariate data collected by multiple sensors, and trained an end-to-end model to classify the driving event. Ohn-Bar *et al.* [24] combined line detection results, vehicle speed information, and steering information to analyze the driving behavior. Driver, vehicle, and environmental data are all involved in some studies of driving behavior analysis. Jin *et al.* [25] studied driving behavior during secondary tasks on both eye motion status and vehicle running status. Their research reflects the relevant importance among different secondary tasks on fused driving modes. Tian *et al.* [26] exploited to classify traffic density levels and its corresponding driver glance behavior by utilizing the surrounding vehicle patterns. It can be found that integrating multiple sources data to identifying driver behaviors is a general way, since driving is a complex task including driver, vehicle and environment factors. However, the driver's eye gaze has not been well considered in these methods.

The visualization of driving behavior is an important approach for the review and investigation of driving behaviors. However, there are rarely studies on the visualization of driving behavior data. Liu *et al.* [5], [6] proposed a method of measuring and visualizing the driving behavior by lots of sensor data, such as speed of wheels, angles of steering and the rate of accelerator opening. They used representative colors extracted by the deep sparse autoencoder to represent typical driving behaviors. Sama *et al.* [27], [28] proposed a driving feature clustering method and visualized different velocity styles of drivers on the map. Driving data, such as steering angle, brake position rate, acceleration, and velocity, were collected for the feature clustering and proactive driving behaviors reproducing. They used nine colors to represent the clustering results on acceleration and deceleration. Unlike these previous methods involving the visualization of driving behavior on steering wheel and velocity, this paper focuses on visualizing multi-modal fusion data from both vehicle source data and driver gaze data.

To analyze the driving behavior, this paper proposes a multi-information fusion and visualization, which contains three main modules: data acquisition, data fusion, and feature visualization. In the step of data acquisition, to generate the eye gaze data, a novel eye gaze prediction method without gaze calibration is proposed, which selects the sparse coefficients encoding eye image by linear combinations of bases from the learned dictionaries. The estimated gaze information is then embedding with vehicle data and normalized by sliding windows to output the data with time attribute. In the step of data fusion and feature visualization, to better exploit the latent features from these data, we utilize FastICA to fuse them and extract the distinct features with three-dimension. At last, these 3D features with GPS cues are mapped into RGB color space to represent various driving behaviors. As shown in Fig. 1, there is a color to represent it for different behavior. For instance, green is the look-straight with high speed driving and dark green represents the look-around
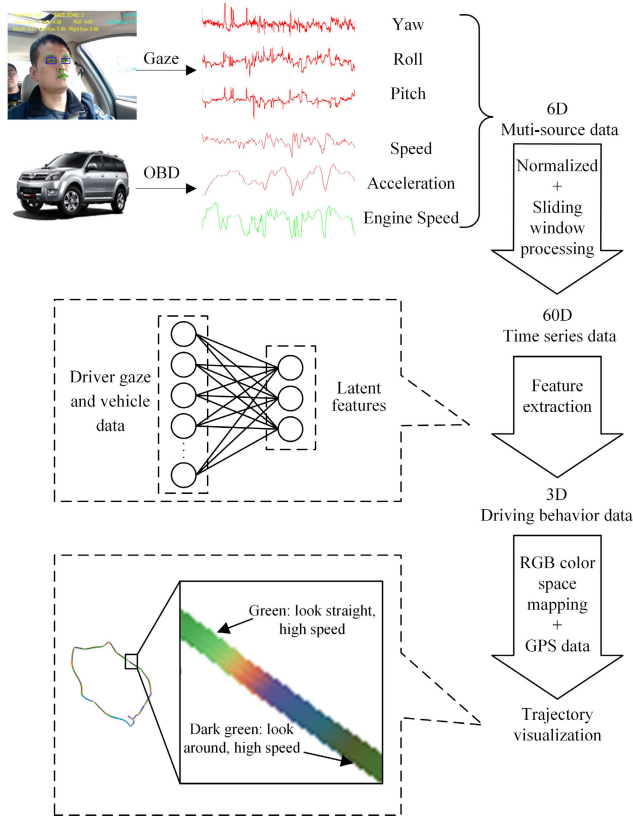
behavior with high speed driving. Our contributions are summarized as follows:

• A multi-information learning and visualization method is proposed to prompt better understanding and analyzing of the driving behavior from time series data, which integrates the information of driver's gaze data, and vehicle data, such as yaw, pitch, roll, speed, acceleration, and engine speed.

• A novel head pose-free eye gaze prediction method without gaze calibration is proposed, which combines head pose information, and multi-scale sparse features of eye images.

• The fused information is extracted to three-dimensional features by FastICA, and mapped into RGB color space for the representation of continuous driving behaviors on the trajectory map.

The rest of this paper is organized as follow. Section II-A introduces the proposed gaze prediction method. In Section II-B, the proposed driving behavior learning and visualization method is described. Several experiments are given in Section III. Section IV gives a brief conclusion.

II. PROPOSED METHOD

In this section, we mainly introduce the proposed multi-source information fusion and visualization method, which contains three main modules: data acquisition, data fusion, and feature visualization, as shown in Fig.1. In this research, a face camera is fixed in front of the windshield to
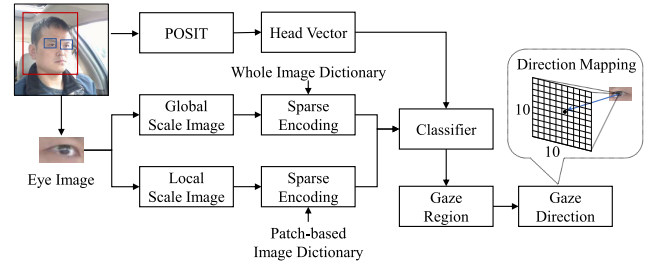

FIGURE 2. The structure of gaze direction estimation.

capture the driver's face data, while an OBD data recorder is used to collect the vehicle data. The driver's data includes yaw, pitch, and roll, is estimated by the eye gaze prediction method. The obtained driver's information is fused with vehicle data (speed, acceleration, and engine speed information). All embedded information is normalized, and processed by sliding windows to output the time attributed data. Then, FastICA is utilized to exploit three-dimensional latent features. The latent features is mapped to RGB color space to produce a colored trajectory of driving behaviors.

A. GAZE DATA ACQUISITION

To acquire the driver's eye gaze data, an appearance-based calibration-free eye gaze prediction method is used, with head vector estimation, and eye image sparse dictionary learning. The overall framework of the proposed gaze estimation method is shown in Fig.2. Firstly, the driver's face region is detected and localized by a landmark detector, in which the eye image is cropped from the detected face region. The head vector is then computed by the geometric method with a weak perspective projection model using POSIT (Pose from Orthography and Scaling with ITerations). The eye images are encoded by global and local sparse dictionaries for extracting detailed structural information. To deal with eye gaze estimation under large head motion, a method based on support vector machine (SVM) is utilized to classify the gaze into small gaze region. Finally, the gaze region is converted to gaze direction by grid mapping on the partitioned gaze zone. Here, we introduce the eye gaze estimation method.

1) MULTI-SCALE SPARSE ENCODING

To achieve more detailed information, the sparse coding with multi-scale is employed, which includes coding global images and local scale images partitioned from global images. Fig.3 shows the examples of global images and local images. To be specific, $X = [x_1, x_2, \ldots, x_n]$ in $\mathbb{R}^{m \times n}$ are input eye image set, where x_n is a column vector at global scale. The local scale set of the corresponding eye image could be described as $X'_n = [x_n^1, x_n^2, \ldots, x_n^p]$, where x_n^p is one distinct part of x_n. Hence, the problem of spare coding could be describe as constructing a dictionary $D = [d_1, d_2, \ldots, d_k]$ in $\mathbb{R}^{m \times k}$ and its corresponding coefficients $a = [a_1, a_2, \ldots, a_n]$ in $\mathbb{R}^{k \times n}$ that can minimize $\|X - Da\|^2$ with $\|a\|_1 < \varepsilon$. As introduced in Ref. [29], it could be
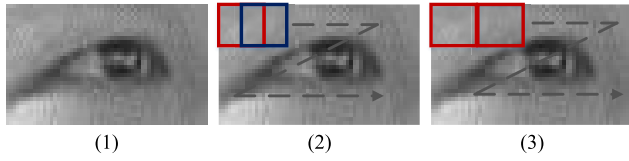
FIGURE 3. Example of eye images partition in (1) the whole image as global scale, (2) overlapped patches as local scale and (3) distinct patches as local scale.
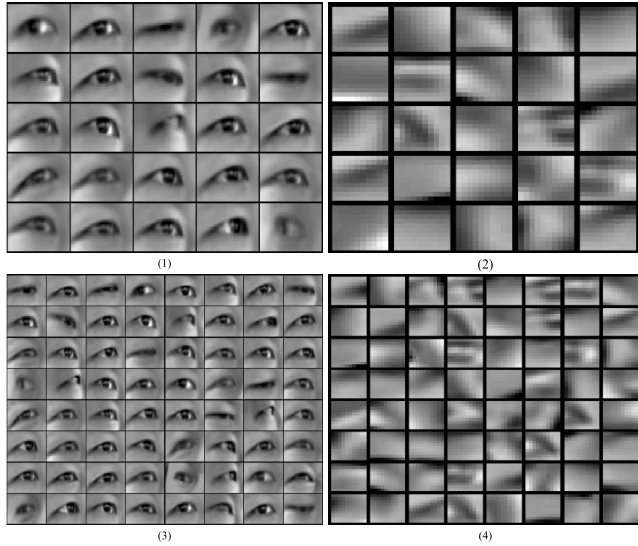

FIGURE 4. Examples of dictionaries constructed by full size images and distinct patch-based images. (1) and (3) are dictionaries with 25 and 64 bases on the global scale. (2) and (4) are dictionaries with 25 and 64 bases on the local scale.
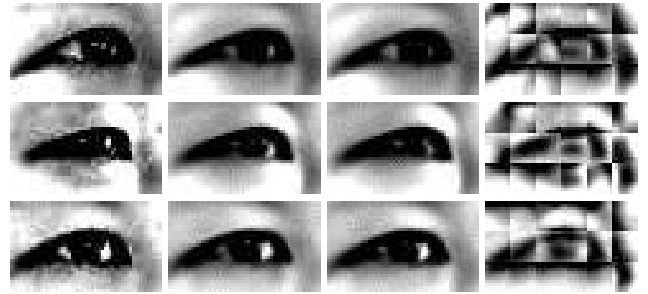

FIGURE 5. Examples of reconstructed images. These columns are eye images, reconstructed images with 25 bases and 64 bases dictionary at the global scale. The approximate result of the 25-bases patch scale dictionary is shown in the last column.

regarded as solving ℓ^1 sparse coding problem by

$$
\begin{aligned}
f(x, D) \\
= \underset{D \in C, a_i \in \mathbb{R}^{k \times n}}{\arg \min} \sum_{i=1}^{n} \|x_i - D a_i\|_2^2 + \lambda \|a_i\|_1 \\
s.t.\ C \equiv \{D \in \mathbb{R}^{m \times k} : \|d_i\|_2 \le 1\ \forall i = 1, 2, \dots, n\}
\end{aligned}
\tag{1}
$$

where λ represents a regularization parameter. Inspired by [29], the dictionary can be solved as the sparse approximation and learned in an online manner.

During the process of learning the dictionary, a and D are updated by the block-coordinate approach [29] in an asynchronous manner that fixes variables besides the updated variable.

$$
\underset{D \in C}{\arg \min} \sum_{i=1}^{n} \|x_i - D a_i\|_2^2 + \lambda \|a_i\|_1
\tag{2}
$$

Similar with the global dictionary, the dictionary of local images D' also could be calculated. As shown in Fig.4, typical dictionaries in different scales that are based on 25 bases and 64 bases are learned. After obtaining the global dictionary D, the reconstruction coefficient a of the input image can be calculated by LARS [30]:

$$
\hat{a} = \underset{a \in \mathbb{R}^{k \times n}}{\arg \min} \|x - D a\|_2^2 + \lambda \|a\|_1
\tag{3}
$$

Similar with the global scale, the reconstruction coefficient $a' = [a'_1, a'_2, \dots, a'_p]$ of the local image $X' = [x'_1, x'_2, \dots, x'_p]$ also could be calculated by LARS [30] with the learned local dictionary D':

$$
\hat{a}' = \underset{a' \in \mathbb{R}^{q \times p}}{\arg \min} \|x' - D' a'\|_2^2 + \lambda \|a'\|_1
\tag{4}
$$

Thus, for each input image x, it can be constructed by $D\hat{a}$. As shown in Fig.5, it gives the results of constructing eye images based on different bases. Significantly, the reconstructed images using our method are close to the original images.

2) GAZE PREDICTION

In the naturalistic driving conditions, the coarse gaze directions are always denoted by gaze regions. Hence, in this paper, inspired by Ref. [31], we build a gaze region estimator to classify the gaze region that combines the features of eye images and the input head vector.

To be specific, the head pose information and the sparse eye features could be defined as $E = [e_1, e_2, \dots, e_n]$, where $e_n = [V_{head}, V_{global}, V_{local}]^T$, V_{head} is the head vector, V_{global} and V_{local} are sparse encoding coefficients of the global scale and the local scale, respectively. Moreover, the gaze region labels of them are $Y = [y_1, y_2, \dots, y_n]$.

Then, SVM is employed to predict the driver's gaze direction, containing several binary classifiers with $G(G-1)/2$ internal nodes and G leaves. G is the number of gaze region classes. According to Ref. [32], for each classifier, it could be calculated by:

$$
\begin{aligned}
\underset{w^{ij}, b^{ij}, \xi^{ij}}{\min} \quad & \frac{1}{2} (w^{ij})^T w^{ij} + Z \sum_t \xi_n^{ij} \\
& (w^{ij})^T \varphi(e_n) + b^{ij} \ge 1 - \xi_n^{ij},\ if\ y_n = i, \\
& (w^{ij})^T \varphi(e_n) + b^{ij} \le -1 + \xi_n^{ij},\ if\ y_n = j, \\
& \xi^{ij} \ge 0
\end{aligned}
\tag{5}
$$

where *i*-th and *j*-th represent different classes in the training dataset, ξ and Z are the penalty items, φ means a linear kernel function. It is equivalent to minimize $\frac{1}{2}(w^{ij})^T w^{ij}$ and maximize $2/\|w^{ij}\|$.

B. DATA FUSION

To remove the correlation of the data, the whitened [33] is employed, which utilizes the standard PCA to decrease the complexity of the subsequent algorithm. After whitened, all data from multi-source driving dataset are normalized, which could be computed by:

$$y_{t,d} = \begin{cases} \dfrac{2x_{t,d}}{\pi}, & 1 \le d \le 3 \\ \dfrac{2(x_{t,d} - x_{d_{min}})}{x_{d_{max}} - x_{d_{min}}} + 1, & otherwise \end{cases} \quad (6)$$

where $x_{t,d}$ is the raw data while the gaze estimation is denoted by $1 \le d \le 3$, $x_{t,d}$. Otherwise, $x_{t,d}$ represents the vehicle data. $x_{d_{max}} = \max(x_{1,d}, x_{1,d}, \ldots, x_{N_X,d})$, $x_{d_{min}} = \min(x_{1,d}, x_{1,d}, \ldots, x_{N_X,d})$. $y_t = (y_{t,1}, y_{t,2}, \ldots, y_{t,N})^T \in \mathbb{R}^D$ represents the data is normalized at time t.

To exploit the features with the time attribute, the paper employs the windows sliding process with outliers reducing. Assume that the window size of the sliding window is w, the dimension of processed time series data is $D_V = w \times D$. Hence, at the time t, the data of the time series could be described as:

$$v_t = (y_{t-w+1}^T, y_{t-w+2}^T, \ldots, y_t^T)^T \in \mathbb{R}^{D_V} (t \ge w) \quad (7)$$

After achieving the fusion data by normalizing these existing multi-source data, such as gaze information, speed data and acceleration information, FastICA [33] is applied to extract the latent features, which searches the latent features in the direction of maximum entropy and its objective function $J(W)$ could be formulated as:

$$J(W) = [E\{G(W^T Z)\} - E\{G(V)\}]^2 \quad (8)$$

where $G(\cdot)$ represents a nonlinear function, where $G(\cdot) = \tanh(\cdot)$. V means a random Gaussian vector. E represents the process of whitened. At the extreme point of $E\{G(W^T Z)\}$, we could get the maximum of $J(W)$, which could utilizing the constraint $E\{(W^T Z)^2\} = \|W\|^2 = 1$ to be calculated by:

$$E\{Zg(W^T Z)\} + \beta W = 0 \quad (9)$$

where $\beta = E\{W_0^T Zg(W_0^T Z)\}$, W is initialized by W_0, and the derivative of $G(\cdot)$ is $g(\cdot)$. Hence, the function could be denoted as:

$$F(W) = E\{Zg(W^T Z)\} + \beta W \quad (10)$$

Solving the above equation with approximated Newton iterative method:

$$W_{k+1} = W - \frac{E\{Zg(W^T Z)\} + \beta W}{E\{g'(W^T Z)\} + \beta} \quad (11)$$

To simplify the iterative function, FastICA could be described as:

$$W_{k+1} = E\{Zg(W^T Z)\} - E\{g'(W^T Z)\}W \quad (12)$$

When the setted dimension of extracted features and the threshold of convergence are reached, the algorithm would

TABLE 1. Average loss of reconstructed images with different dictionaries.

Dataset	Dictionaries					
	F25	F64	F100	F225	F400	F1024
Train	0.17	0.16	0.16	0.15	0.15	0.14
Test	0.23	0.21	0.21	0.20	0.20	0.20
Dataset	Dictionaries					
	S25	S64	S100	S225	S400	S1024
Train	0.20	0.18	0.18	0.17	0.16	0.15
Test	0.23	0.21	0.20	0.19	0.19	0.18

TABLE 2. Average PSNR of reconstructed images with different dictionaries.

Dataset	Dictionaries					
	F25	F64	F100	F225	F400	F1024
Train	20.89	22.20	22.73	23.53	23.97	24.88
Test	17.28	17.93	18.20	18.48	18.61	18.87
Dataset	Dictionaries					
	S25	S64	S100	S225	S400	S1024
Train	23.49	24.71	25.23	26.08	26.58	27.28
Test	21.35	22.32	22.85	23.47	23.91	24.53

be over. Through the symmetric orthogonalization, the separation matrix W could be obtained, which each separation matrix represents one latent feature of driving behavior.

C. FEATURE VISUALIZATION

To better understand and analyze the driver behavior with multi-source data, in our paper, the obtained 3D feature data are represented by mapping them into a color space [6], which can be described as:

$$rgb_{t,d} = \frac{w_{t,d}^{(final)} - w_{min_d}^{(final)}}{w_{max_d}^{(final)} - w_{min_d}^{(final)}} \quad (13)$$

where, at the time t, $rgb_{t,d}$ and $w_{t,d}^{(final)}$ are the d-th element of the 3D features and the 3D eigenvector, respectively. d-th element is the value of *green*, *red* and *blue*. On the d-th element, its maximum and minimum values are $w_{max_d}^{(final)}$ and $w_{min_d}^{(final)}$. Combining the GPS data and the fusion data, the colored driving curve could be realized to represent the driver's various behaviors.

III. EXPERIMENTAL RESULTS

A. GAZE DIRECTION RESULTS

To verify the proposed gaze estimation method, the public eye gaze dataset UTMultiView [34] is utilized to test our proposed method of estimating gaze regions. There are 23,040 training samples and 1,280 testing samples. The number of head pose and gaze directions in the training set are 144 and 160, respectively. The number of head pose and gaze directions in the testing set are 8 and 160, respectively. The angle of horizontal gaze poses is $-55°$ to $+55°$ and the angle of vertical gaze poses is $-55°$ to $+65°$.

In our experiments, the data is divided into several regions to represent the gaze regions. For example, there are 4 regions

TABLE 3. Results of gaze classification (%) with different dictionaries and methods under different setting number of gaze regions.

Gaze zone number	F25	F64	F100	F225	F400	F1024	S25	S64	S100	S225	S400	S1024	F1024 +S25	F1024 +S25+H	HoG +RF	Raw +SM
4	81.40	84.45	85.78	86.09	86.87	87.73	88.90	87.18	83.82	83.82	85.07	85.85	86.32	89.2	84.92	87.5
9	75.39	76.32	77.03	77.03	77.81	79.37	75.15	73.75	71.25	71.56	71.32	70.23	80.54	81.8	72.42	78.81
16	62.34	66.71	68.35	68.04	69.45	71.95	67.5	65.46	62.5	62.89	64.29	63.67	70.15	73.2	65.00	72.89
25	55.39	61.09	65.70	63.12	64.37	65.46	58.67	60.23	58.43	57.42	57.5	54.53	61.48	65.1	59.22	62.39
36	46.64	53.82	56.25	55.46	57.03	58.98	51.32	48.51	43.20	45.23	44.68	41.09	60.93	62.9	51.64	56.40
64	38.43	43.35	47.34	45.46	47.57	49.14	34.29	33.59	35.31	34.76	36.56	33.04	48.28	51.8	42.89	46.56
100	25.85	32.26	36.87	36.32	39.21	40.93	20.78	22.73	22.81	21.79	20.23	19.84	38.75	41.2	29.69	33.06
144	24.29	28.04	31.64	30.85	32.03	36.64	15.39	15.07	15.93	15.70	16.71	14.84	34.21	37.5	27.27	31.23
225	17.18	21.17	24.84	24.76	23.67	26.87	7.89	8.82	8.28	10.07	9.14	10.54	27.34	27.9	16.80	24.76
400	11.79	14.60	16.56	18.20	17.57	19.60	3.28	4.53	4.21	4.06	3.67	4.29	18.35	20.9	10.00	13.04

with the same size $55° \times 60° (horizontal \times vertical)$. Similar with that, if there are k^2 gaze regions, each of them is $(110/k)° \times (120/k)°$. To analyze the proposed method detailedly, the space in our paper is divided into several standards that include 4, 9, 16, 25, 36, 64, 100, 144, 225, and 400 gaze regions that contain 2, 3, 4, 5, 6, 10, 12, 15, 20, respectively.

1) COMPARISONS ON IMAGE RECONSTRUCTION USING DIFFERENT DICTIONARIES

In our experiments, the dictionaries are learned with different base numbers, which include 25, 64, 100, 225, 400, and 1,024 at the scale of global and local. For instance, 'F25' is the dictionary with 25 bases learned from the 'Full' image representing the global scale. 'S64' represents the 64-bases dictionary learned at the local scale from the patched images. Hence, we can learn a series of dictionaries with essential information about input eye images. As shown in Table 1, the ratio of reconstruction error could be quantified with these obtained dictionaries, where the lower value of loss means the higher accuracy of reconstruction.

The loss value is computed by

$$R = \frac{1}{2n}\|x - Da\|_2^2 + \lambda\|a\|_1 \tag{14}$$

The PSNR is employed to evaluate the effect of those reconstructed images by different dictionaries, which could be calculated by the following:

$$PSNR = 20 \times \log_{10}\frac{MAX}{\sqrt{MSE}} \tag{15}$$

where MSE is the mean square error between the reconstructed image and its corresponding original image. MAX represents the maximum value of the reconstructed image. If the PSNR is larger, the reconstruction image performs better. Table 2 shows the results of the evaluation. It is significant that the accuracy of reconstruction increases along with the larger base.

TABLE 4. MAE of gaze estimation (degree) with different setting number of gaze regions.

Gaze zone number	F1024 +S25	F1024 +S25+H	HoG +RF	Raw +SM
4	14.80	14.39	15.44	14.67
9	10.64	10.36	11.51	10.75
16	8.79	8.28	9.44	8.37
25	7.87	7.29	8.24	7.44
36	7.04	6.95	7.99	7.34
64	6.11	5.99	7.29	6.71
100	5.90	5.49	6.88	6.40
144	5.60	5.15	7.11	6.33
225	5.41	5.29	6.95	6.17
400	5.55	5.24	7.03	6.32

2) COMPARISONS WITH BASELINE METHODS

To demonstrate the effectiveness of the proposed method, some existing baseline methods, such as HoG+RF, Raw+SM are compared on the dataset. In Table 3 and 4, 'F1024+S25+H' and 'F1024+S25' represent the proposed method with and without head pose information, respectively. Besides, the accuracy of gaze estimation and the mean absolute error are shown in Table 3 and 4. In this experiment, instead of the ground-truth gaze angle in the dataset, gaze region center angle is employed as the gaze angle for training and testing eye images, which is similar to Ref. [20]. Hence, the gaze estimation error could be calculated by transferring the gaze classification labels into gaze angles.

From these results, it should be noted that, for the problem of gaze region prediction, the accuracy of gaze region estimation that constructs dictionaries with different base numbers varies a lot. Moreover, the dictionaries of full size image perform better than dictionaries of patch-based images. Moreover, when combining the head pose information, it could achieve the highest accuracy among all compared methods.
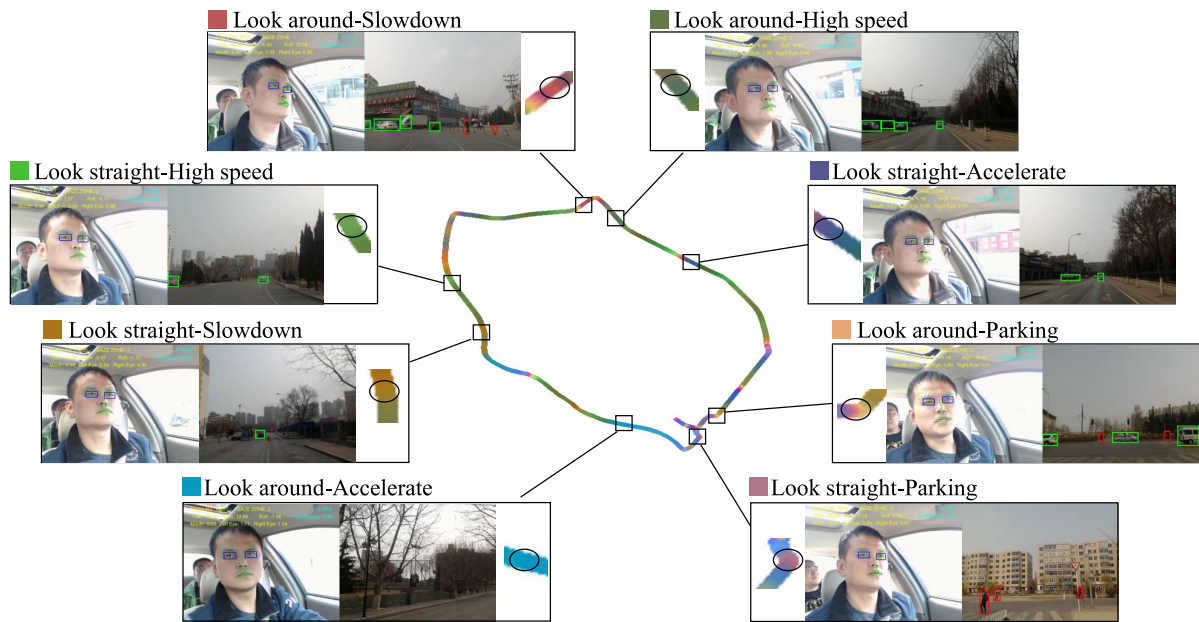
FIGURE 6. Visualization results of the proposed method. The magnified color trajectories are shown with the images captured from the driver face camera and road scene camera. There are eight basic driving behaviors to be regarded as the head orientation {Look straight, Look around} and the vehicle speed {Accelerate, Slowdown, High speed, Parking}. Results of our method are mapped into the RGB space for visualization with the fused driver-related data.

TABLE 5. Variable details of raw data at time t.

Name	Description	Unit	Range
$x_{t,1}$	yaw of gaze angle	degree	(-90,90)
$x_{t,2}$	pitch of gaze angle	degree	(-45,45)
$x_{t,3}$	roll of gaze angle	degree	(-30,30)
$x_{t,4}$	speed	m/s	[0,70)
$x_{t,5}$	acceleration	m^2/s	(-30,30)
$x_{t,6}$	engine speed	revolutions per minute	(0,3000)

B. VISUALIZATION RESULTS

To verify the effectiveness of the proposed visualization method, a naturalistic driving dataset was built. One expert driver was participated to drive an experimental car through two school campuses in a close-loop route. The driver drove the same route five times. During the driving experiment, the driver drove the car in regular operation. Therefore, the driver met different driving situations for car manipulation and attention behavior, such as traffic light, overtake other cars, and parked cars. This paper obtains eight typical driving behaviors according to the driving state on the velocity style and attention style. These driving behaviors can be regarded as the combination of the head orientation (Look straight and Look around) and the vehicle speed (Accelerate, Slowdown, High speed, and Parking).

As the driver drove the car, vehicle data was collected by an OBD reader, with a sampling rate of 50Hz, providing the vehicle's speed, acceleration, and engine speed. The OBD reader has direct access to the vehicle. A portable device was used to collect the GPS data. The driver's images were captured from a normal USB camera, with an image resolution of 640 × 480 and a frame rate of 30Hz. The camera was mounted at a certain angle from the driver, on the windshield's upper left, and close to the driver visor. The driver's gaze information was calculated by the proposed gaze prediction method in Section II-A, providing the driver's gaze angle (yaw, pitch, and roll). The descriptions of raw data are given in Table. 5.

All raw data is normalized and formatted by the sliding window. High-dimensional time series data is extracted and reduced to three-dimensional driving behavior data, as described in Section II-B. The dataset has 12,550 group valid data after raw data timeline alignment and sliding windows processing. In our experiments, the size of the sliding window is set to 100.

In this section, different with the visualization results, F-measure(m_F) is employed to evaluate our method, which could be computed by:

$$m_F = \frac{2pr}{p+r} \tag{16}$$

where, $p = P_T/(P_T + N_F)$ is the accuracy rate, $r = P_T/(P_T + P_F)$ is the recall rate. P_T, N_F and P_F are the number of real class data, negative data and negative positive data, respectively. If F-measure is higher, the correct rate of the classification is also higher.

The compared methods in our experiment employ different feature extraction approaches that include PCA, SAE [6], DSAE and YAS, which could extract 3D features for visualization with GPS data. In Ref. [35], the yaw of the head pose could represent major head orientation when it has a

TABLE 6. The color meanings of driving behavior processed by different methods. The underlined colors are different driving behaviors with similar color.

Driving behavior		PCA	SAE	DSAE	FastICA
Look-straight	Accelerate				
	Slowdown				
	High speed				
	Parking				
Look-around	Accelerate				
	Slowdown				
	High speed				
	Parking				

TABLE 7. F-measure of driving behaviors classified by compared methods.

Driving behavior		PCA	SAE	DSAE	FastICA
Look-straight	Accelerate	**0.798**	0.454	0.594	0.681
	Slowdown	0.357	0.423	0.455	**0.635**
	High speed	0.528	0.417	0.517	**0.705**
	Parking	0.854	0.538	0.406	**0.869**
Look-around	Accelerate	**0.312**	0.299	0.206	0.296
	Slowdown	0.127	**0.317**	0.196	0.281
	High speed	0.133	0.291	**0.486**	0.447
	Parking	0.303	0.095	0.178	**0.315**
Average value		0.427	0.354	0.380	**0.529**

large rotation. Thus, there are eight basic driving behaviors to be regarded as the head orientation {Look straight, Look around} and the vehicle speed {Accelerate, Slowdown, High speed, Parking }. As shown in Fig.6, the results of our method is mapped into the RGB space for visualization with the detected driver-related data. And the related meanings of the colors are shown in Table 6, in which the underlined features of the same feature extraction method are different driving behaviors with a similar color. From the Fig.6, it could be found that, the proposed method is effective to visualize the various driving behaviors and can be used to distinguish the different easy-confused behaviors.

Besides the visualization, Table 7 shows the results obtained by a well-trained SVM classifier. To be specific, we randomly selected 2,000 groups data with corresponding driving behavior labels for testing. The other data in the dataset is utilized as the training dataset. As shown in Table 7, it shows the driving behaviors distinguished by different methods. It should be noted that, for the look-around behavior, PCA and SAE perform better than FastICA. However, for most driving behaviors, it is obvious that FastICA could achieve a higher F-measure value. These results also verify

that the visualization with RGB color is easier to distinguish more driving behaviors for representation.

IV. CONCLUSION

In this paper, we propose a multi-source information embedded method to better understand and analyze the driving behavior. In our method, to estimate eye gaze accurately, a novel eye gaze prediction method without gaze calibration is proposed, which is head pose-free and combines the head pose information and multi-scale sparse features of eye images to determine the gaze direction. The estimated gaze information is embedding with vehicle data to form the fused data. And then, FastICA is exploited to find more driving-related information for visualization and understanding the driving behavior.

REFERENCES

[1] M. Miyaji, M. Danno, and K. Oguri, "Analysis of driver behavior based on traffic incidents for driver monitor systems," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2008, pp. 930–935.

[2] J. Liang, X.-Y. Cheng, and X.-B. Chen, "The research of car rear-end warning model based on MAS and Behavior," in *Proc. Workshop Power Electron. Intell. Transp. Syst.*, Aug. 2008, pp. 305–309.

[3] M. H. Alkinani, W. Z. Khan, and Q. Arshad, "Detecting human driver inattentive and aggressive driving Behavior using deep learning: Recent advances, requirements and open challenges," *IEEE Access*, vol. 8, pp. 105008–105030, 2020.

[4] Y. Liao, M. Wang, L. Duan, and F. Chen, "Cross-regional driver–vehicle interaction design: An interview study on driving risk perceptions, decisions, and ADAS function preferences," *IET Intell. Transp. Syst.*, vol. 12, no. 8, pp. 801–808, Oct. 2018.

[5] H. Liu, T. Taniguchi, T. Takano, Y. Tanaka, K. Takenaka, and T. Bando, "Visualization of driving Behavior using deep sparse autoencoder," in *Proc. IEEE Intell. Vehicles Symp.*, Jun. 2014, pp. 1427–1434.

[6] H. Liu, T. Taniguchi, Y. Tanaka, K. Takenaka, and T. Bando, "Visualization of driving behavior based on hidden feature extraction by using deep learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 9, pp. 2477–2489, Sep. 2017.

[7] T. K. Chan, C. S. Chin, H. Chen, and X. Zhong, "A comprehensive review of driver behavior analysis utilizing smartphones," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 10, pp. 4444–4475, Oct. 2020.

[8] N. Lin, C. Zong, M. Tomizuka, P. Song, Z. Zhang, and G. Li, "An overview on study of identification of driver characteristics for automotive control," *Math. Problems Eng.*, vol. 2014, pp. 1–15, Jan. 2014.

[9] H.-B. Kang, "Various approaches for driver and driving behavior monitoring: A review," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2013, pp. 616–623.

[10] W. Rongben, G. Lie, T. Bingliang, and J. Lisheng, "Monitoring mouth movement for driver fatigue or distraction with one camera," in *Proc. 7th Int. IEEE Conf. Intell. Transp. Syst.*, Oct. 2004, pp. 314–319.

[11] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi, and B. Hariri, "Yawdd: A yawning detection dataset," in *Proc. 5th ACM Multimedia Syst. Conf.*, 2014, pp. 24–28.

[12] C. Yan, F. Coenen, Y. Yue, X. Yang, and B. Zhang, "Video-based classification of driving behavior using a hierarchical classification system with multiple features," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 30, no. 5, Jun. 2016, Art. no. 1650010.

[13] F. Vicente, Z. Huang, X. Xiong, F. De la Torre, W. Zhang, and D. Levi, "Driver gaze tracking and eyes off the road detection system," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 2014–2027, Aug. 2015.

[14] A. Mittal, K. Kumar, S. Dhamija, and M. Kaur, "Head movement-based driver drowsiness detection: A review of state-of-art techniques," in *Proc. IEEE Int. Conf. Eng. Technol. (ICETECH)*, Mar. 2016, pp. 903–908.

[15] M. D. Breitenstein, D. Kuettel, T. Weise, L. V. Gool, and H. Pfister, "Real-time face pose estimation from single range images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[16] L. Fridman, J. Lee, B. Reimer, and T. Victor, "'owl' and 'lizard': Patterns of head pose and eye pose in driver gaze classification," *IET Computer Vision*, vol. 10, no. 4, pp. 308–314, 2016.

[17] Y. Wang, T. Zhao, X. Ding, J. Peng, J. Bian, and X. Fu, "Learning a gaze estimator with neighbor selection from large-scale synthetic eye images," *Knowl.-Based Syst.*, vol. 139, pp. 41–49, Jan. 2018.

[18] F. Lu, Y. Sugano, T. Okabe, and Y. Sato, "Adaptive linear regression for appearance-based gaze estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 10, pp. 2033–2046, Oct. 2014.

[19] Y. Wang, T. Shen, G. Yuan, J. Bian, and X. Fu, "Appearance-based gaze estimation using deep features and random forest regression," *Knowl.-Based Syst.*, vol. 110, pp. 293–301, Oct. 2016.

[20] F. Lu and X. Chen, "Person-independent eye gaze prediction from eye images using patch-based features," *Neurocomputing*, vol. 182, pp. 10–17, Mar. 2016.

[21] B. Lv, R. Yue, and Y. Zhang, "The influence of different factors on right-turn distracted driving behavior at intersections using naturalistic driving study data," *IEEE Access*, vol. 7, pp. 137241–137250, 2019.

[22] X. Wang and H. Wang, "Driving Behavior clustering for hazardous material transportation based on genetic fuzzy C-means algorithm," *IEEE Access*, vol. 8, pp. 11289–11296, 2020.

[23] J. Zhang, Z. Wu, F. Li, J. Luo, T. Ren, S. Hu, W. Li, and W. Li, "Attention-based convolutional and recurrent neural networks for driving Behavior recognition using smartphone sensor data," *IEEE Access*, vol. 7, pp. 148031–148046, 2019.

[24] E. Ohn-Bar, A. Tawari, S. Martin, and M. M. Trivedi, "On surveillance for safety critical events: In-vehicle video networks for predictive driver assistance systems," *Comput. Vis. Image Understand.*, vol. 134, pp. 130–140, May 2015.

[25] L. Jin, B. Guo, Y. Jiang, F. Wang, X. Xie, and M. Gao, "Study on the impact degrees of several driving behaviors when driving while performing secondary tasks," *IEEE Access*, vol. 6, pp. 65772–65782, 2018.

[26] R. Tian, J. Wang, L. Li, and Y. Chen, "Connecting road environment features and driver glance behavior in the macro level: Surrounding vehicle patterns, traffic density, and driver eye-glance behaviors," in *Proc. 17th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2014, pp. 1614–1619.

[27] K. Sama, Y. Morales, N. Akai, H. Liu, E. Takeuchi, and K. Takeda, "Driving feature extraction and behavior classification using an autoencoder to reproduce the velocity styles of experts," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 1337–1343.

[28] K. Sama, Y. Morales, H. Liu, N. Akai, A. Carballo, E. Takeuchi, and K. Takeda, "Extracting human-like driving behaviors from expert driver data using deep learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 9, pp. 9315–9329, Sep. 2020.

[29] J. Mairal, F. Bach, and J. Ponce, "Sparse modeling for image and vision processing," 2014, *arXiv:1411.3230*. [Online]. Available: http://arxiv.org/abs/1411.3230

[30] R. Tibshirani, I. Johnstone, T. Hastie, and B. Efron, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, Apr. 2004.

[31] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, "Appearance-based gaze estimation in the wild," in *Proc. IEEE Conf. Comput. Vis. pattern Recognit.*, Jun. 2015, pp. 4511–4520.

[32] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.

[33] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, nos. 4–5, pp. 411–430, Jun. 2000.

[34] Y. Sugano, Y. Matsushita, and Y. Sato, "Learning-by-synthesis for appearance-based 3D gaze estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1821–1828.

[35] X. Fu, X. Guan, E. Peli, H. Liu, and G. Luo, "Automatic calibration method for driver's head orientation in natural driving environment," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 303–312, Mar. 2013.

GUOLIANG YUAN received the B.E. and M.E. degrees from Dalian Maritime University, in 2005 and 2013, respectively, where he is currently pursuing the Ph.D. degree. His research interests include computer vision, image processing, and eye gaze tracking.
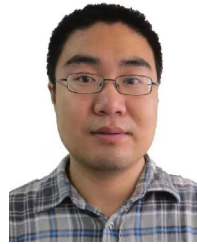
YAFEI WANG received the Ph.D. degree in electronics science and technology from the Dalian University of Technology, Dalian, China, in 2018. He is currently a Postdoctoral Research Fellow with Dalian Maritime University. His research interests include image processing, computer vision, eye gaze tracking, and machine learning.

JINJIA PENG received the B.E. and M.E. degrees from Dalian Maritime University, in 2014 and 2017, respectively, where she is currently pursuing the Ph.D. degree. Her research interests include computer vision and image processing.

XIANPING FU received the Ph.D. degree in communication and information system from Dalian Maritime University, Dalian, China, in 2005. From 2008 to 2009, he was a Postdoctoral Fellow with the Harvard Medical School, Schepens Eye Research Institute, Boston, MA, USA. He is currently a Professor with the Information Science and Technology College, Dalian Maritime University. His research interests include perception of natural scenes in engineering systems, including multimedia, image/video processing, and object recognition.

● ● ●