

Promotion of Proactive Behavioral Change to Prevent Human Error

- Part 4 Estimation of Driver's Head Pose, Gaze, and Driving Posture and Visualization of Gaze Targets Using In-Vehicle Cameras for Driving Behavior Analysis -

Ishihara Ryota¹⁾ Yamasaki Akito¹⁾ Yoshitake Hiroshi²⁾ Shino Motoki²⁾

1) Meijo University, 1-501 Shiogamaguchi, Tempaku-ku, Nagoya, Aichi, 468-8502, Japan

2) Institute of Science Tokyo, 2-12-1 Ookayama, Meguro-ku, Tokyo, 152-8550, Japan

KEY WORDS: Human Engineering, Human Error, Driver Behavior, Driver Sensing/Driver Monitoring, Driving Posture, Visibility, Gaze Estimation, Head Pose Estimation, In-Vehicle Camera (C2)

1. Introduction

Analysis and evaluation of driver behavior are indispensable for designing driving assistance that reduces traffic accidents. At an intersection with poor visibility, the crossing road is occluded, so drivers not only turn their heads but also lean forward to check the blind spot as early as possible. Evaluating such a scene therefore requires a postural index in addition to head pose and gaze, together with identification of the gaze target. This study proposes a method that estimates the head pose, gaze and posture of a driver in a non-contact manner from in-vehicle camera images and visualizes gaze targets by projecting the estimated gaze onto the external environment.

2. Driver State Estimation and Visualization of Gaze Targets

A driver-specific face model, the true-scale three-dimensional positions of 68 facial landmarks in a head coordinate system defined by the eyes and the nose tip, is built in advance from the driving video. The driver is recorded by two cameras with pre-calibrated relative pose; the landmarks are detected in both images by 3DDFA-V3⁽¹⁾ and their positions obtained by triangulation. As frontal frames dominate during driving, the head yaw and pitch angles are divided at 10° intervals and frames with a small reprojection error, the discrepancy between the detected landmarks and the triangulated points reprojected onto both images, are selected from each interval so that face orientations are not biased; the landmarks are then transformed into the head coordinate system and their median over these frames gives the model. In each frame, the head pose (rotation and translation) minimizing the reprojection error of this model against the landmarks of both cameras is obtained; yaw and pitch are defined relative to the median forward direction over the section, and the translational component gives the forward displacement from the normal head position, an index of forward leaning. Gaze is estimated with UniGaze⁽²⁾ for each camera, and the two estimates are integrated by a weighted average based on the frontality of the head.

The gaze origin is the centroid of six landmarks around the eyes and the nose. After transformation into the coordinate system of the forward-looking stereo camera, the points of the three-dimensional point cloud that lie inside a cone around the gaze, whose spread is set from the gaze estimation error and the effective visual field, are extracted, and the nearest one is taken as the gaze point. It is associated with regions detected by YOLO11⁽³⁾; curve mirrors and road signs are handled by detectors built through transfer learning, and regions without clear boundaries are identified by semantic segmentation. The gaze target and the timing and duration of gaze are thus quantified without contact.

3. Driving Behavior Analysis

The method was applied to video of a vehicle passing straight through an unsignalized intersection with a stop regulation, where hedges and walls severely occlude the crossing road. The data were obtained in an experiment approved by the research ethics committee of the University of Tokyo, after the participant had been informed of its content and had given consent. Fig. 1 shows the gaze target, the head and gaze yaw angles, and the forward leaning distance; shading indicates the stop period.

Forward leaning begins at 10.8 s and reaches 11.4 cm at 11.4 s: the driver first checks the left blind spot through a curve mirror, and then leans forward to look at it directly. The leaning is maintained after the start of motion while the gaze target is switched among the curve mirror and both sides, and is released after 15.7 s, when the blind spot disappears. The gaze returns to the front at 15.8 s and the head pose at 16.3 s, which reveals a compound confirmation behavior in which the gaze precedes the head.

4. Conclusions

The proposed method estimates the head pose, gaze and posture of a driver from in-vehicle camera images and identifies gaze targets by projecting the estimated gaze onto the external environment. Evaluation with more participants and projection of gaze points outside the camera field of view by reconstructing the surroundings from adjacent frames remain as future work.

Acknowledgment: This work was supported by JSPS KAKENHI Grant Number 23H00214.

References

- (1) Z. Wang, et al.: 3D face reconstruction with the geometric guidance of facial part segmentation, Proc. IEEE/CVF CVPR, 1672-1682, 2024.
- (2) J. Qin, et al.: UniGaze: Towards universal gaze estimation via large-scale pre-training, Proc. IEEE/CVF WACV, 2026.
- (3) R. Khanam, et al.: YOLOv11: an overview of the key architectural enhancements, arXiv:2410.17725, 2024.

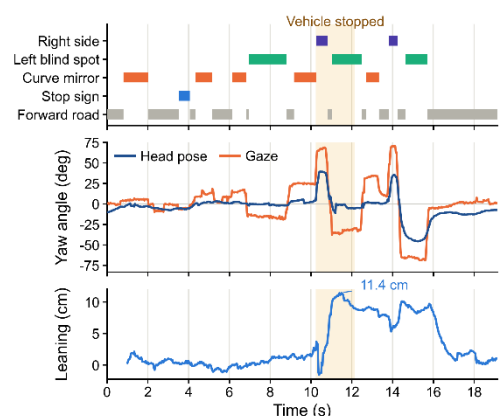


Fig. 1. Gaze target, head/gaze yaw angle, and forward leaning distance while passing through an intersection with poor visibility.