Iowa Initiative for Artificial Intelligence

Final Report

Project title:	Advanced Eye Tracking of Moving Objects in Driving Simulation
Principal Investigator:	Elizabeth O'Neal, Michelle Reyes, & Jodie Plumert
Prepared by (IIAI):	Yanan Liu
Other investigators:	
Date:	
Were specific aims fulfilled:	In part
Readiness for extramural proposal?	N
If yes … Planned submission date	
Funding agency	
Grant mechanism	
If no … Why not? What went wrong?	Due to small training sample, YOLOv3 model is not perfect. Labeling more videos needs additional time.

<u>**Brief summary of accomplished results:**</u>

We developed and validated a YOLO model to detect traffic signs for driving video. We successfully determined whether eye marker overlapped with AOI or not and create a pipeline to calculate eye gaze duration based on that.

<u>**Research report:**</u>

Aims (provided by PI):

Aim 1: Develop a generalizable machine learning algorithm to identify objects or areas of interest (AOIs) at specified timesteps as AOIs move and the distance between the AOI and viewer closes within a simulated driving environment.

Aim 2: Derive measures of eye gaze using the location of AOIs identified under Aim 1 and pupil coordinates.

Data:

Researchers at the National Advanced Driving Simulator currently have eye-tracking data (gathered using Dikablis Eye Tracking Glasses Professional headset and Ergoneers D-Lab software) for 320 hazard-laden simulator drives. Drives last approximately 20 minutes each. In additional to periods of normal driving, each drive contains 19 hazard events.

Input Measures

Driving scenario videos. For each driving session, a forward-facing camera captured video of the driving scenario. Note that the forward-facing camera apparatus was head mounted. Therefore, video capture is not static; it follows the head movements of the subject. Video data is exported as an .mp4 file. Video files can be exported to show only the forward view or to show the forward view with gaze overlay (Figure 1).

Hazard events. Hazard events may contain multiple AOIs, such as moving cars, brake lights, traffic lights, and road signs. Task triggers, created using the eye-tracking software, recorded the beginning and end of each hazard event. Detailed information about each simulated hazard (i.e., overall description of the hazard and relevant AOIs, AOI position within the simulated environment, precise measurements of AOI movement in distance and time, distance between driver and AOI at specific timesteps), and the relevant areas of interest will be provided. Timestamp information regarding the start, critical periods, and end of each hazard can be exported as a .csv or .txt file.

Pupil/eye glance data. The software determines where subjects are looking in the forward scene by recording the x- and y-coordinates of the eyes at each timestep (~.17 ms). Coordinates are available for the left and right eyes, either individually or combined. Eye coordinate data can be exported as a .csv to .txt file. It may be useful to use markers embedded into the scene (Figure 1) as a constant in the video coordinate space due to the nature of the video capture.

Output Measures

Areas/objects of Interest. The location (in x- and y-coordinates or other mapping values) of the AOIs at each timestep (~.17 ms).

Glance at AOI. A count variable of the number of glances to AOIs for each hazard event. Glances are defined as any look to the AOI that exceeds 120ms.

Glance duration. The duration (in ms) of each glance that exceeds 120ms to the AOIs.

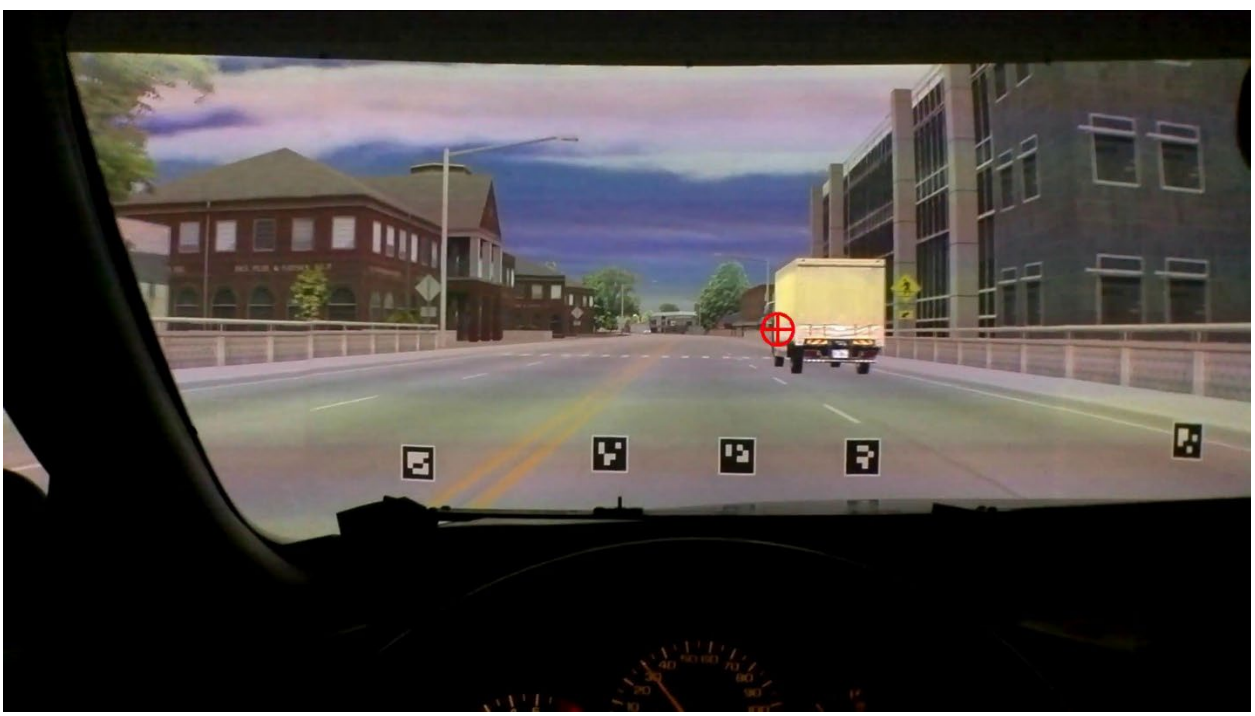

Figure 1. Screenshot of video, gaze overlay, and fixed markers in DLab. Note: Markers (i.e., black and white square images embedded in the scene) stay in a constant position in the driving scene.

AI/ML Approach:

YOLO was used to detect objects (traffic signs) from the frames extracted from driving video. YOLO model was implemented by using ImageAI (https://github.com/OlafenwaMoses/ImageAI). We trained YOLOv3 model for performing detection of traffic signs using our labeled frames. We extracted frames from driving videos and labeled them. Total training sample is 1390. Loss is 7.26.

Results:

Figure 2 shows eye gaze area in red circle, traffic sign detection in white bounding box. They are overlapping with each other. By calculating the Intersection over Union (IoU) of two bounding boxes, we can calculate the duration of eye gaze.

Figure 2. Screenshot of video, object detection (white rectangle), eye gaze (red circle).

From Figure 2, we could calculate eye gaze duration based on IOU of object detection in each frame. For the test video, it contains 1164 frames with 30 fps. YOLO model detects 222 frame which contains ROI (traffic sign). Only 23 frame overlays with eye gaze circle. Based on IOU, 14 frames are the same object. The eye gaze duration for this object is ~0.46 second (1/30 second X 14 frame).

Due to small training sample and only 1 object (traffic sign), YOLO model is not perfect (miss some of the object in frame). More training sample requires a lot of manual labels and takes a lot of time.

=================