Prediction of effort and eye movement measures from driving scene components

Cabrall, Christopher; Happee, Riender; de Winter, Joost

DOI
10.1016/j.trf.2019.11.001

Publication date
2020

Document Version
Final published version

Published in
Transportation Research. Part F: Traffic Psychology and Behaviour

Transportation Research Part F 68 (2020) 187–197

Contents lists available at ScienceDirect

Transportation Research Part F

journal homepage: www.elsevier.com/locate/trf

Prediction of effort and eye movement measures from driving scene components

Christopher D.D. Cabrall *, Riender Happee, Joost C.F. de Winter

Department of Cognitive Robotics, Delft University of Technology, Mekelweg 2, 2628 CD Delft, the Netherlands

ARTICLE INFO

Article history:
Received 14 February 2018
Received in revised form 1 August 2019
Accepted 3 November 2019
Available online 27 November 2019

Keywords:
Driving scenes
Eye-tracking
Workload
Driver assistance
Automated driving
Individual differences

ABSTRACT

For transitions of control in automated vehicles, driver monitoring systems (DMS) may need to discern task difficulty and driver preparedness. Such DMS require models that relate driving scene components, driver effort, and eye measurements. Across two sessions, 15 participants enacted receiving control within 60 randomly ordered dashcam videos (3-second duration) with variations in visible scene components: road curve angle, road surface area, road users, symbols, infrastructure, and vegetation/trees while their eyes were measured for pupil diameter, fixation duration, and saccade amplitude. The subjective measure of effort and the objective measure of saccade amplitude evidenced the highest correlations (r = 0.34 and r = 0.42, respectively) with the scene component of road curve angle. In person-specific regression analyses combining all visual scene components as predictors, average predictive correlations ranged between 0.49 and 0.58 for subjective effort and between 0.36 and 0.49 for saccade amplitude, depending on cross-validation techniques of generalization and repetition. In conclusion, the present regression equations establish quantifiable relations between visible driving scene components with both subjective effort and objective eye movement measures. In future DMS, such knowledge can help inform road-facing and driver-facing cameras to jointly establish the readiness of would-be drivers ahead of receiving control.

1. Introduction

1.1. The potential of eye-based driver monitoring systems

With the advent of automated driving systems, drivers will be removed from sustained involvement in the driving task until being called back in. Consequently, a large body of research has aimed to establish time budget requirements for reaching situation awareness (SA) before the driver is given back control (e.g., Lu, Coster, & De Winter, 2016).

A major disadvantage of current SA measurement methods, such as probe-recall techniques, is that they are disruptive and impractical in real-time applications. Instead of traditional accounts of SA, a directly measurable driver index should be useful for driver monitoring systems (DMS) in transitions of control. Eye-based DMS in automated vehicles have the potential to verify driver readiness ahead of returning control to the human.

Victor (2005) explained that: '*the visual information you explicitly experience is not the same as the visual information that guides your actions . . . Information enters the nervous system and influences action even when it doesn't gain conscious attention*'.

* Corresponding author.
E-mail addresses: C.D.D.Cabrall@tudelft.nl (C.D.D. Cabrall), R.Happee@tudelft.nl (R. Happee), J.C.F.deWinter@tudelft.nl (J.C.F. de Winter).

https://doi.org/10.1016/j.trf.2019.11.001
1369-8478/

Neurological pathways identified by Ungerleider and Mishkin (1982) differentiate between a '*vision-for-action*' sensorimotor process that runs in parallel and fast and a more conscious '*vision-for-identification*' process associated with traditional SA measurement techniques. Moreover, De Winter, Eisma, Cabrall, Hancock, and Stanton (2019) propose a theoretical account of SA that is based on eye movements in relation to the task environment to overcome task interruption of more common SA measurement techniques and thus is more amenable to real-time applications. The contribution of our present approach is to inform an eye-based DMS verification function.

During the automation-to-manual transition phase, we assume the eyes to undertake preparatory activities that can be measured in the laboratory and later checked against measurements on the roads. Our investigation aims to extend the present body of knowledge concerning take-over requests with an elaboration on eye movements between the moment the driver first focuses on the road and the moment the driver resumes manual control. We are concerned with measuring '*vision-for-action*', which is presumed to be available at an earlier point than the '*vision-for-identification*' concepts of conventional SA measurement methods.

1.2. Taking into account the visual demands of the driving scene

The present investigation proposes to capture real-world driving scene variation on a continuous scale. Around the world, affordable dashboard cameras and video sharing sites have enabled access to a large body of realistic driving scenes. Moreover, both crowdsourced (e.g., Cabrall et al., 2018) and semi-automatic annotation (e.g., Jin, Li, Ma, Guo, & Yu, 2017) processes are now available to rapidly perform classifications and thus allow for driving scene libraries of desired contents. Lastly, advances in computer vision enable automatic processing of driving scenes down to the point where scenes can be encoded at a pixel level (e.g., Yu, Xian, Chen, Liu, Liao, Madhavan, & Darrell, 2018). Thus, it is now conceivable to determine a total visible scene percentage per object of interest (e.g., cars, pedestrians, signs/symbols, buildings). Such techniques enable controlled laboratory studies to answer questions such as which scene contents (and in which proportions) might be expected to affect eye movements and perceived driver effort.

Unlike humans who can easily intuit differences between easy and demanding driving scenes, a DMS will require a priori programming that ensures that the recorded eye movements are assessed relative to the visual demands of the driving scene. Previous research about the association between driving scene characteristics and driver workload has relied on variables that are nominal or ordinal in scale. For example, De Waard (1996) evidenced increased driver workload for sections of motorways with entrances/exits over those without, for sections with adjacent noise barriers compared over those without, and for areas of rural vegetation over those of monotonous moorland. Furthermore, Foy and Chapman (2018) evidenced high to low mental demand responses across the following virtual road types : suburban roads, city center, arterial A-roads, and dual carriageway. In lessons learned from developing driving research scenes and scenarios, Papelis, Ahmad, and Watson (2003) argued that '*Often times, specifications about the characteristics of the ambient traffic or ambient environment are missing or incomplete'* and *'it is often the case that variations in these ambient characteristics of a scenario can make a drastic difference on how participants perceive the scenario*'.

More research is needed concerning how eye measures can be predicted from driving scene determinants. Previous eye measurement outcomes of driving scene components stand to be improved upon and extended by continuous scale measurements. For example, drivers' eyes have been found to frequently change their points of reference with different sized cars and turn radiuses (Olson, 1964), in the presence vs. the absence of a lead vehicle (Mourant, Rockwell, & Rackoff, 1969), in relation to signs, other vehicles, and road edge markings (Mourant & Rockwell, 1970), while approaching and transiting curves (Land & Lee, 1994; Laya, 1992), and with varying levels of experience across different types of roads: rural, suburban, and expressway (Chapman & Underwood, 1998).

1.3. Study aims

The present study aims to relate a continuous quantification of driving scene contents (i.e., available from driving scene semantic segmentation) to common eye measures (i.e., pupil diameter, fixation duration, saccade amplitude) and driver effort ratings (i.e., driver visual workload). The present method was designed to support applications in real-time transitions of control from automated driving to human driving via DMS verification of (in)adequacy of situated visual behavior.

2. Methods

We investigated the effects of visible driving scene characteristics on human perceived effort ratings and eye-tracking measures.

2.1. Participants and apparatus

Written informed consent was obtained from all participants, and the research was approved by the Human Research Ethics Committee of the Delft University of Technology under the title 'Driving video ratings' (16 December 2015). The

experiment was completed by 15 participants (six female, nine male) aged between 18 and 36 (M = 26.60, SD = 4.26) with an average driving experience of around seven years since obtaining the driver's license (M = 7.20, SD = 4.20).

The experiment apparatus consisted of a stimulus display monitor, SR Research Eyelink 1000 Plus eye-tracking camera with integrated IR source and dedicated head/chin rest mount, as well as a gaming steering wheel (Fig. 1). The display was a 24-in. (diagonal) BenQ XL2420T-B monitor with a resolution of 1920 × 1080 pixels and a display area of 531 × 298 mm. The display was positioned about 95 cm in front of the participant and about 35 cm behind the eye-tracking camera/IR source. The boundaries of the stimulus display area subtended approximately 31/18 degrees of horizontal/vertical viewing angle per the setup ranges required by guidelines of the eye-tracker. Eye behavior data were recorded after individual participant calibration. The eye event parser was set according to the default psychophysical configuration: saccade velocity threshold of 22 deg/s, saccade acceleration threshold of 3800 deg/s^2, and saccade motion threshold of 0 deg. The steering wheel was an unconnected Logitech G27 and along with an isolating partition was used to facilitate driving video stimulus immersion. A mouse was used for effort ratings.

2.2. Procedure

The height of the head/chin rest mount was adjusted to each participant. Participants kept their heads stationary within the mount except for voluntary rest breaks around every five minutes across about 15 min of driving video viewing and rating trials. Each trial began with a drift correction dot in the center of the screen to which participants needed to fixate and click the mouse at the same time. A 3-second long driving video clip was then played during which participants were tasked to move their hands to the wheel while imagining that they were taking over control (i.e., from automated driving) and that they must drive within that scene.

2.3. Stimuli and measurements

Stimuli consisted of 60 randomly ordered 3-second long dashcam clips (Fig. 2). The clips were shown in two rounds of 60 exposures each separated by a break of about 5 min. To incorporate a variety of different driving scene circumstances, we selected the video clips from different complexity levels, extended from a driving scene library presented in Cabrall et al. (2018). Reasonably, DMS will need to assess drivers at a high frequency, on the order of seconds rather than minutes. Short-lasting clips were used as a tradeoff of informative eye data (i.e., an average fixation duration is around 0.3 seconds, thus in 3 seconds about 10 fixations might occur) while constraining the scene to be fairly similar (e.g., curves might not last more than a few seconds, so longer durations of measurement would need to average large changes of scene components). Furthermore, as a simplified approach, only a representative frame (i.e., the mid-point of the 3 seconds) was used to characterize objective quantifications of driving scene contents.

Driving scene content components were manually outlined and color-coded (cf. CityScapes Dataset, 2018) into five separate categories:

(1) road surface,
(2) road users (vehicles, bicycles, pedestrians),
(3) signs (stop signs, crosswalks, roadway writing, billboard advertisements, etc.),

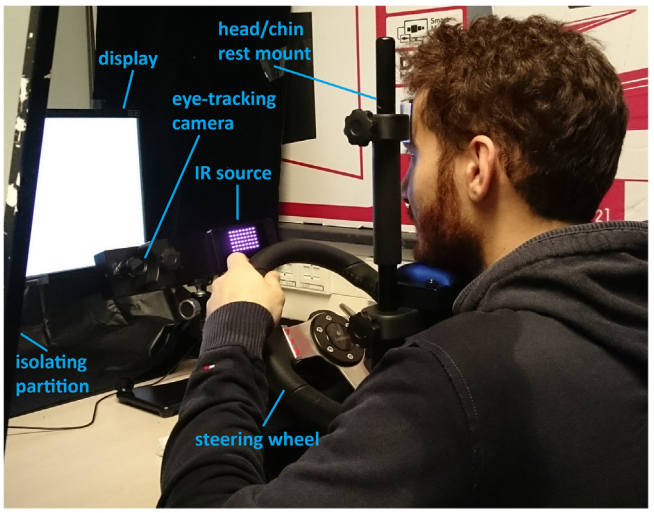

Fig. 1. Experiment apparatus.

Video 23.avi

Visible Driving Scene Context Contents

Road Area	10.71%
Road Users	0.00%
Signs/Symbols	0.00%
Buildings/Infrastructure	0.35%
Vegetation/Trees	23.02%
Approx. Road Curvature	2°

Driver Visual Workload Responses

	p01	p02	...	p15
Effort	8	4	...	0

Driver Eye Measures

	p01	p02	...	p15
Pup. Size	4320.7	3444.6	...	4379.3
Fix .Dur.	871.2	490.3	...	490.9
Sac. Amp.	0.7	0.9	...	1.1

Video 47.avi

Visible Driving Scene Context Contents

Road Area	9.65%
Road Users	7.93%
Signs/Symbols	1.93%
Buildings/Infrastructure	10.99%
Vegetation/Trees	46.07%
Approx. Road Curvature	61°

Driver Visual Workload Responses

	p01	p02	...	p15
Effort	15	10	...	25

Driver Eye Measures

	p01	p02	...	p15
Pup. Size	4466.9	3555.1	...	4558.1
Fix .Dur.	872.8	250.4	...	321.3
Sac. Amp.	0.7	2.6	...	3.3

Video 56.avi

Visible Driving Scene Context Contents

Road Area	6.75%
Road Users	7.13%
Signs/Symbols	1.42%
Buildings/Infrastructure	31.97%
Vegetation/Trees	13.97%
Approx. Road Curvature	78°

Driver Visual Workload Responses

	p01	p02	...	p15
Effort	60	19	...	25

Driver Eye Measures

	p01	p02	...	p15
Pup. Size	4536.2	3498.6	...	4576.6
Fix .Dur.	483.3	477.3	...	291.6
Sac. Amp.	1.9	2.9	...	3.9

Fig. 2. Examples of continuously scaled quantities of scene contents (road area, road users, signs/symbols, buildings/infrastructure, vegetation/trees, road curve angle) along with participants' (i.e., p01, p02 ... p15) measures of effort ratings and eye measures (pupil diameter, fixation duration, saccadic amplitude).

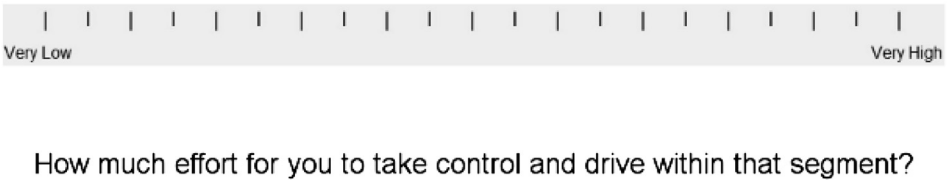

Fig. 3. Driving effort response scale.

(4) buildings (houses, light poles, fences, etc.), and
(5) vegetation (trees, bushes, hedges, etc.).

A free online image processing tool (Krzywinski, 2018) was used to determine a percentage of the windshield view that a specific category covered in terms of pixelated area. A transparent protractor overlay was used to manually approximate the curve angle of the road for a representative frame in the 3-second video clip (i.e., the middle frame). First, a mid-lane position point was picked where the road "disappeared" beneath the vehicle. Next, a furthest visible point in the middle of that same lane of travel along the horizon was picked and a line drawn to connect these points. Lastly, this line was measured in approximated angular difference (degrees) from a straight-ahead line.

The eye-tracking camera recorded the eyes of the participant and delivered measurement outputs as averages across the 3-second duration exposure scene viewings: average pupil diameter (arbitrary units), average fixation duration (ms), and average saccade amplitude (degrees). After the clip finished and disappeared, an effort rating response scale and prompt ("How much effort for you to take control and drive within that segment?") was presented on the upper half of the screen, and participants moved a vertical cursor to click on the scale to input their answer from between "Very Low" to "Very High". Cursor click horizontal positions were divided by the pixel length of the scale and rounded to a single point resolution from 0 to 100. The presented horizontal effort scale contained 21 equally spaced demarcations from left to right following from those described within the seminal NASA-TLX (Task Load Index) subscales (Hart & Staveland, 1988) (see Fig. 3).

3. Results

Data were collected in two rounds. Round 1 data (eye-tracking and effort ratings) were collected during the first exposure period of the 60 randomly ordered driving video clips. Round 2 concerned a repetition of the video clips in a new randomized order after a break of about 5 min.

3.1. Overview of data from first exposures

All continuously scaled quantifications of driving visual scene contents from each video clip are depicted in Fig. 4 as ranked by Round 1 effort rating responses, averaged across the 15 participants (M = 24.40%, SD = 11.10%) along with averages for pupil diameter in arbitrary units (M = 4129.81, SD = 60.42), fixation duration (M = 433.99 ms, SD = 74.27 ms), and saccade amplitude (M = 1.79 deg, SD = 0.67 deg). The standard deviations at the level of 15 participants after averaging the 60 videos were 14.40% for effort, 430.57 for pupil diameter, 126.05 ms for fixation duration, and 0.33 deg for saccade amplitude. At the level of individual videos separately, the standard deviations were 21.54% for effort (n = 900), 444.06 for pupil diameter (n = 900), 218.20 ms (n = 899) for fixation duration, and 1.06 deg (n = 898) for saccade amplitude.

Bivariate correlations between driving scene content variables were computed to examine the presence of multicollinearity. All correlations between predictors were found to be well below a conventionally considered threshold of r = 0.80 (Table 1, top).

Correlations between predictor driving scene variables and the dependent variables of effort ratings and eye measures are given in Table 1 (bottom). For effort ratings, road curve angle, signs, road users, and buildings evidenced moderate positive correlations (r = 0.25 to 0.34). For the pupil diameter measure, correlations were all near zero. For fixation duration, road surface area and vegetation showed a positive correlation, while road curve angle, signs, and road users showed negative correlations; however, associative strengths here were all weak. For saccade amplitude, road curve angle, road users, and signs evidenced moderate-to-strong positive correlations (r = 0.36–0.42) while buildings showed a moderate positive correlation (r = 0.26) and vegetation a weak negative correlation (r = −0.11).

3.2. Linear regression analyses

While Section 3.1 allowed for considering the scene components in an isolated fashion, the present section reports on the scene components taken together. Linear regression analyses were performed to investigate relationships between the scene components (independent variables) and dependent variables of effort ratings or driver eye measures.

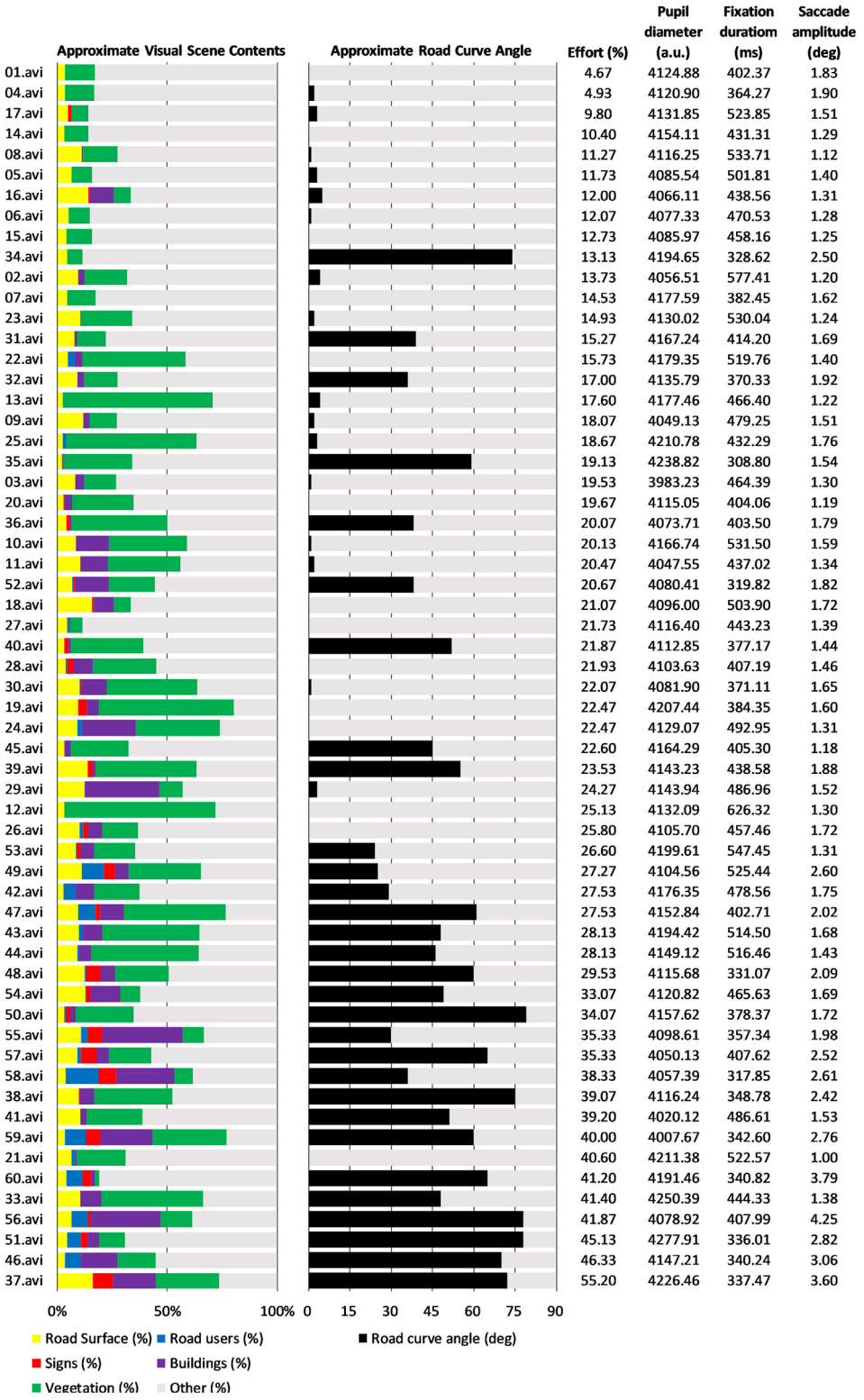

Fig. 4. Driving scene visual characteristics per video clip with averages of effort ratings and eye measures (Round 1).

Table 1
Correlations between driving scene contents, effort ratings, and eye measures (n = 900, 15 participants × 60 videos) (Round 1).

	Road curve angle	Road users	Buildings	Road surface	Signs	Vegetation/trees
Road curve angle (deg)	–	–	–	–	–	–
Road users (%)	0.35	–	–	–	–	–
Buildings (%)	0.20	0.45	–	–	–	–
Road surface (%)	0.02	−0.18	0.31	–	–	–
Signs (%)	0.41	0.51	0.41	0.20	–	–
Vegetation/trees (%)	−0.01	−0.08	−0.09	−0.06	−0.07	–
Effort (%)	0.34	0.26	0.25	0.09	0.28	0.04
Pupil diameter (a.u.)	0.03	−0.01	−0.02	−0.02	−0.01	0.03
Fixation duration (ms)	−0.19	−0.09	−0.07	0.07	−0.14	0.06
Saccade amplitude (deg)	0.42	0.38	0.26	0.02	0.36	−0.11

Note. n = 899 for fixation duration, n = 898 for saccade amplitude.

These regression models were taken first at a general level across all participants and subsequently at an individual level. In all cases, a conventional split off of 66.7% of the data (i.e., data for a random 40 of 60 videos) was used to fit a model on a majority of the data. Next, three separate validation assessments (VA) were performed for ascertaining both repetition (consistency) and generalization (cross-validation) performance. In each case, correlations were calculated between predicted and observed values in the first and second rounds of video data.

(1) VA-1) Model computed with data for 2/3 of Round 1 video clips was used to predict data for the same 2/3 of video clips in Round 2
(2) VA-2) Model computed with data for 2/3 of Round 1 video clips was used to predict data for the remaining 1/3 of video clips in Round 1
(3) VA-3) Model computed with data for 2/3 of Round 1 video clips was used to predict data for the the remaining 1/3 of video clips in Round 2

3.2.1. General level modeling across all participants and validation assessments

The regression models for the randomly selected 2/3 of Round 1 per dependent variable are presented in Tables 2–5. Coefficients of correlation between the predictor variable set of scene components were observed to be moderate to strong (r = 0.46), weak/none (r = 0.06), moderate (r = 0.26), and strong (r = 0.54), respectively for effort ratings, pupil diameter, fixation duration, and saccade amplitude.

Validation assessments of the regression models exhibited moderate repetition and weak generalization capabilities (Table 6). In terms of prediction performance with repetition of the same video clips in the test model as used in the training of the model, effort ratings (r = 0.48) and saccadic amplitude (r = 0.47) had the highest correlative associations.

Table 2
Model summary statistics for prediction of self-reported driver effort (%) from driving scene contents.

Predictor	*B*	*β*	*t*	*p*
Constant	7.258			
Road curve angle (deg)	0.229	0.309	7.626	<0.001
Road users (%)	80.211	0.126	1.944	0.052
Signs (%)	56.200	0.052	0.945	0.345
Buildings (%)	23.576	0.103	1.849	0.065
Road Surface (%)	37.152	0.054	1.252	0.211
Vegetation/trees (%)	13.141	0.108	2.907	0.004

$F(6,593) = 27.03$, $p < 0.001$, $r = 0.46$, $r^2 = 0.21$.

Table 3
Model summary statistics for prediction of mean pupil diameter (arbitrary units) from driving scene contents.

Predictor	*B*	*β*	*t*	*p*
Constant	4128.378			
Road curve angle (deg)	0.559	0.036	0.799	0.425
Road users (%)	−62.268	−0.005	−0.065	0.948
Signs (%)	−224.933	−0.010	−0.162	0.871
Buildings (%)	−168.657	−0.035	−0.568	0.571
Road Surface (%)	−116.379	−0.008	−0.168	0.866
Vegetation/trees (%)	58.160	0.023	0.552	0.581

$F(6,593) = 0.32$, $p = 0.928$, $r = 0.06$, $r^2 = 0.00$.

Table 4
Model summary statistics for prediction of mean fixation duration (ms) from driving scene contents.

Predictor	*B*	*β*	*t*	*p*
Constant	412.582			
Road curve angle (deg)	−1.266	−0.171	−3.874	<0.001
Road users (%)	396.261	−0.062	0.883	0.378
Signs (%)	−1303.525	−0.120	−2.015	0.044
Buildings (%)	−78.374	−0.034	−0.565	0.572
Road Surface (%)	638.426	0.093	1.978	0.048
Vegetation/trees (%)	100.016	0.082	2.034	0.042

$F(6,593) = 7.17$, $p < 0.001$, $r = 0.26$, $r^2 = 0.07$.

Table 5
Model summary statistics for prediction of mean saccade amplitude (deg) from driving scene contents.

Predictor	*B*	*β*	*t*	*p*
Constant	1.741			
Road curve angle (deg)	0.011	0.288	7.461	<0.001
Road users (%)	8.221	0.254	4.120	<0.001
Signs (%)	−4.710	−0.085	−1.638	0.102
Buildings (%)	1.531	0.131	2.484	0.013
Road Surface (%)	−3.363	−0.096	−2.342	0.019
Vegetation/trees (%)	−0.827	−0.134	−3.781	<0.001

$F(6,592) = 39.90$, $p < 0.001$, $r = 0.54$, $r^2 = 0.29$.

In terms of prediction performance with new video clips yet-unseen by the model, correlation coefficients between predicted and observed values for effort ratings, fixation duration, and saccade amplitudes on average between the two generalization assessments were typically weak to moderate (i.e., r between 0.04 and 0.30) while pupil diameter remained weak (i.e., r = 0.03 and 0.05).

3.2.2. Individual level modeling and validation assessments

Table 7 shows that person-specific regression models exhibited 'moderate' to 'strong' repetition and generalization capabilities. In terms of prediction performance with repetition of the same video clips as used in the training of the model, effort ratings (Mean r = 0.58) and saccadic amplitude (Mean r = 0.49) had the highest correlations between what was predicted and what was observed.

In terms of prediction performance with new video clips yet-unseen by the model, correlation coefficients between predicted and observed values for effort ratings were of moderate to high strength (Mean r = 0.49, 0.55) and for saccade amplitude as well (Mean r = 0.36, 0.43). Comparatively, prediction performances for fixation duration and pupil diameter were weaker (i.e., between 0.07 and 0.22).

Table 6
Predictive correlation coefficients between observed scores and those predicted by a regression analysis with scene components as predictor variables.

VA-1: Repetition	Dependent variable	*r*	Mean absolute error
40 random video clips in Round 1 (600 data points), repeated as the 40 same video clips in Round 2 (600 data points)	Effort	0.48	15.85 (%)
	Pupil diameter	0.08	338.55 (a.u.)
	Fixation duration	0.17	152.17 (ms)
	Saccade amplitude	0.47	0.74 (deg)
VA-2: Generalization	Dependent variable	*r*	Mean absolute error
40 random video clips in Round 1 (600 data points), generalized to 20 new video clips in Round 1 (300 data points)	Effort	0.30	15.84 (%)
	Pupil diameter	0.03	327.56 (a.u.)
	Fixation duration	0.18	155.66 (ms)
	Saccade amplitude	0.14	0.74 (deg)
VA-3: Generalization	Dependent variable	*r*	Mean absolute error
40 random video clips in Round 1 (600 data points), generalized to 20 new video clips in Round 2 (300 data points)	Effort	0.25	16.67 (%)
	Pupil diameter	0.05	331.53 (a.u.)
	Fixation duration	0.20	151.72 (ms)
	Saccade amplitude	0.04	0.81 (deg)

Table 7
Predictive correlation coefficients between observed scores and those predicted by person-specific regression analyses with scene components as predictor variables.

VA-1: Repetition	Dependent variable	Mean *r*	*N r*	*SD r*	Mean absolute error
40 random video clips in Round 1, repeated as the 40 same video clips in Round 2	Effort	0.58	15	0.18	10.79 (%)
	Pupil diameter	0.20	15	0.19	145.61 (a.u.)
	Fixation duration	0.15	15	0.21	153.60 (ms)
	Saccade amplitude	0.49	15	0.20	0.69 (deg)
VA-2: Generalization	Dependent variable	Mean *r*	*N r*	*SD r*	Mean absolute error
40 random video clips in Round 1, generalized to 20 new video clips in Round 1	Effort	0.49	15	0.27	10.58 (%)
	Pupil diameter	0.07	15	0.29	125.87 (a.u.)
	Fixation duration	0.15	15	0.27	131.58 (ms)
	Saccade amplitude	0.36	15	0.25	0.73 (deg)
VA-3: Generalization	Dependent variable	Mean *r*	*N r*	*SD r*	Mean absolute error
40 random video clips in Round 1, generalized to 20 new video clips in Round 2	Effort	0.55	15	0.19	11.69 (%)
	Pupil diameter	0.12	15	0.23	155.99 (a.u.)
	Fixation duration	0.22	15	0.20	156.19 (ms)
	Saccade amplitude	0.43	15	0.25	0.71 (deg)

4. Discussion

4.1. Using driving scenes to predict effort ratings

Our experiment exposed participants to a range of driving scene contents, and a range of subjective effort ratings were captured regarding imagined reception of driving control within those driving scenes. With video clips as short as only a few seconds, visible driving scene contents were evidenced to be predictive of subjective effort. As a starting place, our driving scenes featured no emergency conditions. Driving scenes of greater urgency could also be worthwhile for further investigations as suitable to different DMS design assumptions.

Not all visual information in the driving scenes was found to be well correlated with the effort ratings. For example, the amount of vegetation/trees and road surface were weakly associated with effort. So, expecting effort to rise just on account of having more to look at would appear to be an oversimplification. The driving scene features and objects that were found to be correlated with effort rating response appear to be those that are semantically meaningful to the task of receiving driving control. Participant effort ratings appeared to be most closely associated with road curve angle (i.e., an aspect of lateral control).

Within the regression model, at the general level across all participants, effort ratings were explainable from the driving scene contents taken together with a moderately strong correlation ($r = 0.46$). In validation assessments (repetition and generalization), the average observed error for this general level model was around 16 points on the effort scale and dropped to around 11 points from individual-level models. An estimation error of 11% for driver effort sounds sufficient for practical benefit in automated real-time DMS judgments in transitions of control. However, future simulator or on-road studies with prolonged active driving should confirm such assumptions. Regardless, an apparent increase in strength of correlation was evidenced by moving from a general level model prediction of visual workload to individually tailored ones.

4.2. Using driving scenes to predict driver eye behavior

Saccade amplitude appears to be a more relevant eye measure compared to pupil diameter (generally lacking relations) and fixation duration (generally weaker relations) in terms of association with driving scene contents. Vegetation/trees actually produced a negative correlation with saccadic amplitude. Although vegetation/trees can contain varying amounts of visual information, such features are likely interpretable as background with respect to the driving task of taking over control. Greater amounts of vegeration reduce the probability of the presence of other road users and the visual scanning required to detect and track their movements. Predictions of pupil diameter were not statistically distinguishable from non-influence of driving scene components while fixation duration and saccadic amplitude were. Nearly 30% of the variance in saccade amplitude was explainable from the full set of continuously scaled driving scene contents. In validation assessments, around 0.75 degrees of error in saccadic amplitude was observed. Future simulation studies need to validate whether such levels of error might be viable in DMS transitions-of-control applications.

4.3. Potential applications and example use case

The most relevant application areas for the findings of the present investigation and its produced regression model equations are envisioned to support further research and development of real-time DMS applications for the transitional phases between automated and human driving control. Ahead of being given full control of a vehicle, the eyes of the would-be driver

can be compared against stored predicted values based on the present driving scene the contents of which can be quantified on a continuous scale resolution. By periodically viewing various video clips of real-world driving scenes and responding to prompts of subjective effort ratings (i.e., while not driving), automated eye-tracking enabled DMS might pre-train and fit its expectations for specific owner/user behavior towards improved assessment operations. When using only the individual driver as ground truth, such modeling would only be expected to be as good as the driver him/herself and thus vulnerable to maladaptive eye movements, particularly if the driver is unaware of such vulnerabilities (see Kruger & Dunning, 1999). Because standard driver education programs do not appear to adequately address emergent hazards (Fisher, Pollastek, & Pradhan, 2006), supplemental driving video clips with data beyond the individual driver may prove useful for special populations such as learner drivers.

In a hand-over of driving control situation, as the person him/herself begins to assess the driving scene and ascertain task demands, allocate attentional resources, and ultimately regain situation awareness, an automatic DMS safety layer can provide oversight and correct as needed. If the driving scene is one where moderate or high amounts of driving visual workload is expected (e.g., as inferred from a sharp curve, as well as other road users and many signs and symbols to read and interpret, etc.) but the driver's eyes are moving with shorter distances than has been previously computationally predicted (e.g., they are still mentally fixated on that last email they were composing), then any number of different adjustments might be made in terms of automated warnings and/or vehicular control to modulate the potential risks of the transition. One example interface solution might be to begin to highlight relevant missed parts of the driving scene until the driver can unlock full manual control by gazing at these, but of course, there are many alternative design solutions. In any case, a real-time assessment of driver fitness to drive within the present scene through directly observable constructs (i.e., vision-for-action ahead of vision-for-identification aspects typically construed by information processing models of situation awareness) should be desirable and the present study represents a starting point method and resulting models for generating such information.

4.4. Limitations

There are important considerations in common across our measures that should be taken into account. Due to difficulties in reliable manual human annotation, a potentially confounding effect of ego-vehicle speed in the driving video segments was not yet controlled, and we recommend such an aspect as an interesting independent factor to investigate in future studies. Eye measurements were taken while viewing a previously filmed driving video rather than in a full-fidelity environment where additional fields of view might be expected to be present and relevant (e.g., mirrors and periphery). Eye-tracking and scene identification will be limited by the cost and availability of technological software and hardware components (e.g., computer vision and machine learning) although these have been recently undergoing rapid advancements. For example, our first approach of manual annotations of scene contents limited us to quantifying only a single representative frame of each 3-second clip, whereas, in the future, more automatic processes could extend such objective quantification of scene characteristics. As with all models, more data is expected to improve the presently provided regression equations. The current framework could be extended via additional videos of greater variety and increased sets of classifiable items.

5. Conclusions

In conclusion, the present study has established correlations between driving scene components and subjective driver effort as well as objective eye measures of fixation duration and saccade amplitude. In particular, the degree of road curve angle plays a strong role in human judgments of driving scene difficulty and has a prominent influence on eye movement behavior. Saccade amplitude appears to be the most sensitive eye measure (i.e., as compared to pupil diameter and fixation duration). Additionally, we contribute new regression model equations of various driving scene contents on workload and eye measures. From the continuous data level resolution used to train these models, we have been able to perform multiple validation assessments to show that our models have predictive capabilities both in terms of repetition and generalization relevant towards anticipating nominal adaptive eye behavior of would-be drivers.

Supplementary materials

Supplementary materials for this article are available at https://doi.org/10.4121/uuid:b94716d5-8885-4cef-8f49-e707d2293fea.

Acknowledgment

This research was conducted within HFAuto – Human Factors of Automated Driving (PITN-GA-2013-605817).

References

Cabrall, C. D. D., Lu, Z., Kyriakidis, M., Manca, L., Dijksterhuis, C., Happee, R., & De Winter, J. C. F. (2018). Validity and reliability of naturalistic driving scene categorization judgments from crowdsourcing. *Accident Analysis & Prevention, 114*, 25–33.

Chapman, P. R., & Underwood, G. (1998). Visual search of driving situations: Danger and experience. *Perception, 27*, 951–964.

Cityscapes Dataset (2018). Semantic Understanding of Urban Street Scenes. https://www.cityscapes-dataset.com
De Waard, D. (1996). The measurement of drivers' mental workload. Ph.D. Thesis. Traffic Research Centre, University of Groningen. Haren, The Netherlands.
De Winter, J. C. F., Eisma, Y. B., Cabrall, C. D. D., Hancock, P. A., & Stanton, N. A. (2019). Situation awareness based on eye movements in relation to the task environment. *Cognition Technology and Work, 21*, 99–111.
Fisher, D. L., Pollastek, A. P., & Pradhan, A. K. (2006). Can novice drivers be trained to scan for information that will reduce their likelihood of a crash?. *Injury Prevention, 12*(Suppl 1), i25–i29.
Foy, H. J., & Chapman, P. (2018). Mental workload is reflected in driver behavior, physiology, eye movements and prefrontal cortex activation. *Applied Ergonomics, 73*, 90–99.
Hart, S., & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology, 52*, 139–183.
Jin, Y., Li, J., Ma, D., Guo, X, & Yu, H. (2017). A semi-automatic annotation technology for traffic scene image labeling based on deep learning preprocessing. In IEEE International Conference on Computation Science and Engineering (CSE) and Embedded and Ubiquitous Computing (EUC).
Kruger, J., & Dunning, D. (1999). Unskilled and unaware. *Journal of Personality and Social Psychology, 77*, 1121–1134.
Krzywinski, M. (2018). Image color summarizer: RGB, HSV, LCH & Lab image color statistics and clustering – Simple and easy. http://mkweb.bcgsc.ca/color-summarizer/?analyze.
Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature, 369*, 742–744.
Laya, O. (1992). Eye movements in actual and simulated curve negotiation tasks. *IATSS Research, 16*, 15–26.
Lu, Z., Coster, X., & De Winter, J. C. F. (2016). How much time do drivers need to obtain situation awareness? A laboratory-based study of automated driving. *Applied Ergonomics, 60*, 293–304.
Mourant, R. R., Rockwell, T. H., & Rackoff, N. J. (1969). *Drivers' eye movements and visual workload*. Washington, D.C.: Highway Research Record, No. 292.
Mourant, R. R., & Rockwell, T. H. (1970). Mapping eye-movement patterns to the visual scene in driving: An exploratory study. *Human Factors, 12*, 81–87.
Olson, P. L. (1964). The driver's reference point as a function of vehicle type, direction and radius of turn. *Human Factors, 6*, 319–325.
Papelis, Y., Ahmad, O. & Watson, G. (2003). Developing scenarios to determine effects of driver performance techniques for authoring and lessons learned. In Proceedings of the Driving Simulation Conference North America. Dearborn, Michigan, USA.
Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.
Victor, T. (2005). Keeping eye and mind on the road (Doctoral dissertation, Uppsala University, Uppsala, Sweden). http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A167500&dswid=-6616.
Yu, F., Xian, W., Chen, Y., Liu, F., Liao, M., Madhavan, V., & Darrell, T. (2018). BDD100K: A diverse driving video database with scalable annotation tooling. arXiv preprint, arXiv:1805.04687.