Quantifying the Impacts of Situational Visual Clutter on Driving Performance Using Video Analysis and Eye Tracking

SAFER SIM

SAFETY RESEARCH USING SIMULATION

UNIVERSITY TRANSPORTATION CENTER

Chengbo Ai, Ph.D.
Assistant Professor
Civil & Environmental Engineering
University of Massachusetts Amherst

Michael A. Knodler, Ph.D.
Professor
Civil & Environmental Engineering
University of Massachusetts Amherst

Quantifying the Impacts of Situational Visual Clutter on Driving Performance Using Video Analysis and Eye Tracking

Chengbo Ai, Ph.D., PI
Assistant Professor
Department of Civil and Environmental Engineering
University of Massachusetts Amherst
https://orcid.org/0000-0002-3536-9348

Michael Knodler Jr., Ph.D., co-PI
Professor
Department of Civil and Environmental Engineering
University of Massachusetts Amherst
https://orcid.org/0000-0002-6517-4066

Qing Hou
Graduate Research Assistant
Department of Civil and Environmental Engineering
University of Massachusetts Amherst
https://orcid.org/0000-0002-1599-6782

Francis Tainter
Graduate Research Assistant
Department of Civil and Environmental Engineering
University of Massachusetts Amherst
https://orcid.org/0000-0003-0180-6113

A Report on Research Sponsored by

SAFER-SIM University Transportation Center

Federal Grant No: 69A3551747131

June 2021

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, by a grant from the U.S. Department of Transportation's University Transportation Centers Program. However, the U.S. Government assumes no liability for the contents or use thereof.

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. UM-2-Y3	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Quantifying the Impacts of Situational Visual Clutter on Driving Performance Using Video Analysis and Eye Tracking		5. Report Date June 2021 6. Performing Organization Code
7. Author(s) Chengbo Ai, Ph.D https://orcid.org/0000-0002-3536-9348 Qing Hou, Graduate Research Assistant https://orcid.org/0000-0002-1599-6782 Michael Knodler Jr., Ph.D https://orcid.org/0000-0002-6517-4066 Francis Tainter, Graduate Research Assistant https://orcid.org/0000-0003-0180-6113		8. Performing Organization Report No.
9. Performing Organization Name and Address UMass Transportation Center 130 Natural Resources Rd 214 Marston Hall Amherst, MA 01003		10. Work Unit No. 11. Contract or Grant No. Safety Research Using Simulation (SAFER-SIM) University Transportation Center (Federal Grant #: 69A3551747131)
12. Sponsoring Agency Name and Address Safety Research Using Simulation University Transportation Center Office of the Secretary of Transportation (OST) U.S. Department of Transportation (US DOT)		13. Type of Report and Period Covered Final Research Report (September 2019 – June 2021) 14. Sponsoring Agency Code

15. Supplementary Notes

This project was funded by Safety Research Using Simulation (SAFER-SIM) University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program.

16. Abstract

The challenges in investigating the situational clutter are sourced from its complicated constitution of different contributors (e.g., vehicle, other road users, the road infrastructures, etc.) and its dynamically changing manner (e.g., dashboard display, traffic conditions and outlooks of the vehicles, dynamic road, and roadside landscapes, etc.). Although the psychology and cognitive science communities have investigated the situational visual clutter, there lacks effort in studying it in the driving context. The proposed study aims to bridge such a gap.

The objective of this study is threefold: 1) to develop a new video analysis model that can quantify the complex and dynamic driving scene; 2) to employ the developed model to quantify the impact of the situational visual clutter on driving performance, and 3) to demonstrate the potential of employing the driving scene quantification to support other retrospective studies and data mining using the existing driving simulation data.

17. Key Words		18. Distribution Statement	
Highways; Safety and Human Factors		No restrictions. This document is available through the SAFER-SIM website, as well as the National Transportation Library	
19. Security Classif. (of this report)	**20. Security Classif. (of this page)**	**21. No. of Pages**	**22. Price**
Unclassified	Unclassified	29	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Table of Contents

List of Figures

Abstract

The challenges in investigating the situational clutter are sourced from its complicated constitution of different contributors (e.g., vehicle, other road users, the road infrastructures, etc.) and its dynamically changing manner (e.g., dashboard display, traffic conditions and outlooks of the vehicles, dynamic road, and roadside landscapes, etc.). Although the psychology and cognitive science communities have investigated the situational visual clutter, there lacks effort in studying it in the driving context. The proposed study aims to bridge such a gap.

The objective of this study is threefold: 1) to develop a new video analysis model that can quantify the complex and dynamic driving scene; 2) to employ the developed model to quantify the impact of the situational visual clutter on driving performance, and 3) to demonstrate the potential of employing the driving scene quantification to support other retrospective studies and data mining using the existing driving simulation data.

1 Introduction

1.1 Background

Visual clutter and its impact on driving performance have been widely acknowledged. Visual clutter has been taxonomically categorized into three types (Edquist et al. 2007), including 1) the “situational clutter” that is sourced from the interaction among the driver, the vehicle, other road users, and the road infrastructure; 2) the “designed clutter” that is sourced from the existing traffic control devices, e.g., signage, signal, work zone, etc., and 3) the “built clutter” that is sourced from other roadside and roadway objects, e.g., billboard, roadside landscapes, etc. The impacts of both the designed and built clutter have been investigated using naturalistic driving measures and driving simulator designs and measures. Unfortunately, the situational clutter remains an open question, even though such a clutter type is considered to play a more lasting and profound role in impacting the driver’s performance, due to two primary reasons: 1) different sources of contributors, e.g., vehicles, other road users, road infrastructure, etc.; and 2) dynamically changing context, e.g., dashboard display, traffic conditions and outlooks of the vehicles, road conditions, roadside landscapes, etc. (Rosenholtz et al. 2007).

As a proven technology, driving simulation has successfully facilitated numerous driving safety improvement analyses through fully controlled or semi-controlled road contexts and driving conditions in the designed simulation scenarios. Meanwhile, driving simulation technologies have been advanced significantly in recent years through high-fidelity software and hardware, multi-sensor integration and collaborative simulation co-development (Fitzgerald et al. 2010). As the recent driving simulation systems enable a better understanding of the visual clutter and its impact on driving performance through comprehensive, synchronized data acquisition (including calibrated video clips for both interior and exterior, eye-tracking measures, detailed vehicle status, etc.), researchers have been able to achieve a better understanding of the impact of situational clutter.

Unfortunately, many in-service driving simulation systems are either low-cost systems or legacy systems that do not necessarily provide rich functionality for comprehensive, synchronized data collection. Therefore, while the designed clutter and built clutter may still be revealed based on unsynchronized data sources of video clips, vehicle status, simulation scenario designs, etc., the situational clutter that requires investigation on the interactions among the driver, the vehicle, other road users, and the road infrastructure are extremely challenging to be carried out. For example, in an earlier simulation study conducted by the co-PI (Knodler Jr et al. 2019), although video clips of driving participants and the environments, the corresponding eye-tracking data were collected, and the action taken by the participants, it remains challenging to integrate these pieces of information organically because that the data were usually collected independently. There was not any rigorous mechanism to synchronize them after the simulation tests consistently.

Moreover, over the years, researchers have accumulated a significant amount of participant observation data, especially video clips, in different simulation scenarios to have served their original purposes in various studies. There is a need for a method that can leverage the unsynchronized data to quantify the critical interactions among the driver, the vehicle, other road users, and the road infrastructure, i.e., situational clutter. Such a method can further utilize the valuable data, even from low-cost or legacy simulators, to shed light on other insights of the designed scenarios.

1.2 Objectives

The objective of this study is threefold: 1) to develop a new video analysis model that can quantify the complex and dynamic driving scene; 2) to employ the developed model to quantify the impact of the situational visual clutter on driving performance, and 3) to demonstrate the potential of employing the driving scene quantification to support other retrospective studies and data mining using the existing driving simulation data.

2 Methodology

2.1 Overview

To achieve the objectives of this study, this study proposed a methodology that consists of four key steps, including 1) image segmentation, 2) color/object profiling, 3) situational visual clutter quantification, and 4) eye-tracking linkage.

2.2 Data

This study introduced an existing dataset collected through a previous study under SaferSim (Knodler Jr et al. 2019), thanks to the collected rich data items and the comprehensive analysis regarding the driving trust and driver's mental workload. The collected data include vehicle data, eye-tracking data (eye movements), and first-person video clips of the driver. The driving simulator is a fixed-base Realtime Technologies Inc. (RTI) full cab with six screens surrounding it that subtend to 330 degrees of the horizontal field of view and 30 degrees of vertical field of view. The RTI driving simulator automatically recorded vehicle behaviors, and a SensoMotoric Instruments (SMI) head-mounted eye tracker was used to collect eye behavior data during the simulated drives. While the full questionnaire with detailed demographic and driving history information and the physiological sensor (a BioHarness chest strap sensor) were collected previously, this study did not use them due to the constraints of the research schedule. A total of 80 subjects aged 20-30 years participated in this study. All participants were recruited from the University of Massachusetts Amherst and the local area and were compensated for their time. All the participants had a U.S. driving license with a minimum of two years of driving experience. Figure 2.1 shows an example of the collected video frame and the visualization of the eye-tracking data (i.e., eye-gazing vectors).

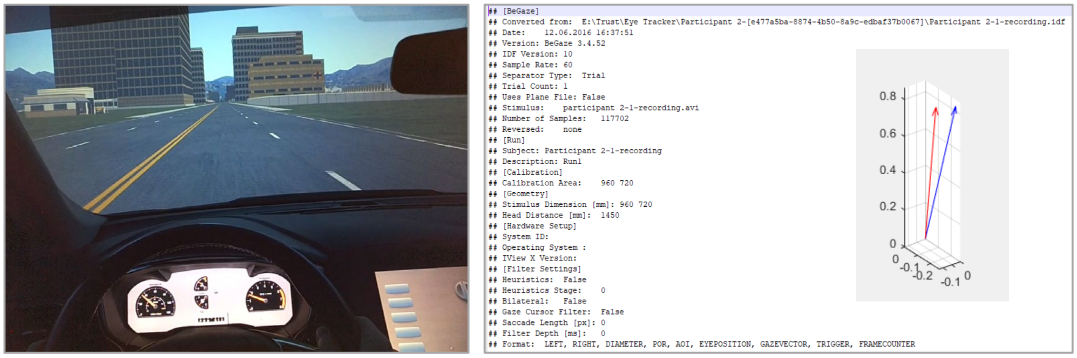

Figure 2.1 - An example of the collected video frame and the eye-tracking data

2.3 Step 1 - Image Segmentation

The objective of the image segmentation step is to create a coarse classification of the scene from the video frame efficiently and effectively. As discussed in Section 1, it remains challenging to create an end-to-end solution for accurate semantic segmentation. It is especially so when using relatively low-resolution video frames captured during the driving simulation. Therefore, the philosophy of this proposed step is not to achieve pixel-accurate results for the entire simulated driving scene, but instead to create a reasonably good segmentation result that can roughly separate the scene into four classes, including the interior of the vehicle, vehicle screen, pavement, and other exterior features. Figure 2.2 shows an example of the segmentation results.

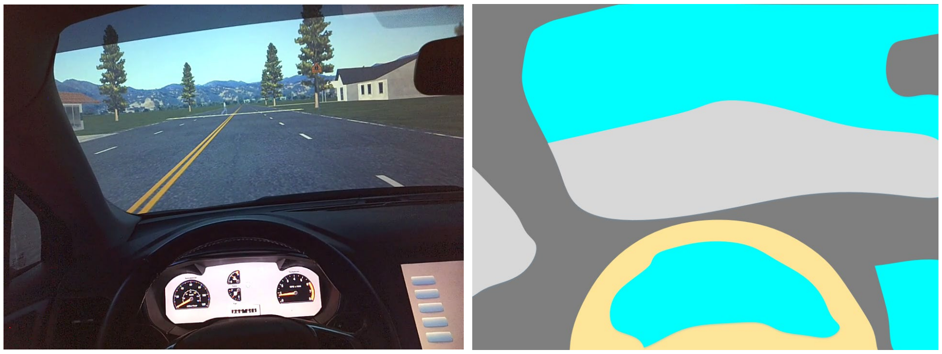

Figure 2.2 - An example of the image segmentation results

In this study, the fully convolutional network (FCN) (E. Shelhamer et al. 2017) architecture was introduced for the image segmentation thanks to its reported good performance, as well as the ease of implementation (with existing open-sourced toolboxes and pre-trained models). The main idea of FCN, unlike a typical convolutional neural network (CNN), is to perform convolution and resampling operations so that it does not use any fully connected layers, which will purposely destroy the local spatial information. In this study, the original implementation of FCN (E. Shelhamer et al. 2017) was introduced by employing the segmentation-equipped VGG net (FCN-VGG16) and 8x upsampling prediction (FCN-8).

The training dataset was derived from the Cityscape Dataset that contains 5000 images of street scenes with fine annotations and an addition of 5000 images of simulated street scenes from the video data captured in the current dataset (Knodler Jr et al. 2019). The mix of the training dataset contains both naturalistic and simulated driving scenes that provide a good diversity of the data for the trained model. Manual digitization was conducted to label the four classes at a pixel level using the Computer Vision Annotation Tool (CVAT) (Sekache et al., 2019). Figure 2.3 shows the interface of the labeling tool, CVAT.

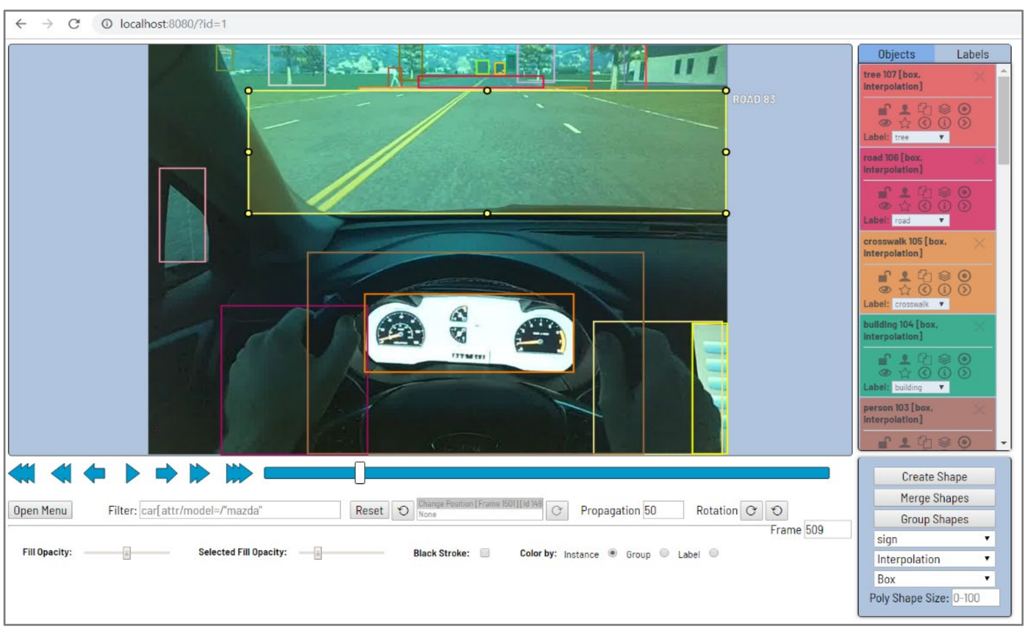

Figure 2.3 - The interface of the labeling tool - CVAT

2.4 Step 2 - Color/Object Profiling

The objective of the color/object profiling step is to create a fine classification for the regions that different objects of interest may occur. The output of Step 1 will create a coarse segmentation map with the four main classes, i.e., the interior of the vehicle, vehicle screen, pavement, and other exterior features. While the classes for the vehicle's interior and the pavement do not render any additional information, the classes for the vehicle screen and other exterior features often include different objects of interest. Therefore, the detailed color/object profiling process focuses on these two classes. In the segmented results, these two classes were further defined as Region 1 (i.e., other exterior features), Region 2 (i.e., vehicle dashboard), and Region 3 (i.e., vehicle central control screen). Figure 2.4 illustrates the regions of interest for processing in this step.

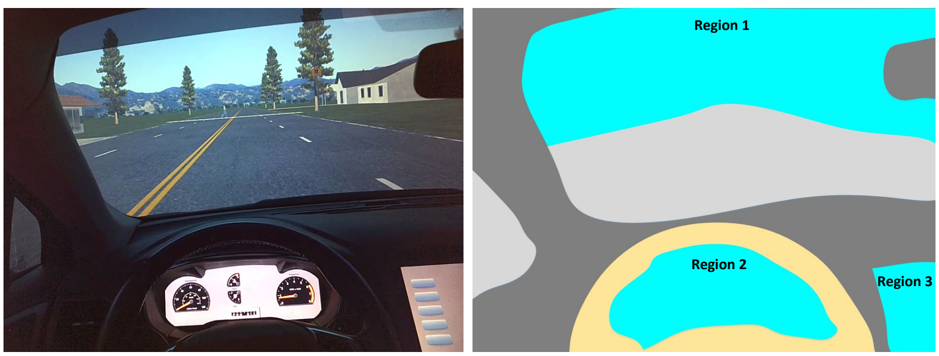

Figure 2.4 - An illustration of the regions of interest for processing

For Region 1, different objects, including signage, tree, house, ground, and sky, were defined to differentiate the potential gazing area by the driver. For Regions 2 and 3, different regions, including different message blocks of the screen from the dashboard (vehicle message) and the central control screen (instructional message), were defined to differentiate the potential message feeds for the driver. It can be observed that the image blobs for the "objects" of interest for Regions 1, 2, and 3 contain different levels of details, which makes it extremely challenging to employ a single model to identify the "objects" accurately. Therefore, two sets of strategies were employed for profiling Region

1 and Region 2/3 separately. For Region 1, where the "objects" are common roadway and roadside objects represented by complicated color and texture features, the FCN-VGG16 model employed in Step 1 was introduced here again. For Region 2/3, where the "objects" are just different color blocks that renders different vehicle and instructional messages, a simple color segmentation model developed by Ai and Tsai (Ai and Tsai 2016) were employed.

Figure 2.5 shows an example of the profiling results. It can be observed that, through Step 1 and Step 2, each of the original video frames was segmented into different image blobs. Each blob represents a self-contained "object" that may attract the driver's attention.

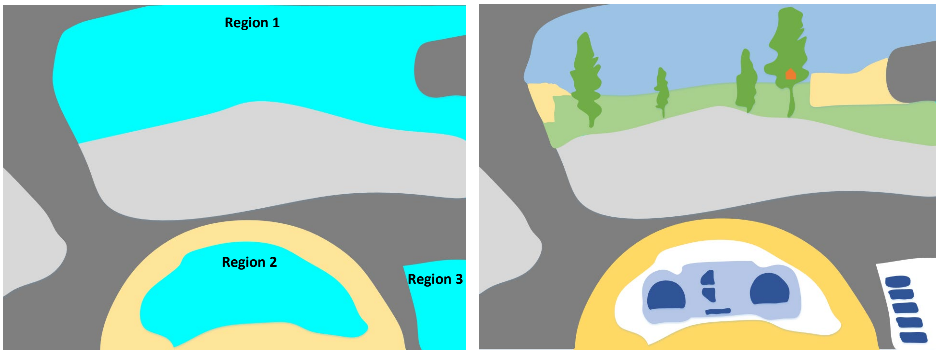

Figure 2.5 - An example of the color/object profiling results

2.5 Step 3 - Situational Visual Clutter Quantification

The objective of the situational visual clutter quantification step is to create a quantitative metric for each image frame (i.e., the instantaneous driving scene for the driver) that can consistently quantify the situational visual clutter. When driving, the driver relies on collecting information from both the interior and exterior of the vehicle to maintain course and conduct maneuvers. Therefore, the situational visual clutter will affect driving behaviors in two ways: 1) the scene itself is cluttered to understand, and 2) the object in the scene is complicated to recognize. In this study, the number of blobs and the blobs' complexity in the scene is defined to quantify the effects of situational visual clutter, respectively.

- The number of blobs: The number of blobs in each image frame quantifies the number of objects that appeared in this scene. A larger number of objects means that a more cluttered scene for the driver to understand. To differentiate the effect of the interior and exterior of the vehicle, the number of blobs were summarized individually, N_{ex} and N_{in} respectively.
- The complexity of blobs: The complexity of the blob represents how difficult for the driver to recognize the objects that appeared in the scene. While many metrics can be employed for quantifying the complexity, e.g., color variance, texture pattern and variance, entropy, etc., a simple mean of the gradient is defined in this study. The magnitude of the image gradient for each blob is defined as $|\nabla I(i,j)| = \sqrt{I_x^2(i,j) + I_y^2(i,j)}$, where the $I_x^2(i,j)$ and $I_y^2(i,j)$ are the approximate values of the directional derivatives in the horizontal and vertical directions. The magnitudes of the gradient within the blob are then averaged by the total number of pixels of the blob. A larger complexity of the blob means a more cluttered object to recognize. To differentiate the effect of the interior and exterior of the vehicle, the complexity of blobs was summarized individually, M_{ex} and M_{in} respectively.

Figure 2.6 shows an example of the quantified situational visual clutter. It can be noticed that the exterior of the vehicle renders a less cluttered scene than the interior of the vehicle (i.e., the number of blobs is smaller). In contrast, the objects in the vehicle's exterior are more complex (i.e., the complexity of blobs is larger).

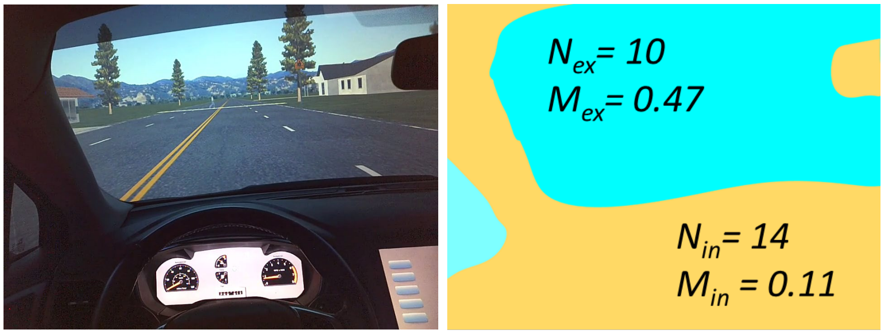

Figure 2.6 - An example of the quantified situational visual clutter in a frame

2.6 Step 4 – Eye Tracking Linkage

The objective of the eye-tracking linkage step is to map the vector of the eye tracker to the image coordinates. Hence, the association between the eye-tracking results can be associated with the image analysis results with quantified situational visual clutter values. However, as discussed in Section 1, the eye-tracking data may not be calibrated with the video data, making the association challenging. Therefore, this study only linked the eye-tracking results with the large regions of the scene, i.e., the exterior and interior of the vehicle. However, if accurate calibration between the eye-tracking devices and the video frames, the linkage between the vector of the eye tracker and the detailed objects captured in the video frame can be established, hence the object-based analysis of the visual clutter effect on driving behaviors may be conducted. Figure 2.7 shows an example of the results with accurate calibration. The red and blue crosses in the image frame represent the intersection between the vectors of the eye trackers and the objects in the image frame (i.e., pavement).

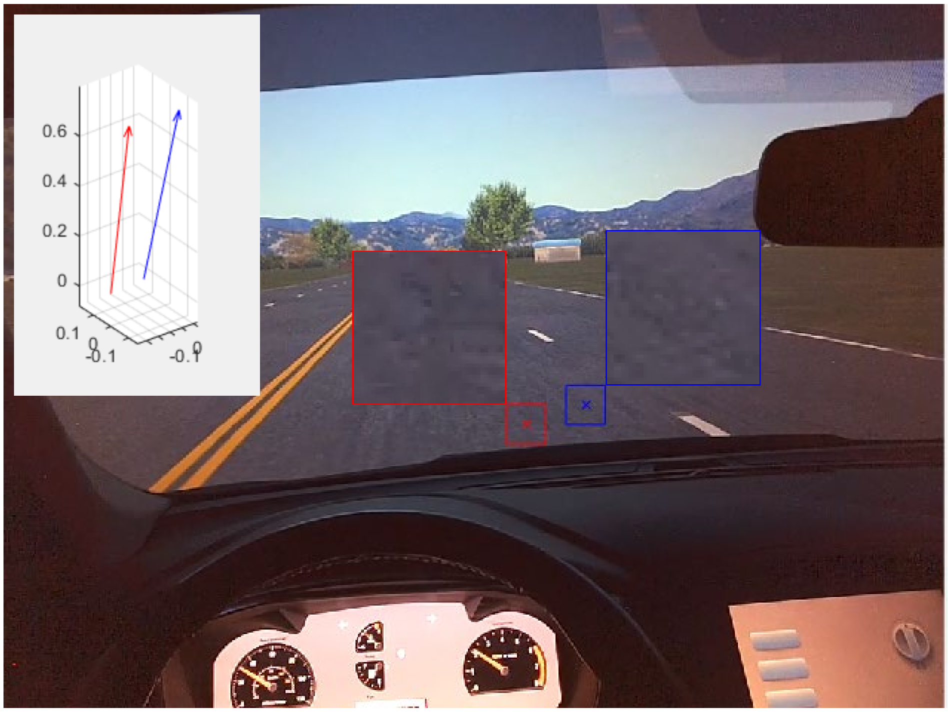

Figure 2.7 - An example of the results with accurate calibration

2.7 Summary

This study proposed a situational visual clutter quantification methodology based on a new video analysis model using the VGG16-FCN image segmentation algorithm and its subsequent object and color profiling methods. The complexity of the scene and the objects within the scene were quantified based on the number of blobs and the complexity of blobs derived from the image. By linking the eye-tracking results that are indicative of the driver's condition and engagement with the quantified results of the driving scene, the effect of the situational visual clutter on the driving performance can be revealed.

It should be noted that, in this study, the situational visual clutter was quantified based on the number of blobs and the complexity of blobs that are aggregated in two broader regions, i.e., exterior and interior of the vehicle. The results' aggregation was used because of the practical consideration of the image segmentation and object detection accuracy (in Steps 1 and 2) and the availability of the eye-tracking calibration (in Step 4), constrained by the existing dataset. However, the proposed methodology is general enough to be adopted for object-based analysis if high-resolution image data and the calibration with eye-tracking are available.

3 Results and Discussion

3.1 Video Analysis Results

In this study, a total of 5,953,786 image frames were processed using the proposed image segmentation and detection methods (Steps 1 and 2 of the proposed methodology). To quantify the overall performance of the algorithms, a dataset of 5000 image frames was randomly selected and manually digitized as ground truth. The performance evaluation metric of Intersection-Over-Union (IoU) was used. The segmentation IoU was reported as 0.857, while the detection IoU was reported as 0.821. While the overall performance did not achieve similar accuracies reported in previous studies using the same deep learning approach (Cordts et al. 2016; Niemeijer et al. 2017), the results are considered reasonable for the subsequent analysis of visual clutter quantification where aggregated results were needed.

3.2 Visual Clutter Quantification Results

It is well understood that the situational visual clutter may affect the driver's comprehension of the scene and subsequently affect the driving performance. To achieve this goal, the eye-tracking events captured in the previous study (Knodler Jr et al. 2019) are compared with the quantified visual clutter results using the corresponding video streams. By linking the quantified visual clutter and the eye-tracking events, the potential effect of the situational visual clutter may be revealed. This study attempted to answer the following three questions.

3.2.1 Do drivers obtain information differently from different regions of the scene?

There were two types of eye-tracking events captured in the dataset (Knodler Jr et al. 2019), including the fixation and saccade. "*Saccades are the type of eye movement used to move the fovea rapidly from one point of interest to another, while a fixation is the period of time where the eye is kept aligned with the target for a certain duration,* *allowing for the image details to be processed.*" (Rayner 2009). In the context of driving, saccades are generally considered as the driver is searching for information, while fixations are generally considered as the driver is acquiring detailed information. Figure 3.1 shows the results of the proportion of the two typical eye-tracking events that occurred at different regions of the scene. The results show that drivers are more likely to search and acquire information in Region 1, whereas they are less likely to search and acquire information in Region 2. The results can be intuitively understood, as Region 1 covers the roadway and roadside environment (except the pavement surface) that includes many objects, e.g., signage, signals, vegetations, buildings, etc. so that the drivers need to process unknown information. In contrast, Region 2 renders similar information, e.g., speed, RPM, etc., without any sudden changes so that the drivers do not need to process intensive information search and acquisition tasks. For Region 3, as the information may be unknown to the drivers but remained the same throughout most driving sessions, the drivers only need to process the median load of search and acquisition tasks.

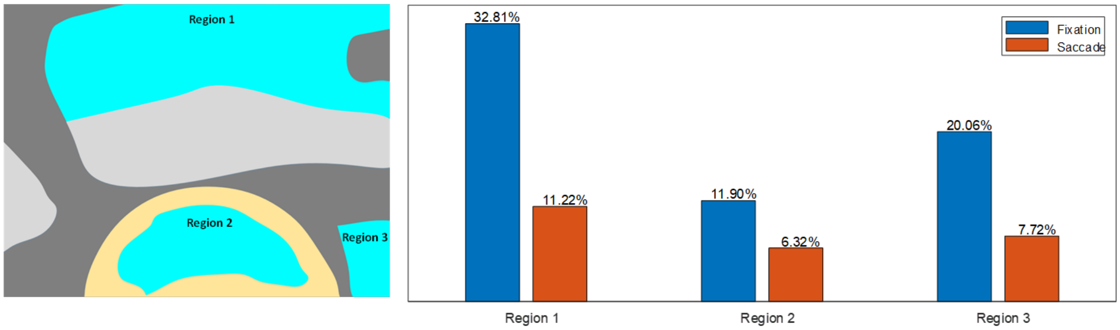

Figure 3.1 - Results of the eye-tracking events in different regions

In summary, drivers obtain information from different scene regions with different levels of frequencies and intensities. In particular, drivers attempt to search and acquire more frequently and intensively from the regions where unknown information may be present. At the same time, they spend less effort in the regions where anticipated or known information may be present.

3.2.2 How do drivers search/acquire information when the scene is cluttered?

Figure 3.2 shows the proportion of eye-tracking events that occurred in the interior and exterior of the vehicles when different numbers of blobs (objects) are present in the corresponding regions.

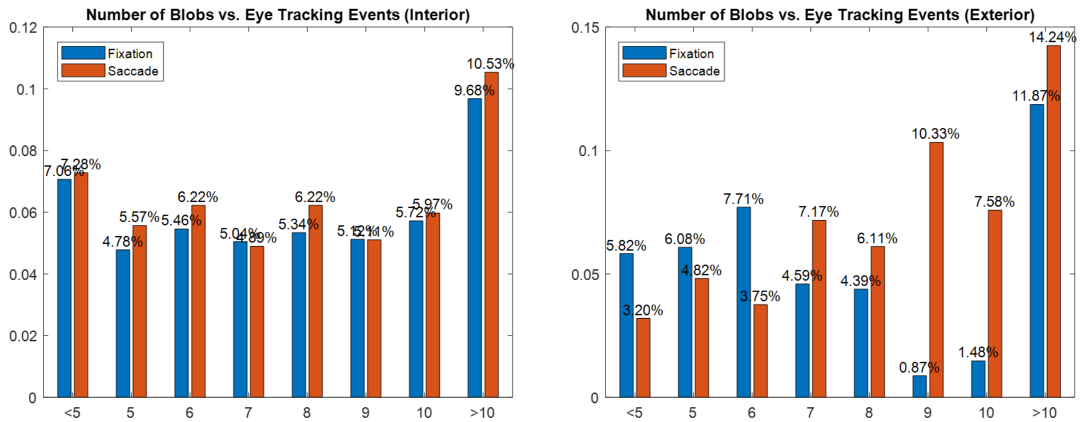

Figure 3.2 - Results of the eye-tracking events with different numbers of blobs

For the interior, the results show that, regardless of the number of blobs (i.e., how cluttered the interior of the scene is), drivers search and acquire information consistently, unless the scene becomes extremely cluttered (i.e., number of blobs >10). Such a phenomenon can be attributed to the reason that a driver should be familiar with the vehicle's interior, including the information displayed in the dashboard or central screens, and they do not expect to search or acquire any additional information that they are not anticipating. However, suppose the scene becomes extremely cluttered. In that case, drivers may need to engage more attention to identify if any additional information that is critical to their driving might occur, e.g., a warning message from the dashboard, etc., or much unknown information start to appear, e.g., instructional messages from the central control, etc.

For the exterior, the results show that drivers tend to search for information more frequently and intensively if the number of blobs increases (i.e., the exterior scene becomes more cluttered). In contrast, they tend to dwell on certain objects to acquire detailed information with a larger number of blobs. Such a phenomenon can be attributed to the reason that drivers are not necessarily familiar with the surroundings of the driving course, where they may expect unknown information to process. Therefore, if the scene is cluttered and complex, drivers will spend more effort searching for such unknown information. However, such a tendency does not necessarily translate to more effort on acquiring detailed information because 1) with a large number of blobs (i.e., the exterior scene becomes more cluttered), drivers can no long fixate on a single object for more detailed information, but rather browse for more information broadly; 2) most of the objects may be familiar or unimportant to the drivers so that they do not necessarily need to spend much effort on the details.

In summary, drivers do not necessarily search or acquire more information when the scene is more cluttered. In particular, drivers do not increase the frequency and level of intensity in searching or acquiring information if the more cluttered scene is familiar to them. However, drivers tend to increase the frequency and intensity in searching information if the more cluttered scene is not familiar to them or unknown information is anticipated in a familiar scene. In addition, in an unfamiliar but cluttered scene or if unknown information is anticipated in a familiar but cluttered scene, drivers tend to dwell less on a single object for more detailed information. Still, they spend more effort on browsing all the objects that appeared in the cluttered scene.

3.2.3 How do drivers search/acquire information when objects are complex in the scene?

Figure 3.3 shows the proportion of eye-tracking events that occurred in the interior and exterior of the vehicles when objects with different levels of complexity are present in the corresponding regions.

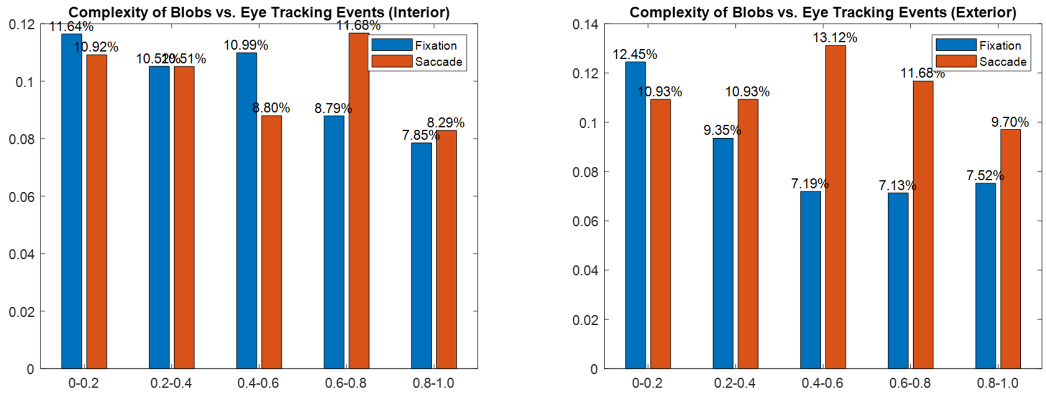

Figure 3.3 - Results of the eye-tracking events with different complexities of blobs

For the saccade events, drivers consistently search for important and unanticipated information, regardless of the complexity of the blob (i.e., how complex the object is in the scene). Such a phenomenon was observed for both the interior and exterior of the vehicle. For the interior, it can be attributed to the reason that drivers should be familiar with the interior of the vehicle, and they do not expect to search for any additional information; for the exterior, it can be attributed to the reason that drivers will continue searching for unanticipated information from the roadside or roadway. The complexity of individual objects in the scene does not affect how the drivers search for more information for the exterior (or how the drivers are reluctant to search for more information from the interior).

For the fixation events, drivers tend to reduce the frequency and intensity for acquiring information from the scene with more complex objects for both the interior and exterior of the vehicle. In contrast, they spend more effort on objects that are easier to understand. Such a phenomenon can be attributed to the reason that acquiring information from complex objects may not be feasible or comfortable for the drivers to achieve. Therefore, they will not pursue the information any longer. This is consistent with the philosophy of why simple and unambiguous designs are critical for traffic control devices, operational control of the vehicles, etc.

In summary, drivers do not necessarily increase or decrease their frequency and level of intensity for search unanticipated information regardless of the complexity of the object in the scene. However, drivers tend to increase the frequency and intensity of acquiring information if the object is less complex.

4 Conclusion and Future Studies

Visual clutter and its impact on driving performance have been widely acknowledged. Unfortunately, as a critical component of visual clutter, situational clutter remains an open question, even though such a clutter type is considered to play a more lasting and profound role in impacting the driver's performance. There is a need for a method that can leverage the unsynchronized data to quantify the critical interactions among the driver, the vehicle, other road users, and the road infrastructure, i.e., situational clutter.

In this study, a new situational visual clutter model that objectively quantifies the complex and dynamic driving scene based on video analysis and the linkage with eye-tracking data was developed. The video analysis tool was established upon the emerging deep learning framework, i.e., VGG16-FCN, and the image color and texture analysis algorithms. The number and complexity of the blobs in the segmented image frames were defined to quantify the level of the clutter of the scene and the object that appeared in the scene, respectively. The potential usage of the developed model to support other retrospective studies and data mining was explored using the existing driving simulation data captured (Knodler Jr et al. 2019).

This study shows that the newly developed model for quantifying situational visual clutter can potentially reveal insight on how drivers search and acquire information from the scene and how the clutter may impact the effectiveness of the search and acquisition. More importantly, the developed methodology for video analysis provides an effective means for leveraging the existing driving simulation data and quantifying the possible factors impacting driver's behavior and performance, including situational visual clutter.

For future studies, the following directions are recommended to improve the performance of the proposed methodology further: 1) to evaluate other deep learning and image processing approaches for better (consistent) quantification of situational visual clutters; 2) to investigate the impact of uncalibrated eye-tracking data on the performance of the proposed method; and 3) to reveal more insights on how situational visual clutters may affect the driving performance by incorporating vehicle status data.

5 References

1. Ai, C., and Tsai, Y. J. (2016). "An automated sign retroreflectivity condition evaluation methodology using mobile LIDAR and computer vision." Transportation Research Part C: Emerging Technologies, 63, 96–113.
2. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., and Schiele, B. (2016). "The Cityscapes Dataset for Semantic Urban Scene Understanding." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
3. E. Shelhamer, J. Long, and T. Darrell. (2017). "Fully Convolutional Networks for Semantic Segmentation." IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(4), 640–651.
4. Edquist, J., Horberry, T., Regan, M., and Johnston, I. (2007). "'Visual clutter' and external-to-vehicle driver distraction."
5. Fitzgerald, J., Larsen, P. G., Pierce, K., Verhoef, M., and Wolff, S. (2010). "Collaborative Modelling and Co-simulation in the Development of Dependable Embedded Systems." Integrated Formal Methods, D. Méry and S. Merz, eds., Springer Berlin Heidelberg, Berlin, Heidelberg, 12–26.
6. Knodler Jr, M., Christofa, E., Hajiseyedjavadi, F., Tainter, F., and Campbell, N. (2019). "To Trust or Not to Trust? A Simulation-based Experimental Paradigm." Harvard Dataverse.
7. Niemeijer, J., Fouopi, P. P., Knake-Langhorst, S., and Barth, E. (2017). "A Review of Neural Network-based Semantic Segmentation for Scene Understanding in Context of the self-driving car."
8. Rayner, K. (2009). "The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search." Quarterly Journal of Experimental Psychology, SAGE Publications, 62(8), 1457–1506.

9. Rosenholtz, R., Li, Y., and Nakano, L. (2007). “Measuring visual clutter.” Journal of Vision, 7(2), 17–17.