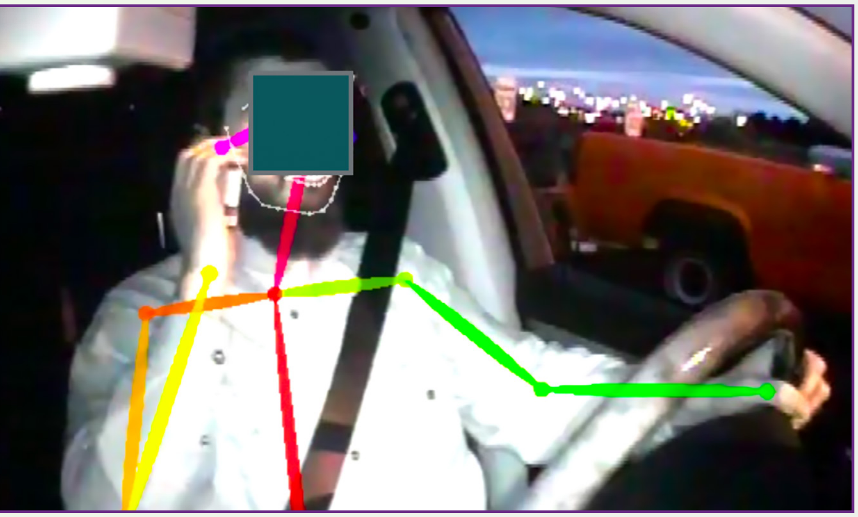
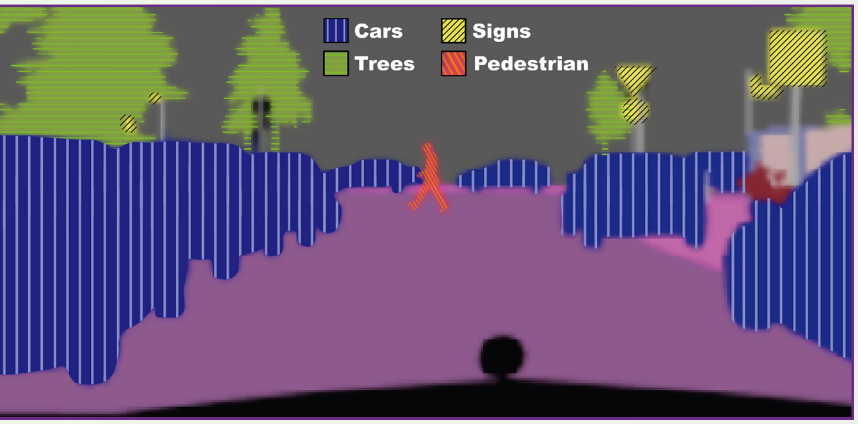
Exploratory Advanced Research Program

Using Video Analytics to Automatically Annotate Driver Behavior and Context in Naturalistic Driving Data

Research Summary Report

EXPLORATORY ADVANCED RESEARCH





U.S. Department of Transportation
Federal Highway Administration

Cover photo credits clockwise from top left:

Technical Report Documentation Page

1. Report No. FHWA-HRT-24-078	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Using Video Analytics to Automatically Annotate Driver Behavior and Context in Naturalistic Driving Data		5. Report Date December 2024 6. Performing Organization Code
7. Author(s) Abhijit Sarkar (ORCID: 0000-0003-0525-5240), Lynn Abbott, Jeffrey Hickman, Ioannis Papakis, Calvin Winkowski, Debanjan Datta (ORCID: 0000-0001-5997-5142), Akash Sonth (ORCID: 0000-0002-5045-5906), Hirva Bhagat, Shreyas Bhat, Sandesh Jain (ORCID: 0009-0002-5274-3693), Andrei Svetovidov, Rich Hanowski (ORCID: 0000-0002-3937-3061), Omkar Kaskar		8. Performing Organization Report No.
9. Performing Organization Name and Address Virginia Tech Transportation Institute 3500 Transportation Research Plaza Blacksburg, VA 24061 Schatz Publishing Group 11950 W. Highland Ave. Blackwell, OK 74631		10. Work Unit No. 11. Contract or Grant No. 693JJ319C000004
12. Sponsoring Agency Name and Address Office of Research, Development, and Technology Federal Highway Administration 6300 Georgetown Pike McLean, VA 22101-2296		13. Type of Report and Period Covered Research Summary Report; August 2019–November 2022 14. Sponsoring Agency Code HRT-10
15. Supplementary Notes The contracting officer's representative was Charles Fay (HRSO-20; ORCID: 0000-0002-2498-9795).		
16. Abstract Naturalistic driving data provides a wealth of information for researchers studying driver behavior and distracted driving. However, manually annotating the videos to extract the data is costly and time consuming. This project's research team set out to develop a system to analyze videos from the second Strategic Highway Research Program Naturalistic Driving Study dataset and automatically produce annotations and descriptors for events, behavior, and driving scenarios related to transportation safety.[1] The project had four objectives: characterize high-level driver behavior, such as eating or using a cellphone; classify the environment outside the vehicle, such as the position of roadway objects, work zones, and intersections; understand interactions and dependencies between drivers and the surrounding environment, such as looking at a billboard or a passing vehicle; and demonstrate how the video analytics techniques used in this study can help human factors researchers address research questions in novel ways. The researchers developed and tested advanced computer vision algorithms, including deep neural network (DNN)-based methods to capture spatial and temporal information embedded in the naturalistic driving videos. The DNN models included convolutional neural network models for image recognition and transformer-based models to process sequential data. All the codes developed as part of this project are open sourced.		
17. Key Words Naturalistic driving, computer vision, video analytics, driver behavior, object detection, gaze fixation, scene perception, object saliency, event correlation, human factors	18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, VA 22161. https://www.ntis.gov	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 24 22. Price N/A

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

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.755	cubic meters	m^3
		NOTE: volumes greater than 1,000 L shall be shown in m^3		
		MASS		
ounces		28.35	grams	g
pounds		0.454	kilograms	kg
short tons (2,000 lb)		0.907	megagrams (or “metric ton”)	Mg (or “t”)
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
foot-candles		10.76	lux	lx
foot-Lamberts		3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
poundforce		4.45	newtons	N
poundforce per square inch		6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or “t”)	megagrams (or “metric ton”)	1.103	short tons (2,000 lbs)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	2.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbfin^2

*SI is the symbol for International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.
(revised March 2003)

TABLE OF CONTENTS

INTRODUCTION 1

PROJECT OVERVIEW 3

- DATASETS 3
- ANNOTATING THE DATA 3

PROJECT METHODOLOGY 4

- DRIVER BEHAVIOR 4
  - Body Pose Estimation 4
  - Head Pose and Gaze Direction 5
  - Classification of Secondary Behaviors 5
  - In-Vehicle Object Detection 6
  - Passenger Detection 6
- CONTEXT OUTSIDE THE VEHICLE 7
  - Work Zones 9
  - Scene Perception 10
  - Object Detection and Tracking 11
- INTERACTIONS BETWEEN DRIVER AND DRIVING ENVIRONMENT 12
  - Event Correlation and Object Saliency 12
  - Gaze Fixation to Outside Objects 13
  - Safety Analysis 13
- HUMAN FACTORS RESEARCH APPLICATIONS 14

CONCLUSION 17

REFERENCES 19

LIST OF FIGURES

Figure 1. Illustration. Many things in the driving environment compete for a driver's attention.[4] 1

Figure 2. Graph. Combined mean recall values for all keypoints by action classes. 4

Figure 3. Flowchart. Processing pipeline for detecting vehicle occupants.[12] 7

Figure 4. Illustration. Semantic segmentation versus instance segmentation.[14] 8

Figure 5. Photos. Work zone scenes from a naturalistic driving dataset.[17] 9

Figure 6. Photo. MOT assigns a unique ID number to each object in a scene.[21] 11

Figure 7. Image. Driver gaze direction zones.[26] 12

Figure 8. Photo. VDHA helps identify secondary behaviors such as using a cellphone. 14

LIST OF TABLES

Table 1. Examples of secondary behavior descriptions using a visual dictionary. 5

LIST OF ABBREVIATIONS

3D three dimensional

ADAS advanced driver assistance systems

AI artificial intelligence

CNC crashes and near crashes

CNN convolutional neural network

CV computer vision

DNN deep neural network

EAR Exploratory Advanced Research

FHWA Federal Highway Administration

FOV field of view

HPV Head Pose Validation

ID identification

mAP mean average precision

mAR mean average recall

ML machine learning

MOT multiobject tracking

NDS Naturalistic Driving Study

PoG point of gaze

SHRP2 second Strategic Highway Research Program

VDHA visual dictionary of human action

VRU vulnerable road user

VTTI Virginia Tech Transportation Institute

Distracted driving threatens the safety of drivers, passengers, and anyone outside the vehicle and contributes significantly to crashes and near crashes (CNC). In 2022, 3,308 people were killed, and 289,310 were injured in motor vehicle crashes involving distracted drivers.[2] Distractions can come from inside the vehicle, such as the driver using a cellphone, eating, or talking to passengers, and outside the vehicle due to pedestrians, cyclists, work zones, billboards, and more (figure 1).

Naturalistic driving data captures real-world driver behavior and helps researchers understand and mitigate the effects of distracted driving. However, the wealth of information in large datasets, such as from the second Strategic Highway Research Program (SHRP2) Naturalistic Driving Study (NDS), must be annotated for researchers to study specific driver information.[1] Typically, data are manually annotated, but manual annotation is prone to human bias, time consuming, and costly for large datasets. Moreover, manual annotation does not address how drivers interact with other road users, roadside objects, and infrastructure and how these interactions relate to crashes. An automated annotation system would address these issues and enable researchers to access unprecedented levels of data.

To develop such a system, the Federal Highway Administration's (FHWA) Exploratory Advanced Research (EAR) Program supported the project Video Analytics for Automatic Annotation of Driver Behavior and Driving Situations in Naturalistic Driving Data.[3] Led by the Virginia Tech Transportation Institute (VTTI), the project's research team adapted and developed tools from computer vision (CV) and machine learning (ML) to automatically annotate driver behavior, driving context, and interactions between the driver and the environment in the SHRP2 NDS database.[1] The researchers developed and evaluated a series of deep neural network (DNN) models—artificial intelligence (AI) that recognizes patterns in data—to capture spatial and temporal information embedded in the videos. The DNN models included AI image recognition models called convolutional neural networks (CNNs) and transformer-based models that process sequential data. The researchers then used CV methods to automatically generate annotations and descriptors for transportation safety-related events, behavior, and driving scenarios.

The study aimed to accomplish the following:

1. Characterize high-level driver behavior inside the vehicle (e.g., eating, looking at a cellphone, or fixating on an object) by estimating head and body pose and gaze direction.
2. Classify the environment and context outside of the vehicle (e.g., work zones, intersections, and vulnerable road users (VRUs)).
3. Examine interactions and dependencies between drivers and the driving environment and effectively predict gaze fixation to outside objects (e.g., a passing vehicle or a billboard).
4. Demonstrate how the video analytics techniques developed apply to human factors research.

In addition to these objectives, the researchers wanted to identify locations and situations that may pose challenges to advanced driver assistance systems (ADAS) or autonomous vehicle navigation systems. For example, some ADAS detect driver drowsiness but not distraction from secondary behaviors (e.g., texting, talking to a passenger). If a system can detect driver distraction, it can warn the driver to take preventive measures.

The automated annotations developed for this study will allow researchers to explore questions about driver distractions that are currently beyond investigators' reach. This project showed that CV methods can do the following automatically:

- Detect passengers and objects.
- Determine a driver's head and body pose, scene awareness related to safety, and gaze direction.
- Characterize the driving environment by detecting work zones and intersections.

The algorithms the researchers used can process 37 different annotations in the SHRP2 NDS data dictionary.[1] The researchers were also able to correlate a driver's gaze direction with objects detected outside a vehicle. This result has the potential to reveal how drivers behave in various traffic situations and contribute to the development of robust ADAS systems that can prevent traffic crashes. The video analytics tools the researchers developed to detect cellphone use, eating, driving behavior in work zones, and gazing at signs along roadways (particularly billboards) may help human factors researchers identify these behaviors and develop automated monitoring systems that alert drivers to take corrective actions. To ensure the public benefitted from these findings, the codes and CV models developed for this project were all made open source.

Figure 1. Illustration. Many things in the driving environment compete for a driver's attention.[4]

Because this project addressed multiple areas of CV, such as object detection, pose estimation, and action recognition, the researchers conducted a detailed literature review for each project task to identify the best methods and associated algorithms. They then used evaluation metrics to capture the performance of the algorithms they developed. The researchers tested methods for detecting driver behaviors, evaluating context and tracking objects outside the vehicle, correlating driver gaze to external objects, and generating high-level video descriptors. Then, they examined how the methods from those tests could be applied to human factors research for automotive safety.

One challenge the researchers faced was that CV and ML techniques often do not generalize well for situations other than those for which they were developed. The researchers assessed several CV models for their ability to analyze naturalistic driving videos and perform detection and estimation tasks. Many existing CV methods required additional training or fine-tuning to perform well with driving-specific datasets.

Most of the researchers' methods employed transfer learning, a field of ML that transfers knowledge from one domain application to another. Transfer learning uses some of the learning from one database or domain to train on a target dataset. The researchers used pretrained models and retrained, or fine-tuned, them with annotations for the targeted case. CNNs contain millions of parameters, and transfer learning can fine-tune all of the parameters or a subset. The researchers found that fine-tuning the whole network performed better than only a subset of parameters.

DATASETS

The research team used several datasets for this project depending on the task, ease of access, and availability of existing annotations. For images and videos, they primarily used SHRP2 NDS, which contains data from more than 3,400 drivers with 32 million miles driven over 5.5 million trips.(1) The dataset includes the speed, acceleration, braking, turn signal, and distance to other objects in the scene for each trip. In the study, four video camera views and one wide-angle still camera captured images inside and outside the vehicles. The SHRP2 NDS dataset contains manual annotations for approximately 42,000 CNC and baseline events. The researchers for this study performed additional annotations for gaze fixation and object tracking.

Other datasets the researchers used focused on head pose, driver actions, and pixel-level annotations. The researchers also developed a new dataset specifically targeted for inside-the-vehicle passenger detection.(5)

ANNOTATING THE DATA

The SHRP2 NDS data dictionary includes more than 75 reduction variables describing the driver or driving environment attributes.(1) For this project, data reductionists manually added many annotations for images or events from the datasets, which accomplished the following:

- Included attributes such as lighting conditions, driving situations, weather conditions, and driver impairment.
- Aided the researchers' understanding of the variability of the overall dataset used in their investigations.
- Helped the researchers test and evaluate the algorithm's performance under diverse conditions.

Ideally, an algorithm should perform consistently regardless of the variability in the data. For example, naturalistic driving videos present technical challenges, such as variations in illumination inside and outside the vehicle, and an object detector should detect roadway objects, such as cars and pedestrians, regardless of lighting conditions.

This project required the following types of annotations:

- Image-based classification—A single image annotated to a specified class (e.g., a city scene versus a highway scene or a work zone versus a nonwork zone).
- Sequence-based classification—A series of images annotated to capture the start time and end time of a video sequence to a single class (e.g., eating or texting).
- Pixel-based classification—A pixel-level understanding of an image is needed, such as for object detection and tracking (e.g., a car).
- Attribute-based classification—Ground-truth annotations for image-based attributes that require additional annotations or measurements for the image or a section of the image (e.g., a driver's three-dimensional (3D) head pose angle measurement).

The following sections summarize the investigations the researchers performed to achieve the project's objectives.

DRIVER BEHAVIOR

The researchers identified several factors from the driver's behavior inside the vehicle that could either indicate or lead to distractions. They investigated methods for detecting the following in naturalistic driving videos:

- Body pose estimation.
- Head pose and gaze direction.
- Classification of secondary behavior.
- In-vehicle object detection.
- Passenger detection.

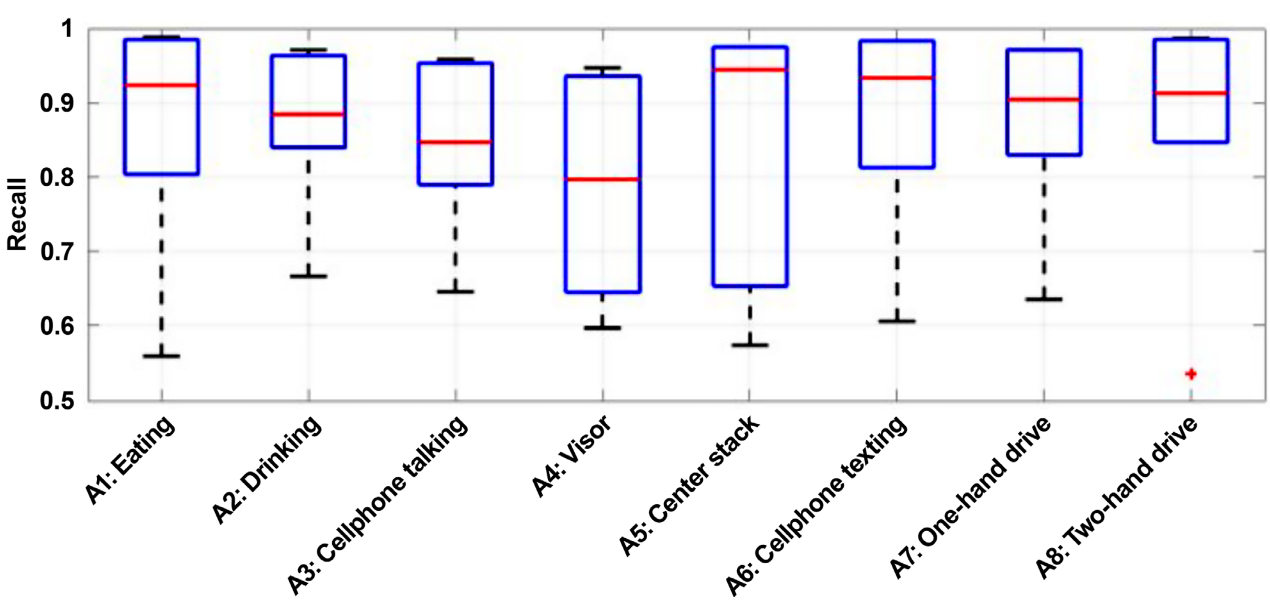
Body Pose Estimation

Identifying a driver's body pose is important for determining the driver's activity, including engagement in secondary behaviors and driver distraction. Techniques for determining body pose usually focus on locating the positions of keypoints related to prominent body joints (e.g., shoulders, elbows, wrists, hips, knees, and ankles) and the head.

The researchers tested three deep-learning methods to identify the driver's body pose inside the vehicle using the VTTI Machine Learning Pose dataset, which provides 80,000 images from 25 participants in naturalistic driving and simulated naturalistic driving conditions.(6) Some drivers in the videos performed secondary tasks while driving. This study used approximately 5,500 images from the dataset comprising 8 different action classes, including eating with the right hand while the left hand is on the steering wheel, talking on the phone with the right hand against the ear while the left hand is on the steering wheel, driving with both hands on the steering wheel, etc.

Of the three models they tested, the researchers obtained the best results with HRNet, which uses a top-down approach to first detect people and then estimate keypoint types and locations.(7) The model most reliably detected the driver's right shoulder and right elbow. The driver's left wrist and left elbow were not identified consistently, which was expected since those joints are often blocked from the camera's view because the camera was mounted near the rearview mirror.

Across the eight action classes, several activity types had recall values above 90 percent, including eating, interacting with the center stack (the portion of the dashboard where navigation, entertainment, and climate controls are often located), texting, one-handed driving, and two-handed driving (figure 2). Drinking and talking on a cellphone had recall values between 80 and 90 percent. The lowest recall value,



Source: FHWA.

Figure 2. Graph. Combined mean recall values for all keypoints by action classes.

approximately 80 percent, occurred for driver interactions with the visor.

Head Pose and Gaze Direction

Understanding driver gaze behavior is essential because CNC events increase significantly when drivers take their eyes off the road. According to the National Highway Traffic Safety Administration, drivers taking their eyes off the road for 5 s at 55 mph is equivalent to driving the length of a football field with their eyes closed.(8) The ability to predict gaze fixation may help ADAS systems warn distracted drivers and predict the takeover time in partially automated vehicles.

The researchers used an Oak Ridge National Laboratory dataset containing videos of drivers' faces and ground-truth head pose measurements.(9) Ten participants were asked to look at 16 different locations inside a stationary vehicle. Later, these same participants performed the same tasks while driving. Two cameras recorded the scenarios, and each participant also wore a head-mounted gyroscope that collected data for head-pose angles.

To predict gaze location, the researchers developed two deep-learning algorithms to estimate a driver's head pose from in-vehicle camera images.(10) One algorithm used a single image-based CNN model. The second used a recurrent neural network model—a type of AI neural network that uses sequential or time-series data—to study the temporal patterns of the 3D head pose angles (roll, pitch, and yaw) and predict a driver's gaze locations, such as forward, rearview mirror, left mirror, radio, speedometer, etc.

The temporal model performed better overall than the CNN model for predicting a gaze fixation target. To the best of the researchers' knowledge, this project is the first to use large-scale naturalistic data to study the gaze locations of drivers. The implications go beyond safety events to facilitate continuous data processing. The algorithm processed individual images at 375 frames per second. At that rate, processing all 7.7 billion frames in the SHRP2 dataset would take less than 250 d on a single graphics processing unit.(1)

Classification of Secondary Behaviors

Secondary behaviors while driving are distracting by definition because they take away from the primary task of driving. Understanding the role each secondary task plays in distraction is problematic because these tasks overlap with the primary task. In this investigation, the researchers introduced a set of visual attributes that are semantically explainable to humans and easily detectable by a machine. Specifically, the researchers looked at driver behavior inside a vehicle by representing secondary behaviors using a method based on a visual dictionary. A visual dictionary of human action (VDHA) is a collection of the temporal sequence of human behaviors and interactions between objects and body parts captured in a systemic way to uniquely represent any human action.

Table 1 shows how microactions by the driver (i.e., movement of the head, hand, and mouth), spatial relations between parts of the driver's body (i.e., body pose, head pose, and hand-to-face distance), and the driver's interactions with the surroundings (i.e., at least one hand off the wheel and the presence of an object) can uniquely represent five

Table 1. Examples of secondary behavior descriptions using a visual dictionary.

Secondary Task Name	Head Movement	Hand Movement	Mouth Movement	At Least One Hand Off Wheel	Body Pose Off Normal	Head Pose Off Normal	Hand-to-Face Distance	Presence of Object
Cellphone, talking	Maybe	No	Yes	Yes	No	No	Close	Yes
Adjusting radio	Yes	Yes	No	Yes	Maybe	Maybe	Far	No
Applying makeup	Maybe	Yes	Maybe	Yes	Maybe	Maybe	Close	Yes
Eating with utensils	Maybe	Yes	Yes	Yes	No	No	Close	Yes
Looking at pedestrian	Yes	No	No	No	No	Yes	N/A	No

N/A = not applicable.

secondary tasks. “Maybe” indicates that the task could be performed whether or not the dictionary attribute was present. For example, a driver could adjust the radio with or without the head being away from the forward view (off normal).

The researchers used the visual dictionary elements to encode 52 secondary behaviors chosen by manual annotation from the SHRP2 dataset.[(1)] Next, they fine-tuned four existing CNN action recognition models to identify drivers’ secondary activities. The research team then demonstrated how each of the secondary behaviors can be automatically processed from videos using CV. The visual dictionary’s associative nature made concurrent behavior modeling possible when a driver is involved in multiple actions at the same time. The CV methods employed in this study could be used for further study of drivers’ actions and visual dictionary elements.

In-Vehicle Object Detection

For this task, the researchers chose videos from driver-facing cameras to detect and identify in-vehicle objects that could distract the driver. Using a third-party dataset with cab-facing images manually annotated for objects visible inside the vehicle, the researchers worked with 5,021 images with approximately the same distribution of object classes. The dataset was assigned 25 object classes with labels such as “grocery bag,” “dog,” “backpack,” “bottle,” and “phone.”

The researchers used a CNN object detector that assigned class labels and bounding boxes to objects detected in an image. To evaluate the network’s performance, they computed the mean average precision (mAP) and mean average recall (mAR) for each object the network detected.[(11)] All classes combined had mAP and mAR values of 38.4 and 45.5 percent, respectively. The highest precision and recall values were for the object categories “face,” “person,” and “logo,” which may be a result of emphasizing the face and person classes during the network’s pretraining. The “phone” class had mAP and mAR values of 56.0 and 52.8 percent, respectively. “Food” and “bottle” classes obtained lower values, perhaps because the driver’s hand partially blocked these objects from view.

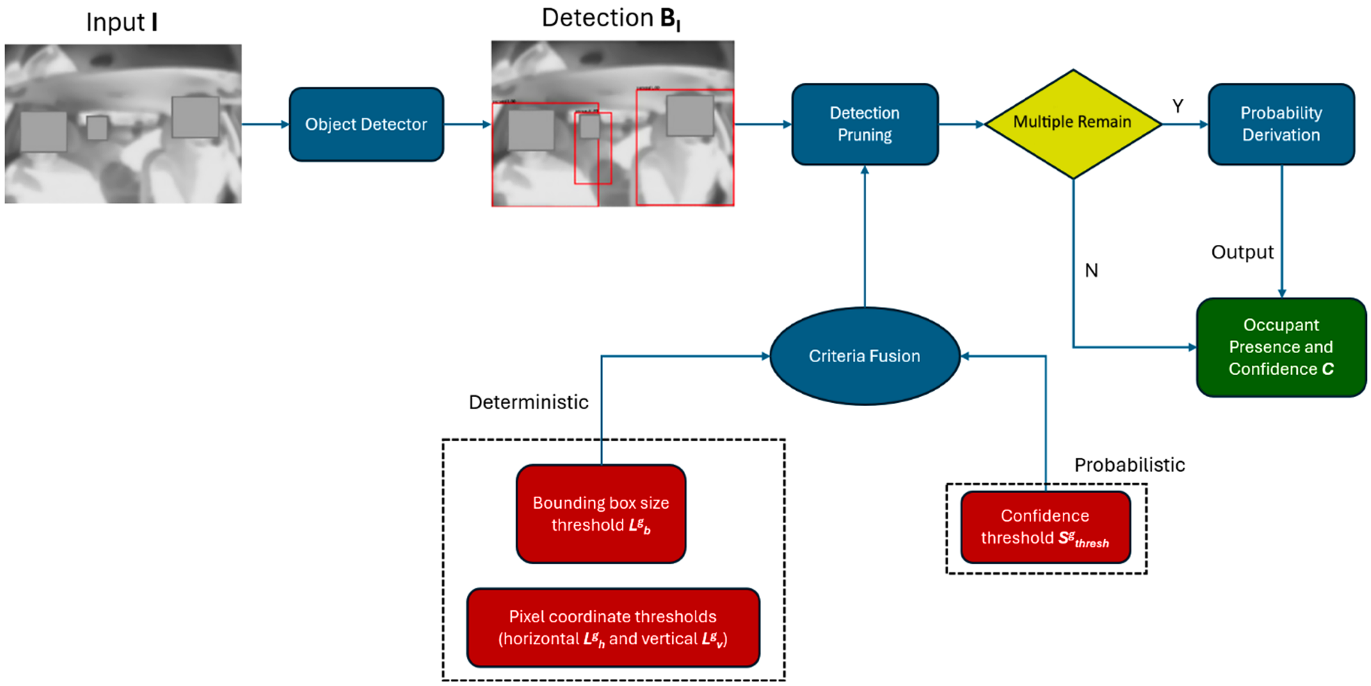
Passenger Detection

By developing a system to detect passengers in naturalistic driving videos automatically, the researchers aimed to facilitate studies related to passenger-driver interactions and vehicle safety. For this task, they used two SHRP2 datasets containing in-vehicle images and annotations for three object categories: driver, front-seat passenger, and back-seat passenger.[(1)] For privacy, the SHRP2 videos have blurred facial images, making automated analysis more difficult. The researchers found that standard person-detection methods often failed. However, a passenger’s presence and location in a vehicle tend to be consistent throughout a video, which helped the researchers automatically detect vehicle occupants.

The occupant-detection algorithm for this task used a two-step method. First, a CNN-based object detector drew a bounding box around each person detected in the image. Second, the system assigned each person to one of three subclasses—driver, front-seat passenger, or back seat passenger—based on their locations in the vehicle and the sizes of the bounding boxes from step one. Figure 3 shows the process for detecting individuals in the vehicle.

Two different test sets for occupant detection resulted in 94.5- and 98.1-percent accuracies. One of the SHRP2 test sets for the classification of occupants resulted in a 99.5-percent accuracy for drivers, 97.3-percent for front-seat passengers, and 94.3-percent for back-seat passengers. In the second SHRP2 test set, all drivers and front-seat passengers were classified correctly, and the classification accuracy of back-seat passengers was 97.7-percent. The researchers also analyzed the system’s performance on day and night images for front-seat passengers. The scores were slightly higher for daytime images (98.5-percent accuracy) than nighttime images (96.7 percent).

This system automatically detected and classified occupants in vehicles from cabin images. The researchers expect this work will facilitate new research involving passengers and in-vehicle occupants, including interactions between drivers and passengers in studies related to safety. The model and image annotations the researchers used are available as open-source resources.[(12)]



Source: FHWA.

Figure 3. Flowchart. Processing pipeline for detecting vehicle occupants.[13]

CONTEXT OUTSIDE THE VEHICLE

In addition to investigating driver behavior, context, and distractions inside the vehicle, the researchers developed methods to automatically detect work zones, the driving environment, and objects outside the vehicle. The researchers used semantic image segmentation, where a class label is assigned to every pixel in an image. Relevant classes in these systems were included for processing roadway images, such as “pedestrian,” “traffic sign,” and “car.”

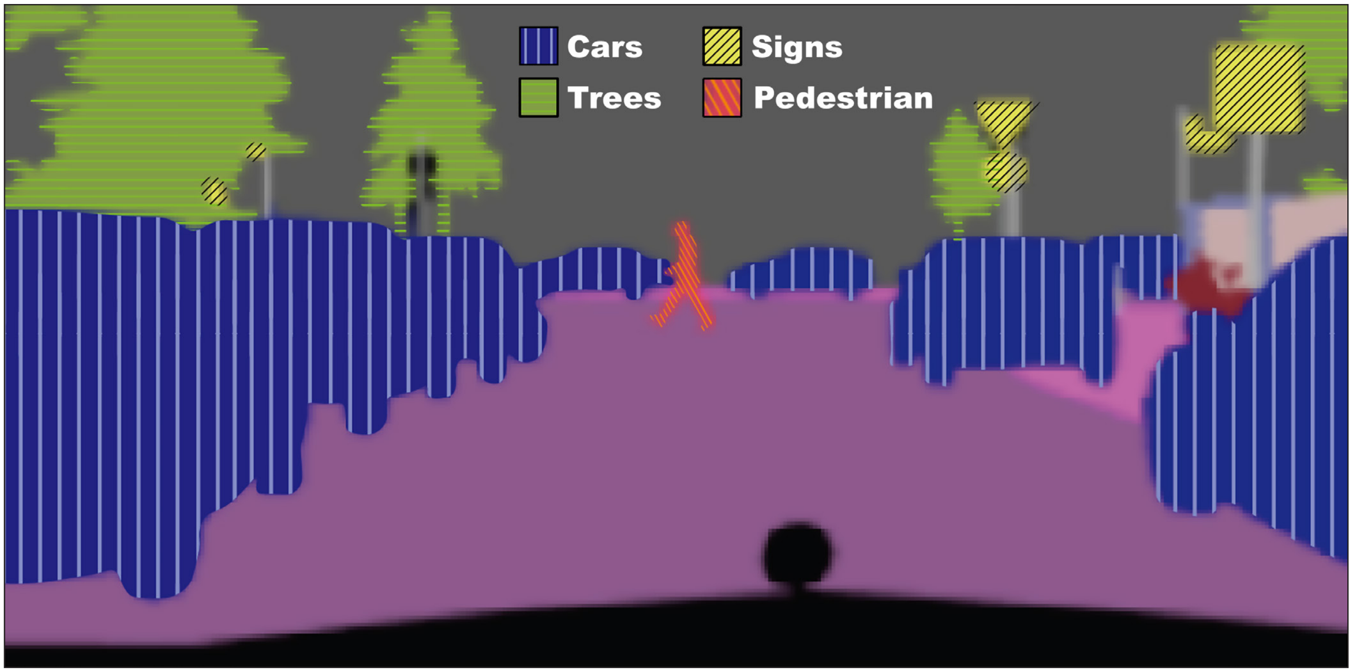
The main difference between object or instance segmentation and semantic image segmentation is that object detection systems assign a bounding box to each object identified in an image, but semantic segmentation systems identify similar objects as belonging to the same class. For example, in semantic segmentation, all cars in an image might be labeled “car” even if the pixels belong to different cars (figure 4).

The researchers evaluated several publicly available, state-of-the-art image segmentation systems, emphasizing systems trained using the CityScapes dataset.[14] CityScapes is a large dataset with urban scenarios that has been annotated for classes of interest to the transportation community. The researchers tested the segmentation systems’ performance on a new third-party dataset. Two of the segmentation models were virtually tied for best performance for daytime images. For all of the models, the daytime image performance was better than nighttime image performance.

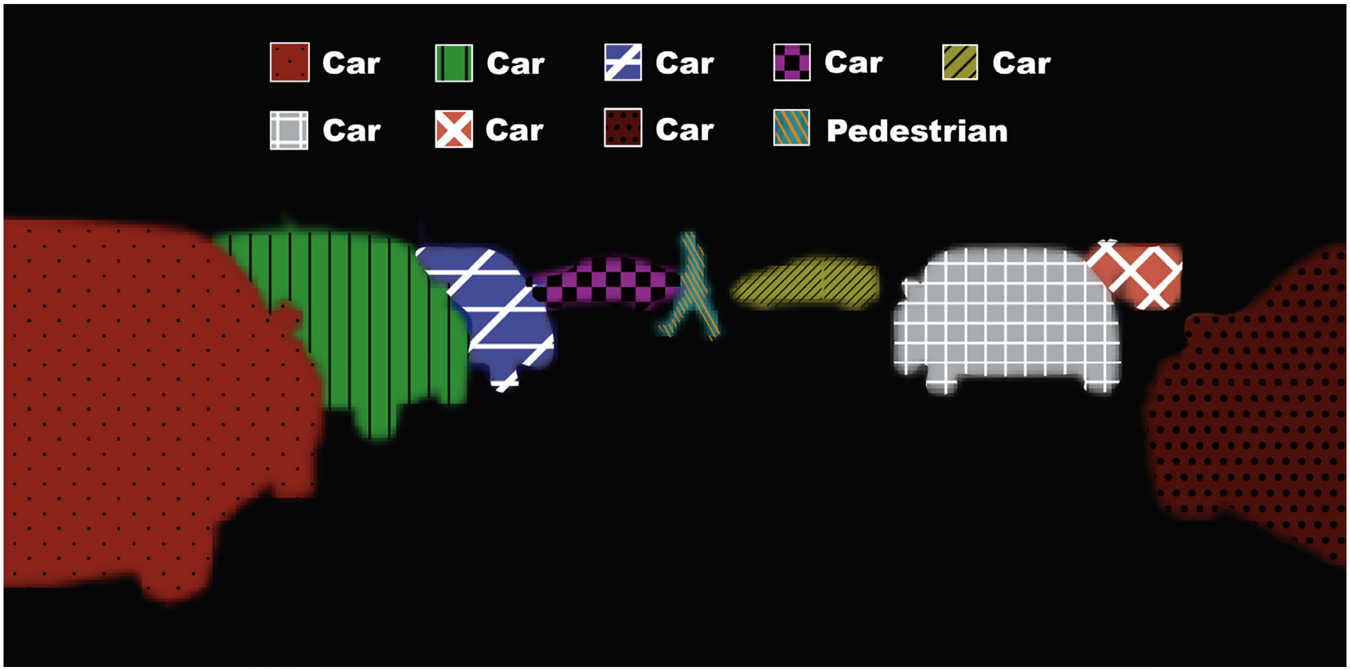
The researchers conducted investigations into methods for detecting the following:

- Work zones and related traffic-control objects.
- Driving environment based on the type of roadway, intersections, and lane markings.
- Billboards.
- Traffic density estimations.
- Object tracking outside of a vehicle.
- VRUs.

For these investigations, the researchers used videos from forward-looking cameras in vehicles and adapted recently developed CV models to automatically determine the external context, such as traffic density and the presence of work zones.



A. Semantic segmentation.



B. Instance segmentation.

Figure 4. Illustration. Semantic segmentation versus instance segmentation.(14,15)

Work Zones

Work zones pose unique and challenging scenarios for drivers because they introduce temporary, possibly unexpected, changes to the roadway. Even if drivers or automated driving systems receive warnings about work zones, these areas remain high risk because the distribution of construction-related objects on the road can change. Detecting and localizing objects in work zones is imperative for developing effective driver safety assistance systems.

For this task, the researchers aimed to detect work zones in naturalistic driving images to identify potential challenges to ADAS or autonomous vehicle navigation systems. They used a new dataset called VTTI-WZDB2020 derived from a large-scale naturalistic study of more than 1,000 truck drivers over millions of miles of driving.(16) The dataset consists of work zone and nonwork zone scenes in various conditions, a variety of roadway conditions (e.g., highway, driveway, rural, and urban), and a range of weather and illumination conditions (figure 5). Manual annotations helped the researchers choose events in work zones and nonwork zones, including CNC, hard-braking, and swerving events.

A. Daytime work zone.

B. Work zone in wet conditions.

Figure 5. Photos. Work zone scenes from a naturalistic driving dataset.(17)

The researchers manually annotated images they selected from the dataset to identify work zone objects, such as drums, cones, barricades, signs, barriers, construction equipment, people, and trailers. The final dataset for the task consisted of images from work zones and nonwork zones. Most of the work zone images included two to seven objects in each image.

The researchers combined image segmentation with object detection and used two ML models to detect work zones.[18] The first model used whole-image classification to determine if an image showed a work zone. This system correctly classified work zone scenes for 95 percent of the test cases and nonwork zones for 97 percent of the test cases. This test set used 1,576 images. Of the 1,004 nonwork zone images in the set, only 31 were misclassified as work zones. For the 572 work zone images, 28 were misclassified as nonwork zones.

The second model used image segmentation to detect and localize work zone objects on the same set of 1,576 images. Using this model, only 32 work zones were misclassified. Through saliency analysis (i.e., attention of the neural network while making a prediction), the researchers determined that some work zone objects were not contributing to the classification, and some nonwork zone objects (e.g., traffic signs and vehicle taillights) were being detected as a work zone. More examples with diverse work zone objects and using prior information could help the network correctly classify more scenarios.

The system learned to emphasize objects that commonly appear in work zones (e.g., cones, drums, and signs) even though this training procedure provided no explicit information related to those objects. The model also picked up various work zone objects in different lighting conditions. When objects were close to each other, the model sometimes had difficulty segmenting the objects. Reflections of work zone objects on the road during rainfall also caused the model to detect objects incorrectly. Adding more data points and exploiting the spatial distribution of work zone objects would help reduce the number of false detections.

Scene Perception

Understanding the driving environment, including roadway features and the presence of other vehicles, is crucial to transportation safety since drivers must maneuver according to the scene and react to changes. Therefore, the researchers investigated systems to evaluate the scene outside the vehicle.

According to FHWA, more than 50 percent of automotive crashes involving injury or death occur at or near intersections.[19] The researchers explored CV methods for detecting and classifying intersections versus nonintersections; day versus night; and urban, residential, or interstate/highway environments. They manually selected and labeled video frames from de-identified video files in the SHRP2 dataset that contained intersections and those that did not.[1] The annotated subset was used to train the system.

The system correctly classified intersections 97 percent of the time by finding indicators such as traffic lights, stop signs, and lines on the road surface.[20] The absence of these indicators helped the network determine no intersection was present. The network was less conclusive for identifying the driving environment. It correctly classified interstate/highway environments 83 percent of the time; the lower overall accuracy resulted from misclassification of urban and residential driving environments.

The researchers gained two primary insights from this experiment. First, the system they used was computationally light with a fast inference time, and a similar system could be used in active and passive driver assistance systems. Second, systems based on deep-learning models need careful attention during the design phase and rigorous testing before they can be used in the real world.

Another aspect of the scene perception that the researchers worked with was estimating traffic density near a vehicle. Using views from forward-facing cameras and an image segmentation system that detects individual vehicles, the researchers identified a trapezoid-shaped region of interest in front of the primary vehicle corresponding to the rectangular region on the road surface (assuming the surface is flat). Then they counted the vehicles that overlapped the region of interest partially or entirely and computed the traffic density by dividing the number of vehicles detected by the estimated road surface area.

Next, the researchers categorized the traffic into density classes:

- Free flow with very little traffic.
- Flow with light restrictions.
- Stable flow but with more restrictions on speed and maneuverability.
- Unstable flow with severe restrictions or stoppages.

The image segmentation approach allowed the researchers to manually adjust the traffic density classification without having to retrain a CNN model. The system can be tailored to adapt to pedestrian or bicycle traffic density. A disadvantage of the image segmentation approach is that nearby vehicles block more distant vehicles from view, restricting the system's ability to estimate density from a single image. Using a sequence of frames and refining density estimates over time might overcome this issue.

The researchers also investigated the performance of lane-detection algorithms. For this experiment, the researchers annotated 700 images with lane lines and tested three lane-detection algorithms to determine how well the algorithms could detect lanes under different operating conditions. The results revealed that lane detection is one of the most challenging issues. Most of the algorithms successfully detected solid lines; however, many struggled to detect dashed lines and lines that were at least two lanes away from the vehicle. The best algorithm achieved a recall value of 0.42.

Object Detection and Tracking

Using SHRP2 NDS and other third-party NDS data, the researchers annotated approximately 4,000 images.[(1)] They used transfer learning to train a new CNN-based model to use with the SHRP2 data to detect and track objects in a scene.[(11)] The model performed particularly well in identifying pedestrians and cyclists. The researchers analyzed the performance of the object detectors in different conditions. The model performed well for daytime images and images with less occlusion. (An object labeled as occluded is more than 50 percent blocked from the camera's view.)

As part of the object detection task, the researchers developed an algorithm to detect billboards in an image.[(11)] Using data from a third-party NDS, the researchers selected 700 images that had been annotated for billboards. To the researchers' knowledge, this is the first work to target billboard detection in NDS videos. Identifying billboards in naturalistic driving data could help human factors researchers study the impact of billboards on distracted driving, as discussed in the human factors applications section of this summary.

Along with object detection, the researchers examined object tracking, a field of CV research focused on tracking an object over time. As the number of objects in a scene increases, so does the complexity of the task. An ideal object-tracking algorithm in a driving scenario needs to track all the objects simultaneously, regardless of the challenges posed by differences in categories, size, appearance, trajectories, and movement speed. A new CV field called multiobject tracking (MOT) addresses this challenge. An object detector finds all the objects in a scene, and then the tracking algorithm assigns a unique identification (ID) number to each object in a video sequence, similar to the image in figure 6 captured from a closed-circuit television camera. The researchers used these models for multiple classes, including VRUs.

© Multiple Object Tracking Benchmark. Modifications by FHWA.

Figure 6. Photo. MOT assigns a unique ID number to each object in a scene.[(21)]

The researchers tested several MOT algorithms for low-quality, high-compression videos similar to the videos in the SHRP2 NDS.(1) Using 184 events from the SHRP2 NDS that included object categories such as cars, trucks, pedestrians, stop signs, and traffic signals, the researchers used the object detector they developed and ran the MOT method.(11) The algorithm developed by the researchers performed much better than the other models.

INTERACTIONS BETWEEN DRIVER AND DRIVING ENVIRONMENT

To investigate the high-level goal of analyzing interactions between drivers and the driving environment, the researchers examined drivers' head and eye movements and their gaze directions. Previous studies by Yarbus demonstrated that eye movements are task dependent.(22) Other studies have shown correlations between head and eye movements and driver drowsiness and between recurring gaze patterns and driver activities, which may predict vehicle maneuvers.(23,24) To illustrate how event-correlation models can reflect a driver's attention and actions relative to driving situations, the researchers examined the direction of a driver's gaze during common driving events and regarding different objects outside the vehicle. The researchers also performed safety analysis based on drivers' gaze fixation.

Event Correlation and Object Saliency

Using videos from the VTTI Head Pose Validation (HPV) dataset—which provides information about the yaw and pitch angles of the driver's head—the researchers categorized the driver's gaze direction as central, left, right, or left or right outside the camera's field of view (FOV), as shown in figure 7.(25) Next, they manually examined 23 short HPV video segments with 1 event per sample for left and right turns at stop signs and left and right turns at traffic lights. In a separate step, the researchers located the salient object closest to the fixation point in the image.

Figure 7 shows a driver's gaze direction zones. The results showed the highest concentrations of gaze direction toward the turn directions (e.g., gazing to the right for a right turn). Although drivers focused sharply on stop signs, their gaze often wandered to other objects once they recognized the stop sign. Other vehicles received more gaze fixation

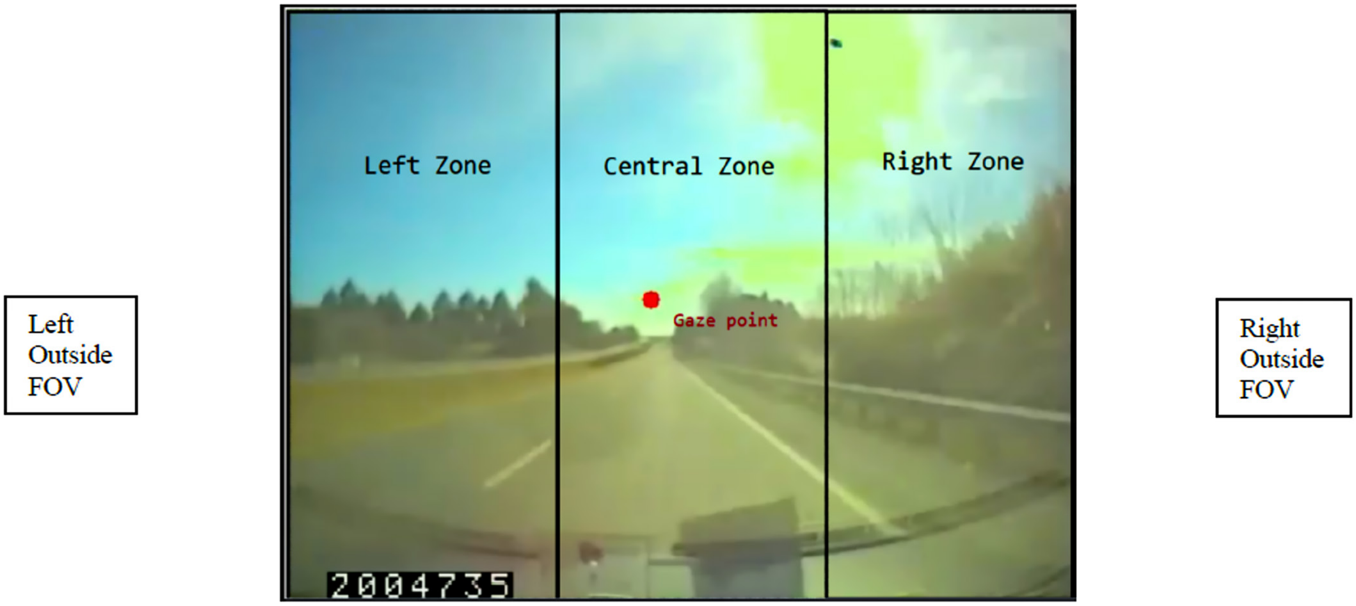


Figure 7. Image. Driver gaze direction zones.(26)

than traffic lights when drivers were making right turns at signalized intersections. The researchers also looked at merge events and lane changes and found that driver gaze direction was more evenly distributed.

This work is the first to consider the development of a correlation model that relates to driver behavior with automatically detected salient objects and events.[27] Correlations between gaze direction, salient objects, and the driving scenario can potentially reveal how drivers behave in different traffic situations and may help improve automated safety-improvement systems by predicting driver behavior.

Gaze Fixation to Outside Objects

A driver's gaze direction is an essential indicator of attention, which can impact traffic safety. Therefore, the researchers considered the gaze direction of drivers in several scenarios to support the event correlation model and developed a new algorithm to predict a driver's point of gaze (PoG) to outside objects.[28]

For this task, the researchers automatically analyzed images from a camera aimed at the driver to obtain the driver's gaze direction. They mapped PoG positions to the vehicle's frame of reference to identify objects outside the vehicle as possible fixation points. In driving situations, short sequences of fixations occur in patterns depending on the scenario, such as making a turn at an intersection. Based on the estimated gaze direction of the driver, the researchers' system selected the closest object in an image (within 20 pixels) from the forward-facing camera (dimension: 480 by 360 pixels). The system results proved highly accurate, suggesting that the research team's approach, in many cases, can automatically determine objects outside a vehicle that have attracted a driver's attention.

In addition to testing drivers' PoG on static objects, the researchers considered temporal sequences of fixation points by plotting eye movement, such as when a driver approaches a signalized intersection and must repeatedly give attention to the traffic lights to detect signal changes.

Safety Analysis

A driver must look at the driving scene and safely maneuver the vehicle. Predicting what object or event outside the vehicle will distract the driver is challenging. One possible way to predict distractions is by understanding which objects in a scene attract attention; the visual saliency of an object helps researchers predict locations that can attract driver attention. For this investigation, the researchers set out to automatically characterize objects that may attract attention and understand gaze fixation patterns.

The SHRP2 NDS dataset contains approximately 9,000 CNC events and approximately 32,000 baseline driving events that have been manually annotated.[1] The annotations include drivers' gaze location for all events and a frame-by-frame annotation of a driver's gaze at the level of the rear-view mirror, forward, left mirror, right windshield, left windshield, etc. The researchers selected 666 CNC events and 446 baseline events in which the driver's gaze was directed through the right windshield and fixed for 2–5 s.

The annotations focused on the following:

- The type of object fixated on (e.g., car, pedestrian, billboard).
- The vehicle's motion (e.g., traveling straight, changing lanes).
- The motion of the object fixated on (e.g., stationary on the roadside, moving right to left).
- The driving environment (e.g., city, intersection, highway).
- The time the gaze fixation started and ended.
- The driver's head and eye movements (e.g., whether the driver had any visible head movement or only used eye movements).

The researchers performed additional manual annotations on the gaze fixation events using both the forward-facing video and the driver's face video, resulting in 617 CNC events and 410 baseline events. Statistical analysis on both sets of events revealed that drivers' gazes were more firmly fixed on roadway objects and their movements during CNC events and more on roadside or stationary objects during baseline events. In other words, gaze fixation on dynamic objects contributes more to safety-critical events than static roadside objects.

Some of the specific findings are as follows:

- CNC events showed driver head movement 73 percent of the time compared to 66 percent for baseline events.
- CNC events had an average object gaze fixation time 0.3 s higher than baseline events.
- Drivers spent more time perceiving information on billboards and at intersections during CNC events.
- Drivers in baseline events were most often looking at stationary objects (89 percent of the time). In CNC events, the driver was mostly looking at the movement of other roadway objects, such as cars and pedestrians, and only at stationary objects 44 percent of the time.
- Driver distraction for all events was exceptionally high when drivers were turning, negotiating with VRUs, and maneuvering to avoid objects.

HUMAN FACTORS RESEARCH APPLICATIONS

One of the research team's objectives was to demonstrate how the video analytics methods from this project can help human factors researchers address automotive safety questions in new ways. The researchers looked at four driver behaviors detrimental to automotive safety: using a cellphone, eating, driving in work zones, and gazing at signs along roadways (particularly billboards). Automated monitoring systems that detect these behaviors and alert drivers to take corrective actions, such as braking or steering, would improve safety by preventing accidents. The researchers propose that these four behaviors could be automatically detected using CV systems.

This study demonstrated that CV techniques can automatically detect cellphone use while driving in some situations. The researchers developed a classifier to detect secondary behavior, including using a cellphone (figure 8). In-vehicle object detection experiments resulted in 56.0 percent mAP and 52.8 percent mAR detecting cellphones.

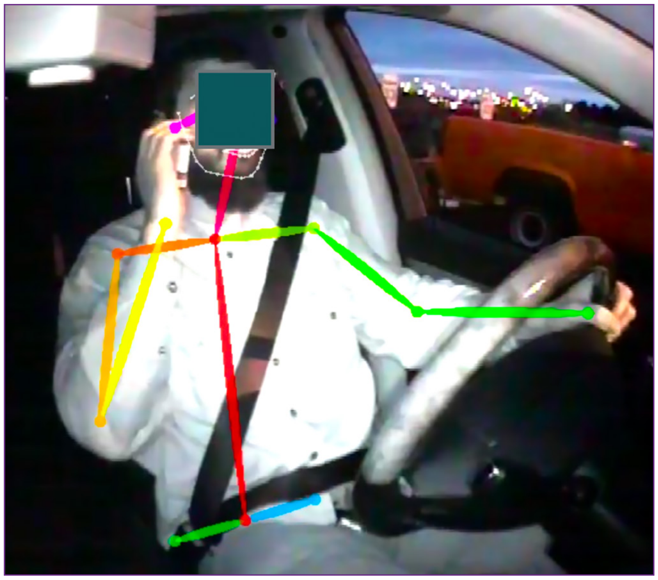


Figure 8. Photo. VDHA helps identify secondary behaviors such as using a cellphone.

More work is needed to improve performance, but these results indicate that cellphones can be detected in many driving situations in naturalistic data.

National crash statistics on eating while driving are not reported; however, eating while driving involves a combination of one or more distractions, such as unwrapping food packaging and holding the food with one hand while driving. A combination of the methods used in this study can provide a detailed understanding of eating behavior while driving, what the driver was eating and for how long, how long the driver's hands were off the steering wheel, and other factors. The secondary behavior algorithm the researchers used to identify driver behavior inside the vehicle can be used to study the time sequence when a driver is eating.[29] In-vehicle object detection can detect food and bottles in many driving situations, and body pose estimations can help researchers understand the driver's control of the vehicle.

In 2022, an estimated 96,000 crashes occurred in work zones, resulting in 37,00 injuries and 891 fatalities.[30] Of these fatalities, 136 involved a pedestrian or construction worker. Researchers need better information about the reasons for these crashes to develop effective countermeasures to reduce work zone crashes. The researchers in this study used a whole-image classification system and a segmentation-based system to automatically detect work zones using forward looking video cameras. The ability to detect work zones with these systems was relatively high, but combining these approaches may achieve even better results. In addition, human factors researchers can use the gaze detection algorithm from this study to examine driver distraction patterns. The gaze fixation and gaze saliency-based methods can also be used to understand drivers' work zone perception and their focus on certain roadway elements. Lastly, kinematics data can show how drivers navigate a work zone by changing lanes, slowing down, or other measures.

Research on the correlation between billboards and increased crash risk has been inconclusive. One problem is that billboards vary in how they capture a driver's attention. Oviedo-Trespalacios et al. determined that changeable billboards attract more attention than static billboards, and billboards closest to the road attract the most attention.(31) The billboard's content also affects the driver's attention.

The researchers in this study detected outside objects, including billboards, in a scene using instance segmentation and object detection methods. In addition, the researchers determined the driver's gaze behavior by processing driver-facing videos from naturalistic driving datasets. The gaze saliency and gaze fixation methods the researchers developed can be used to determine how long a driver looks at a billboard. Together, these tools can help human factors researchers study the effects of billboards on transportation safety.

CONCLUSION

This study aimed to assess the capability of automated systems to analyze naturalistic driving videos and produce annotations and descriptors for events, behavior, and driving scenarios related to transportation safety. The research team adapted and developed CV and ML techniques to automatically produce annotations that will allow investigators to explore interactions between drivers and other road users, road infrastructure, and roadside objects. Automatically producing annotations reduces the number of costly and labor-intensive manual annotations needed for large datasets such as the SHRP2 NDS.(1)

The researchers concluded that CV methods can perform several automated tasks, including detecting passengers and objects, estimating the driver's head and body pose, detecting driver actions related to safety, and estimating gaze direction inside the vehicle. Outside the vehicle, CV methods also detected work zones and intersections. Objects outside the vehicle were detected and associated with the driver's gaze direction. When analyzed over time, event correlation models can be developed that reflect patterns in a driver's attention and actions relative to driving situations.

This project showed that CV-based approaches can efficiently process continuous video data to identify and retrieve key annotations, such as in-vehicle object and passenger detection, secondary behaviors, and more. The algorithm the researchers developed can address 37 annotations from the SHRP2 reduction dictionary. The researchers developed new models to address the problems of action detection and behavior classification in the context of driving. This work is the first to explicitly train an ML model to detect these behaviors.

The researchers also adapted CV models to automatically determine external context, such as traffic density and the presence of work zones. The project resulted in multiple open-source codebases and CV models for the SHRP2 NDS.(1)

This study is the first to consider the development of a correlation model that relates driver behavior with automatically detected salient objects (e.g., billboards) and events (e.g., car cutting into a driver's lane). The correlations between driver gaze fixation and salient objects determined by the research have the potential to reveal how drivers behave in various traffic situations and aid the development of ADAS systems that help prevent accidents in highly automated vehicles.

Human factors researchers could use the video analytics methods from this project to approach automotive safety research in new ways.

REFERENCES

1. Virginia Tech Transportation Institute. 2020. “InSight Data Access Website: SHRP2 Naturalistic Driving Study” (website). https://insight.shrp2nds.us/, last accessed April 29, 2024.

2. National Highway Traffic Safety Administration. 2024. *Distracted Driving in 2022*. Publication No. DOT HS 813 559. Washington, DC: National Highway Traffic Safety Administration. https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813559, last accessed April 16, 2024.

3. FHWA. n.d. “Video Analytics for Automatic Annotation of Driver Behavior and Driving Situations in Naturalistic Driving Data” (web page). https://highways.dot.gov/research/projects/video-analytics-automatic-annotation-driver-behavior-driving-situations-naturalistic-driving-data, last accessed April 29, 2024.

4. FHWA. 2022. *The Role of Artificial Intelligence and Machine Learning in Federally Supported Surface Transportation 2022 Updates*. Report No. FHWA-HRT-22-026. Washington, DC: Federal Highway Administration. https://highways.dot.gov/sites/fhwa.dot.gov/files/FHWA-HRT-22-026.pdf, last accessed May 31, 2024.

5. Sarkar, A., I. Papakis, A. Svedovidov, J. Hickman, and A. L. Abbott. 2020. *Annotation of Blurred Cabin Imagery for Passenger Detection in SHRP2 NDS Data* (dataset). Version 1. https://doi.org/10.15787/VTT1/WS8ORW, last accessed May 31, 2024.

6. Krothapalli, U., L. Stowe, Z. R. Doerzaph, and A. Petersen. 2018. *Application of Proximity Sensors to In-vehicle Data Acquisition Systems*. Report No. 18-UT-066. Blacksburg, VA: National Surface Transportation Safety Center for Excellence.

7. HRNet. n.d. “HRNet” (HRNet software and configuration files in GitHub repository). https://github.com/HRNet, last accessed May 21, 2024.

8. National Highway Traffic Safety Administration. n.d. “Distracted Driving” (web page). https://www.nhtsa.gov/risky-driving/distracted-driving, last accessed May 21, 2024.

9. Ferrell, R., D. Aykac, T. Karnowski, and N. Srinivas. 2021. “A Publicly Available, Annotated Dataset for Naturalistic Driving Study and Computer Vision Algorithm Development.” Report No. ORNL/TM-2019/1106. Oak Ridge, TN: Oak Ridge National Laboratory.

10. VTTI. n.d. “head-pose-detection” (head pose model and codebase files in GitHub repository). https://github.com/VTTI/head-pose-detection, last accessed May 31, 2024.

11. VTTI. n.d. “object-detection” (configuration files and trained models in GitHub repository). https://github.com/VTTI/object-detection, last accessed May 31, 2024.

12. VTTI. n.d. “occupant_detection_classification” (occupant detection model and code files in GitHub repository). https://github.com/VTTI/occupant_detection_classification, last accessed June 25, 2024.

13. Papakis, I., A. Sarkar, A. Svetovidov, J. S. Hickman, and A. L. Abbott. 2021. “Convolutional Neural Network-Based In-Vehicle Occupant Detection and Classification Method Using Second Strategic Highway Research Program Cabin Images.” *Transportation Research Record* 2675, no. 8: 443–457. https://journals.sagepub.com/doi/10.1177/0361198121998698, last accessed June 25, 2024.

14. CityScapes Dataset. 2024. “CityScapes Dataset” (web page). https://www.cityscapes-dataset.com/, last accessed May 20, 2024.

15. Chang, B., M. D. Collins, Y. Zhu, T. Liu, T. S. Huang, H. Adam, and L-C. Chen. 2020. “Panoptic-DeepLab: A Simple, Strong, and Fast Baseline for Bottom-Up Panoptic Segmentation.” *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Washington, DC: IEEE Computer Society. pp. 12475–12485.

16. Sundharam, V., A. Sarkar, A. Svetovidov, J. S. Hickman, and A. L. Abbott. 2023. “Characterization, Detection, and Segmentation of Work-Zone Scenes from Naturalistic Driving Data.” *Transportation Research Record* 2677, no. 3: 490–504. https://doi.org/10.1177/03611981221115724, last accessed June 25, 2024.

17. Krum, A. J., R. J. Hanowski, A. Miller, A. Schaudt, and S. Soccolich. 2019. “FAST DASH: Program Overview and Key Findings From Initial Technology Evaluations.” *Accident Analysis & Prevention* 124: 113–119.

18. VTTI. n.d. “Segmentation-and-detection-of-work-zone-scenes” (codebases and models in GitHub repository). https://github.com/VTTI/Segmentation-and-detection-of-work-zone-scenes, last accessed May 31, 2024.

19. FHWA. 2023. “Intersection Safety” (web page). https://highways.dot.gov/research/research-programs/safety/intersection-safety, last accessed May 29, 2024.

20. VTTI. n.d. “Intersection-Detection” (codebases and models in GitHub repository). https://github.com/VTTI/Intersection-Detection, last accessed May 31, 2024.

21. Multiple Object Tracking Benchmark. n.d. “Welcome to MOTChallenge: The Multiple Object Tracking Benchmark!” (website). https://motchallenge.net/, last accessed June 25, 2024.

22. Kellnohofer, P., A. Racasens, S. Stent, W. Matusik, and A. Torralba. 2019. “Gaze360: Physically Unconstrained Gaze Estimation in the Wild.” *2019 IEEE/CVF International Conference on Computer Vision*, pp. 6912–6921. New York, NY: IEEE.

23. Cornia, M., L. Baraldi, G. Serra, and R. Cucchiara. 2018. “Predicting Human Eye Fixations Via an LSTM-Based Saliency Attentive Model.” *IEEE Transactions on Image Processing* 27, no. 10: 5142–5154.

24. Klauer, C., T. A. Dingus, V. L. Neale, J. D. Sudweeks, and D. J. Ramsey. 2006. *The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data*. Report No. DOT HS 810 594. Washington, DC: National Highway Traffic Safety Administration.

25. Sudweeks, J., A. Sarkar, J. P. Plummer, J. McClafferty, M. A. Perez, J. Hankey. 2016. *Mask Head Pose Validation Study Dataset*. Blacksburg, VA: Virginia Tech Transportation Institute. https://doi.org/10.15787/VTT1/DAFUH5, last accessed May 28, 2024.

26. Bhagat, H., S. Jain, L. Abbott, A. Sonth, and A. Sarkar. 2023. “Driver Gaze Fixation and Pattern Analysis in Safety Critical Events.” *2023 IEEE Intelligent Vehicles Symposium (IV)* Piscataway, NJ: IEEE, 1–8.

27. VTTI. n.d. “Event-Correlation” (codebases and models in GitHub repository). https://github.com/VTTI/Event-Correlation, last accessed May 31, 2024.

28. VTTI. n.d. “gaze-fixation-and-object-saliency” (codebases and models in GitHub repository). https://github.com/VTTI/gaze-fixation-and-object-saliency, last accessed May 31, 2024.

29. VTTI. n.d. “Virginia Tech Transportation Institute Repositories” (GitHub repository). https://github.com/orgs/VTTI/repositories, last accessed May 31, 2024.

30. American Road & Transportation Builders Association. 2024. “National Work Zone Safety Information Clearinghouse” (web page). https://workzonesafety.org/work-zone-data/#national, last accessed May 20, 2024.

31. Oviedo-Trespalacios, O., V. Trulove, B. Watson, and J. A. Hinton. 2019. “The Impact of Road Advertising Signs on Driver Behaviour and Implications for Road Safety: A Critical Systematic Review.” *Transportation Research Part A* 122: 85–98. https://www.sciencedirect.com/science/article/pii/S0965856418310632, last accessed May 21, 2024.

Getting Involved With the EAR Program

To take advantage of a broad variety of scientific and engineering discoveries, the EAR Program involves both traditional stakeholders (State department of transportation researchers, University Transportation Center researchers, and Transportation Research Board committee and panel members) and nontraditional stakeholders (investigators from private industry, related disciplines in academia, and research programs in other countries) throughout the research process.

Learn More

For more information, see the EAR Program website at https://highways.dot.gov/research/research-programs/exploratory-advanced-research/exploratory-advanced-research-overview. The site features information on research solicitations, updates on ongoing research, links to published materials, summaries of past EAR Program events, and details on upcoming events.

EAR Program Results

As a proponent of applying ideas across traditional research fields to stimulate new problem-solving approaches, the EAR Program strives to develop partnerships with the public and private sector. The program bridges basic research (e.g., academic work funded by National Science Foundation grants) and applied research (e.g., studies funded by State DOTs). In addition to sponsoring projects that advance the development of highway infrastructure and operations, the EAR Program is committed to promoting cross-fertilization with other technical fields, furthering promising lines of research and deepening vital research capacity.

EXPLORATORY ADVANCED **RESEARCH**

U.S. Department of Transportation
Federal Highway Administration

Recommended citation: Federal Highway Administration, *Using Video Analytics to Automatically Annotate Driver Behavior and Context in Naturalistic Driving Data* (Washington, DC: 2024) https://doi.org/10.21949/1521750

FHWA-HRT-24-078
HRT-10/12-24(WEB)E