Exploratory Advanced Research Program

Video Analytics Research Projects
Notice

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The U.S. Government assumes no liability for the use of the information contained in this document. This report does not constitute a standard, specification, or regulation.

The U.S. Government does not endorse products or manufacturers. Trademarks or manufacturers’ names appear in this document only because they are considered essential to the objective of the document.

Quality Assurance Statement

The Federal Highway Administration (FHWA) provides high-quality information to serve Government, industry, and the public in a manner that promotes public understanding. Standards and policies are used to ensure and maximize the quality, objectivity, utility, and integrity of its information. FHWA periodically reviews quality issues and adjusts its programs and processes to ensure continuous quality improvement.

Cover photos show, left, hands on steering wheel segmentation using skin color, from Carnegie Mellon University’s research into an automated real-time system to analyze emotional states of the driver (© Carnegie Mellon University); and, right, an overview of the system that will extract driver behaviour features and recognize various actions from SRI International’s research into a comprehensive automatic coding system (© SRI International).
Contents

Introduction 2
Data Bottleneck 2
Funding Research 2
Research Goals 3
Initial Research 4
Machine Learning for Automated Analysis of Large Volumes of Highway Video 4
Automated Feature Extraction 6
DCode: A Comprehensive Automatic Coding System for Driver Behavior Analysis 6
DB-SAM: CMU Driver Behavioral Situational Awareness System 7
Quantifying Driver Distraction and Engagement Using Video Analytics 8
Automated Identity Masking 10
Automation of Video Feature Extraction for Road Safety-Automated Identity Masking 10
DMask: A Reliable Identity Masking System for Driver Safety Video Data 11
Benchmarking Research Progress 12
Getting Involved with the EAR Program inside back cover
Learn More inside back cover
EAR Program Results back cover
TRANSPORTATION RESEARCH BOARD'S SECOND STRATEGIC HIGHWAY RESEARCH PROGRAM (SHRP 2) STUDY DEMONSTRATES THE IMMENSE SCALE OF DATA BEING GATHERED. THE FLAGSHIP DATASET TO EMERGE FROM THE SHRP 2 STUDY IS THE NATURALISTIC DRIVING STUDY (NDS). THIS REPRESENTS THE LARGEST EVER STUDY DESIGNED TO BETTER UNDERSTAND THE INTERACTION AMONG VARIOUS FACTORS INVOLVED IN HIGHWAY CRASHES—DRIVER, VEHICLE, AND INFRASTRUCTURE—SO THAT BETTER SAFETY COUNTERMEASURES CAN BE DEVELOPED AND APPLIED TO SAVE LIVES. THE STUDY IS A FOCUS ON DRIVER BEHAVIOR AND ADDRESSES THE NOTION THAT IT IS POSSIBLE TO OBTAIN MORE AND BETTER INFORMATION ON WHAT PEOPLE DO WHEN THEY DRIVE—NOT JUST IN THE MOMENTS BEFORE THEY GET INTO A COLLISION BUT ON A DAY-TO-DAY BASIS. NDS HELPS DELIVER THE SOLID BASELINE REFERENCE THAT IS REQUIRED TO PROVIDE CONTEXT TO ASSESS WHAT IS HIGH-RISK BEHAVIOR AND WHAT IS CONSIDERED NORMAL.

DATA BOTTLENECK

Researchers for the NDS have gathered over 1.2-million hours of data, collected from the vehicles of approximately 3,000 volunteers going about their regular activities. Each of those vehicles was equipped with four cameras, a Global Positioning System (GPS), and many other sensors. Over 2 petabytes (2,000 terabytes) of data has been generated over a 2-year period, a majority of which comes from video captured by the onboard cameras. The massive size of the video data creates a serious data bottleneck for researchers and makes traditional methods for identifying features in the data, such as objects, behaviors, roadside design details, and surrounding vehicles, completely inadequate. The traditional way to extract features of interest from video is for a researcher to physically sit in front of a monitor and manually log the location in the video data where each feature of interest is found; however, it is estimated that it would take almost 600 technicians a full year, working 40 hours per week, to go through all the video in the NDS.

FUNDING RESEARCH

Several video analytics research projects, funded by the EAR Program, aim to make data processing and analysis more practical by focusing on increasing the automation of video data decoding. The goal of these EAR Program projects is to assist researchers through the use of an algorithm designed to narrow down the amount of data researchers need to manually review. This will enable researchers to quickly and accurately extract the required details from large data files.

There are six ongoing video analytics projects funded by the EAR Program, all tasked with developing technology to make it feasible for a researcher to quickly and flexibly get what they need from an extremely large dataset. The first video analytics research project was awarded to Carnegie Mellon University’s National Robotics Engineering Center to investigate machine learning to teach an algorithm how to find and classify features of interest in the forward facing video data.
Learnings from this first award were examined and used to subsequently award Carnegie Mellon University, SRI International, and University of Wisconsin–Madison, three additional research projects to investigate how to automate the identification and classification of driver behavior, but focused on events inside the vehicle.

FHWA awarded two further EAR Program projects to Carnegie Mellon University and SRI International to focus on automated driver identity masking. The NDS dataset includes personally identifiable information related to the drivers, including video looking right at their faces and GPS locations of each trip taken. Protection of the privacy of the volunteers means that if this sensitive personal information is to be made available to the researcher, the researcher will have to travel to a secure enclave, which is equipped to prevent any unauthorized viewing or recording of the volunteers’ faces. These two research teams are working on algorithms that can automatically find faces in the video data and obscure the actual face in a non-reversible way, while still retaining certain points on the face that allow the researcher to identify and track expression. If this effort is successful, many researchers will be freed from the significant burden that use of the sensitive personal information would entail.

**Research Goals**

The Government’s goals are both short and long term. In the short term, the Government wants to extract value from the NDS data. In the long term, the Government wants to ensure that the data being collected will improve transportation safety to the maximum possible extent. The EAR Program addresses the need for longer term, higher risk research with the potential for long-term improvements to transportation systems—improvements in planning, building, renewing, and operating safe, congestion-free, and environmentally sound transportation facilities. The EAR Program seeks to leverage advances in science and engineering that could lead to breakthroughs for critical current and emerging issues in highway transportation—where there is a community of experts from different disciplines who likely have the talent and interest in researching solutions and who likely would not do so without EAR Program funding.

The following pages contain summary descriptions of the EAR Program-sponsored research projects investigating video analytics.
The first video analytics research contract was awarded by the EAR Program to Carnegie Mellon University in October 2012. The following summary provides an overview of this research activity.

**PROJECT:** Machine Learning for Automated Analysis of Large Volumes of Highway Video  
**INSTITUTION:** Carnegie Mellon University  
**PERIOD:** October 2012–September 2015  
**OBJECTIVE:** To develop a prototype tool that enables machine-learning algorithms to extract useful roadway features from video.  
**CONTACT:** Lincoln Cobb, Office of Safety Research and Development

**SUMMARY:** The SHRP 2 NDS is an unprecedented, extremely large dataset which includes over 1-million hours of video collected from vehicles traveling on U.S. highways. The NDS is being used for a wide range of research activities. However, as stated in the previous pages, such massive video datasets make data management and access a time-consuming and expensive challenge. The objective of this research project is to develop a highly usable prototype software tool that integrates and develops computer vision and machine-learning algorithms to automatically extract useful features from forward-looking video cameras mounted on volunteers’ vehicles.

A machine-learning-based approach enables researchers to harness the power of massive, diverse, and complex datasets to improve the performance of the feature extraction algorithms, instead of being overwhelmed by the quantity of data. Such a tool will make it feasible for a researcher to efficiently extract exactly what they need from a large video dataset. Though the primary application of the results of this work is the SHRP 2 NDS, the research team focused much of their work on the Roadway Information Database (RID), the second element of the SHRP 2 Safety program. The RID includes forward video collected by specialized data collection vehicles on some of the roads travelled by the NDS volunteers. That video information, and other data collected by the video van, is analyzed to generate a great deal of geometric and asset inventory data, including curve dimensions, grade, and cross slope. In addition, the RID includes information provided by local transportation agencies, such as the location and duration of work zones and weather information. The research team also worked with the SHRP 2 NDS forward-looking video streams; however, the lower-resolution NDS dataset represents significant challenges when attempting to identify certain features within a scene.

The researchers for this project integrate a range of previously developed analysis techniques with traditional object detection methods. They combined several state-of-the-art machine-learning algorithms to address a range of detection requirements; for example, a segmentation algorithm was used to help understand the context of a scene and a separate vehicle detector algorithm was used to detect cars and trucks. Learning algorithms can be trained on very large datasets and are capable of learning with high accuracy exactly which features are important, which provide less useful data, and how features can be used in combination to accurately recognize and classify the desired targets. This means that the great diversity provided by large and complex datasets actually adds to the power and robust performance of the system. The researchers developed a simple graphical user interface (GUI) to wrap up all of the individual analysis tools and enable even a novice user to quickly label data, train the detectors, and evaluate the performance of newly trained or existing detectors.

The research team evaluated multiple solutions to confront a range of challenges, including how to handle overlapping objects, as shown in figure 1; identify different roadway features (e.g., car and truck detection); understand a scene and distinguish between vegetation, sky, and highway shoulder; detect traffic signs; and detect and estimate the status of traffic lights. Efforts to overcome these challenges led to the development and refinement of several innovative components, including the development of a new traffic light detection and signal estimation system focused on NDS data, and a new configurable traffic sign detection system. Moving forward, the research team has proposed a hybrid approach to supplement the NDS data with higher-quality data, such as from the RID. In accordance, the research team recommended using NDS data to detect dynamic information, such as traffic light state, vehicles, weather, construction zones, and vehicle–lane positioning, and then using higher-quality datasets from the same location to detect static information, such as roadway characteristics, signs, and traffic light location.
This research project is directed at the development of tools to automate the initial coding of this video data—that is, the identification and classification of information of interest to safety researchers. This toolset relies on advanced machine-learning principles, including a strong reliance on context, to allow the analyst to find the features needed to answer particular research questions. It is based on the concept that the overall video data are not as useful as understanding the actual scene content. For example, if a researcher is interested in driver behavior at an intersection, then all the non-intersection video content is irrelevant. Identifying images of interest based on the task at hand, and automatically identifying the scene content, has the benefit of significantly reducing costs and enables researchers to do more with the data. Automating the identification of interesting images and automatically extracting the desired scene content is therefore a crucial component that is necessary to fully leverage the power that such large datasets have to offer transportation researchers.

For more information on this project, visit www.fhwa.dot.gov/research/tfhrcc/projects/projectsdb/projectdetails.cfm?projectid=FHWA-PROJ-12-0070.

**IMPACT:** This research will provide an improved understanding of how computer vision techniques could be better used to support transportation research. The prototype system demonstrates the effectiveness of advanced machine-learning techniques applied to large, diverse datasets. This research has a particular focus on techniques which may compensate for poor quality data with poor resolution, and lays the groundwork for future development of a comprehensive library of data processing and analysis tools. Fully understanding these data will enable researchers to establish meaningful answers to a broad range of challenging questions, without the need to perform expensive data collection. This research is expected to open further avenues for active learning for traffic and congestion control measures, driver behavior training, and road safety mechanisms.

Figure 1. A typical challenging situation for overlapping objects (left vehicles) and occluded objects (right vehicle) in an urban environment.
Automated Feature Extraction

In 2014, the EAR Program sponsored three additional research awards. The awards were made to SRI International, Carnegie Mellon University, and the University of Wisconsin–Madison. These projects reflected lessons learned from the initial research project conducted by Carnegie Mellon University, “Machine Learning for Automated Analysis of Large Volumes of Highway Video.” The objective of these new research projects is to advance efficient and cost-effective methods and tools to analyze the large amounts of video-related safety data generated by studies such as the SHRP 2 NDS, but focused particularly on the driver and the interior of vehicles involved. Each project brings different technical strengths and approaches, matching computer science with highway engineering and applying the approaches to different features of interest for highway safety. Such tools will reduce the time and costs required to pull needed features from these rich datasets, and thus expand the pool of researchers able and willing to use data, such as the NDS, in their research. In time, as the safety community’s understanding of the relationships between driver behavior and safety increases, these EAR Program-sponsored research projects could enable accelerated development of safety improvements in many areas of transportation.

The following pages provide a brief summary of each of these projects.

**PROJECT:** DCode: A Comprehensive Automatic Coding System for Driver Behavior Analysis  
**INSTITUTION:** SRI International  
**PERIOD:** March 2014–March 2016  
**OBJECTIVE:** To develop a comprehensive automatic coding system to assist in the coding of features in the SHRP 2 NDS data relevant to safety researchers.  
**CONTACT:** Lincoln Cobb, Office of Safety Research and Development

**SUMMARY:** Understanding the interaction among various factors involved in highway crashes is one of the key goals of the SHRP 2 NDS. Improved understanding in this area will ultimately permit the development of better safety countermeasures. The volume of data gathered in this study creates a need for a system that can automatically annotate the data in an accurate and expedient manner. SRI International is developing a comprehensive automatic coding system, known as DCode, to assist in the coding of features in the video data relevant to safety researchers interested in using the SHRP 2 NDS data.

The system will extract driver behavior features, including head pose, gaze direction, eye blinks, mouth movements, facial expressions, upper-body and hand positions, as well as recognize various actions (e.g., using a cell phone) and gestures made by the driver, as shown in figure 2. The system will also track features from the driving environment, including passengers in the vehicle, pedestrians, vehicles, and traffic and weather conditions outside the vehicle.

According to the research team, a comprehensive driving behavior study must take into account not only the actions and behaviors of the driver but also the context in which those actions are performed. The term context refers here to factors outside the vehicle, such as weather, road, and traffic conditions, in addition to actions and signals from nearby vehicles, traffic lights, and road signs. It also

Figure 2. The DCode System will extract driver behavior features and recognize various actions. It will also track contextual features.
refers to factors inside the vehicle, such as passengers and any distractions they may cause, vehicular alarms, satellite navigation and other gadget-induced distractions, and various other objects the driver may interact with. The system under development uses existing software and algorithms to track features on the driver’s body, including the head, face, and hands, as well as features from outside the vehicle that are relevant to safety research, such as pedestrians, vehicles, vehicle signals, and weather conditions.

The research team aims to (1) retrain current algorithms with data that better matches the SHRP 2 NDS data and also explore a suite of novel feature extraction and tracking approaches better suited to this data; (2) extract features related to the driver’s behavior, as well as to the driver’s environment inside and outside the vehicle, to provide valuable contextual features; (3) develop a multitiered feature extraction pipeline where the core layer will track all directly observable features, such as the head pose, facial features, upper-body and hand positions, and pedestrian and vehicle locations—while the upper layers will use these features to identify various actions and gestures as well as monitor the driver’s state based on machine-learning techniques; and (4) develop algorithms in a manner that makes them scalable so that it may be run on a distributed computational architecture. Application programming interfaces and data interfaces for feature extraction will also be developed by the research team that are amenable to remote Web-based access and processing of large volumes of data without any humans in the loop.

PROJECT: DB-SAM: CMU Driver Behavioral Situational Awareness System

INSTITUTION: Carnegie Mellon University

PERIOD: March 2014–March 2016

OBJECTIVE: To develop an automated real-time system to analyze emotional states of the driver to determine if he or she is fatigued or distracted.

CONTACT: Craig Thor, Office of Safety Research and Development

SUMMARY: Driver safety is a major concern in today’s world and the ability to use image processing, computer vision, and machine-learning algorithms to automatically assess whether a driver is in good condition to drive is of extreme importance. The research team for this project will develop an automated real-time system to analyze the emotional state of the driver to determine if he or she is tired or starting to fall asleep, distracted by a conversation, or using a handheld device. The system will also continuously estimate the head pose of a driver to see if they are paying attention to the road and surroundings. The research team’s previous experience in detecting and tracking faces in a crowd in complex environments is highly relevant to this project. This is because a similar challenging scenario arises when a person drives through various scenes with difficult lighting conditions and visual noise.

For this project, the research team plans to delineate regions of interest on the face that need to be analyzed. To realize the goal of a fully automated real-time driver safety analysis system, the research team is using a landmarking tool, known as a modified active shape model, to quickly and reliably provide accurate landmarks irrespective of pose variations and partial occlusions of the face. The research team is also using a fast and robust three-dimensional (3-D)-based reconstruction technique that uses a single view instead of multiple views, known as a 3-D generic elastic model. Texture-based classifiers will be built to determine if a driver’s mouth is open or closed and subsequently used to indicate if the driver is talking. Likewise, the eyes will be tracked to determine if they are open, closed, or partly closed, and used to indicate driver fatigue (figure 3).

Instead of pursuing methods to determine the emotional state of a driver based on facial expressions, this research is focused on the fundamental question of whether a driver is in a condition to safely drive the vehicle. The research team claims this question can be best answered by simply monitoring the eyes and lips for signs of fatigue or distraction. Segmentation-based algorithms will also be used to check if a driver’s hands are on the steering wheel, as shown in figure 4, and compressed sensing-based algorithms will be used to determine whether a driver is talking on a cellphone held next to their ear. To detect if a driver is wearing a seatbelt, a specific region in the image will be located
for the presence of seat belts. The image can be automatically processed and a linear regression-based model applied to detect the presence of a seatbelt. In addition, the ability to detect soft biometric information could eventually be used to improve and tailor behavioral advice; for example, identifying whether a driver is over 18 years, or if a driver is wearing sunglasses when driving at night.

Figure 4. Hands on steering wheel segmentation using skin color.
used as a binary image to determine if the hand is on the wheel or not. In addition, the software is able to detect eye and mouth movement and then use a driver’s gaze pattern to reveal their state of mind during driving; however, this is not a straightforward process because staring straight ahead may not necessarily imply engagement in driving, just as looking right and left over the shoulder could be a sign of alertness rather than distraction. To truly capture a driver’s gaze status, the research team is focusing on whether the eyes are closed for a prolonged period of time and observing mouth movement to focus on conversations that last more than 5–10 seconds. Talking may also involve head movements and hands off the wheel, which can also be measured as further evidence of driver distraction. The ability to detect the location of the vehicle offers valuable contextual information to infer a driver’s behavior. Including information on driving speed and GPS data would enable researchers to determine whether a vehicle is stopped at a traffic light or a stop sign and see if a driver swings the head to look left and right when the vehicle stops.

The final phase of this research includes a quality assessment to validate the accuracy of the extracted features as extraction proceeds. Inconsistencies will be annotated and made available for human operator review. The software will include a fully automatic feature-coding mode and an interactive, human-assisted mode. The interactive mode will consist of visualization of the extracted features, and a GUI that will access a database consisting of the raw SHRP 2 data and extracted features. Users will be able to review the raw SHRP 2 composite video, together with a graphic animation of the extracted features, as well as a display of the multimodal driver states, environmental variables, and driving conditions.

Figure 5. Regions of interest, including the driver’s face, hands, and front view, are monitored by the system.
Automated Identity Masking

The EAR Program supported two additional research awards, one each to Carnegie Mellon University and to SRI International. These two institutions were charged with investigating automated identity masking and both are working on individual approaches to ultimately make it easier for a wide range of researchers to preserve privacy issues while making use of the NDS and related datasets. Development of facial masking that provides underlying information about the driver (e.g., head pose, mouth, and eye movement) while precluding personal identification, would allow for increased access and accelerated conduct of research using naturalistic driving data.

The following pages provide a brief summary of these two projects.

PROJECT: Automation of Video Feature Extraction for Road Safety—Automated Identity Masking

INSTITUTION: Carnegie Mellon University and University of Pittsburgh

PERIOD: February 2014–April 2015

OBJECTIVE: To develop an automated facial masking technique to de-identify face images.

CONTACT: James Pol, Office of Safety Research and Development

SUMMARY: The researchers for this project are developing an automated facial-masking technique to de-identify face images while preserving the facial behaviors of the drivers. The facial de-identification process will be non-reversible so that the driver identity cannot be re-established once masked. At the core of the research is a new concept referred to as facial action transfer (FAT), which clones the facial actions from the video of one person to another person, as shown in figure 6.

FAT represents several innovations when compared to traditional image distortion methods. The concept replaces person-specific facial features (i.e., identity information) of the subject to be protected with those of the target. It also preserves facial actions by generating video-realistic facial shape and appearance changes on the target face. The photo-realistic and video-realistic de-identified video preserves spontaneous and subtle facial movements, while de-identifying the identity of the driver.

The software uses only one picture of the mask face to produce realistic videos where the facial appearances of the driver are replaced by the mask face. The research team has already built a fully automatic GUI to track the facial features in the video and detect facial landmarks in the mask face. The GUI is then able to output videos in which the driver face is de-identified. The research team is now exploring algorithms designed to smooth the de-identified facial actions over time to prevent issues occurring when there are sudden changes of ambient light or movement of the driver’s head.

Figure 6. The personalized facial action transfer process.
PROJECT: DMask: A Reliable Identity Masking System for Driver Safety Video Data  

INSTITUTION: SRI International  

PERIOD: February 2014–February 2016  

OBJECTIVE: To develop an automated, complete, and irreversible identity masking system for processing realistic, low-resolution video in the SHRP 2 dataset.  

CONTACT: Aladdin Barkawi, Office of Safety Research and Development  

SUMMARY: The research team for this project is using innovative technologies to replace the driver’s head in the full SHRP 2 dataset with the head of a computer-generated avatar. The system will preserve information about the driver’s head pose, facial expressions, the state of their eyes and mouth, and the direction of their eye gaze.  

This project uses a four-layered processing framework to detect and track the driver’s head and facial features, interpolate the driver’s head position in frames that the tracker misses, replace the head with a rendered avatar in all frames to mask the driver’s identity, and evaluate the confidence of this identity masking (figure 7).  

In addition, the research team will develop an approach to synthesize facial motions on the computer generated avatar that closely tracks the driver’s facial motions. This enables researchers working with the identity-masked videos to still have access to behaviorally relevant cues on the driver’s face. The research team will also develop a method to detect (and not mask) non-facial elements obscuring the face (e.g., the driver’s hand), so that information loss is minimized in the videos.  

The final step in the process is the development of a GUI tool that displays the face-masked version along with a measure of confidence of masking over the entire video. This will allow a human user to quickly view the lowest confidence frames in the masked video and, with very little effort, re-annotate some key facial points manually in frames where the tracking and or masking was incorrect, thus allowing the system to recover from those mistakes.
FHWA is also working with the Department of Energy’s Oak Ridge National Laboratory to develop tools for benchmarking research progress. A brief summary of this collaboration is included below.

**INSTITUTION:** Oak Ridge National Laboratory  
**PERIOD:** June 2013–May 2017  
**OBJECTIVE:** To develop calibration and measurement techniques that will aid the broader community of researchers that want to work with NDS data.  
**CONTACT:** Lincoln Cobb, Office of Safety Research and Development

**SUMMARY:** FHWA is working with the Oak Ridge National Laboratory to develop calibration and measurement techniques that will enable benchmarking of research progress and technical assessment of EAR Program-sponsored research teams. Tasks include creating camera calibration models, testing baseline algorithms for proposed feature extraction and identity masking, creating partitions of data into training and validation sets, developing evaluation method for automated identify masking, and conducting computational load projections and data sharing architectures.
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 Web site at www.fhwa.dot.gov/advanced research/cvtr.cfm.
EAR Program Results

The EAR Program strives to develop partnerships with the public and private sectors because the very nature of exploratory advanced research is to apply ideas across traditional fields of research and stimulate new approaches to problem solving. The program bridges basic research (e.g., academic work funded by National Science Foundation grants) and applied research (e.g., studies funded by State departments of transportation). In addition to sponsoring exploratory advanced research 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 through dissemination and continued investigations, and deepening vital research capacity.