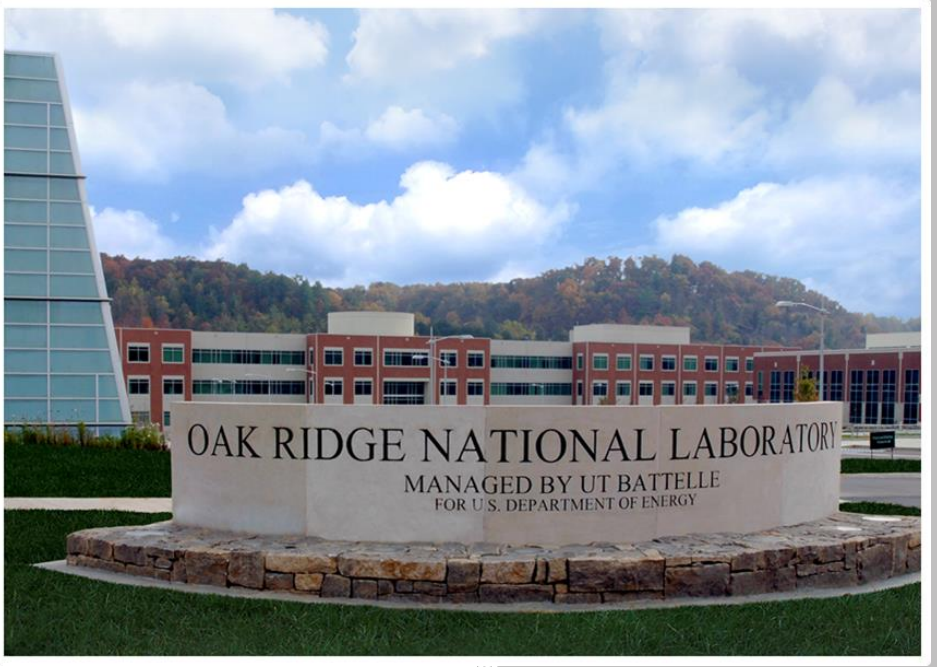
ORNL/TM-2019/1106

A Publicly Available, Annotated Data Set for Naturalistic Driving Study and Computer Vision Algorithm Development



Regina Ferrell
Deniz Aykac
Tom Karnowski
Nisha Srinivas

January 2021

OAK RIDGE NATIONAL LABORATORY
MANAGED BY UT-BATTELLE FOR THE US DEPARTMENT OF ENERGY

DOCUMENT AVAILABILITY

Reports produced after January 1, 1996, are generally available free via US Department of Energy (DOE) SciTech Connect.

Website www.osti.gov

Reports produced before January 1, 1996, may be purchased by members of the public from the following source:

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone 703-605-6000 (1-800-553-6847)
TDD 703-487-4639
Fax 703-605-6900
E-mail info@ntis.gov
Website http://classic.ntis.gov/

Reports are available to DOE employees, DOE contractors, Energy Technology Data Exchange representatives, and International Nuclear Information System representatives from the following source:

Office of Scientific and Technical Information
PO Box 62
Oak Ridge, TN 37831
Telephone 865-576-8401
Fax 865-576-5728
E-mail reports@osti.gov
Website http://www.osti.gov/contact.html

1. OBJECTIVE OF COLLECTION

Oak Ridge National Laboratory developed and implemented a data collection effort to create a dataset for use in evaluating and testing algorithms for analyzing driver behavior under controlled settings for support of the Federal Highway Administration's Exploratory Advanced Research Program. This collection is called the ORNL Naturalistic Driving Study Sample (ONDSS). The dataset is designed to emulate aspects of the Second Strategic Highway Research Project (SHRP2), which contained a massive naturalistic driving study (NDS) with over 3000 drivers between 2010 and 2013 using their personal vehicles, with over 4300 person-years of data collected [HANKEY]. However, due to privacy constraints the video data can only be obtained under strict constraints, and data from the face camera is not available outside secure enclaves at Virginia Tech Transportation Institute in Blacksburg, VA and Turner-Fairbank Research Center in McLean, VA.

The data was collected under Institutional Review Board approval with the understanding that the data will be published and freely available to interested parties. Several data collection goals were accomplished:

1. To create a controlled set of video data where driver head pose was measured by sensors during a set of prescribed motions.

2. To create a set of video data with drivers attempting to convey a particular set of emotions.

3. To create a set of videos with annotated and prescribed usage of a cell phone while simulating driving.

4. To create sets of videos where drivers perform secondary actions that are unrelated to the task of driving and could be indicative of a certain level of distraction. Examples include eating or drinking, dancing, and adjusting various controls and components of the car unrelated to driving such as rolling down the window, talking on the phone, etc.

5. To create a set of videos with the participant using a particular hand or hand combination to drive.

6. To create a set of videos with the participant removing or putting on his/her seat belt.

These activities were grouped into the overall video categories shown in Table 1. All activities were performed while participants were asked to pretend to drive.

Table 1. Explanation of sets of videos generated

Video Category	Description	Special Notes	Annotation State
HeadPose	Participant turns head to directly face designated places in the car relevant to various driving maneuvers.	Has accelerometer head pose data; attempts sequence five times	Complete
Check	Participant looks at designated places (same descriptions as in HeadPose but the participant is not required to turn the head to face those locations) in the car relevant to various driving maneuvers.	Has accelerometer head pose data; attempts sequence five times	Complete

Table 1. Explanation of sets of videos generated (continued)

Video Category	Description	Special Notes	Annotation State
ActionsShorter	Participant engages in a number of nondriving activities while pretending to drive including singing, blinking, yawning, grooming, etc.	Attempts sequence three times	Complete
GlassAndCell	Participant simulates using a cell phone in various modes including talking, texting, and web surfing.	Attempts sequence three times	Complete
Driving	Participant actually drives a short route.	Has accelerometer head pose data	None
Conversations	Participant engages in conversation with people inside and outside the vehicle	Attempts sequence three times	Complete
Eating	Participant picks up food and eats	Attempts sequence three times	Complete
Expressions	Participant simulates various emotional responses including anger, contempt, disgust, fear, happy, sadness, and surprise	Attempts sequence five times	Complete
Hands	Participant moves through a sequence of combinations of hands on steering wheel	Attempts sequence three times	Complete
HighlyEngaged	Participant rifles through a bag of items in passenger seat while pretending to drive	Attempts sequence three times	Complete
Seatbelt	Participant removes and rebuckles seatbelt	Attempts sequence three times	Complete
VisorAndWindow	Participant follows sequence of actions involving the visor and rolling down the window	Attempts sequence three times	Complete

2. EQUIPMENT AND SOFTWARE

All testing was done in the same 2004 Honda Civic. This vehicle was instrumented with a loaned set of camera and Data Acquisition System (DAS) used in the Naturalistic Driving studies performed by Virginia Tech Transportation Institute (VTTI) [INSIGHT]. The rear camera of the system was not mounted in the rear window as in the SHRP-2 and Head Pose Validation studies; instead, it was used to help monitor stages of the test. Figure 1 shows the mounting of the unit housing the front, face, and hands camera. This unit produces a single output video, as shown in Figure 2, which is a combination view from all the cameras. The separate views from each camera were extracted and rotated as needed and saved for the data set. In addition, another annotated view was generated combining the camera views with the expanded hand view and the orientation flipped as first performed by SRI International in their "the SRI-DCODE method of dual hand-face tracking." An example of an annotated video view can be seen in Figure 3.

In addition to the VTTI camera suite, a small GoPro camera was mounted on the dash directly in front of the driver's seat to acquire concurrent, higher resolution video of the same events. For three of the experiments—HeadPose, Check, and Driving—the driver wore a headpiece mounted with a VectorNav

VN-100 sensor [VNAV] that rested on the top of the head. This sensor was used to measure the yaw, pitch, and roll pose angles of the driver's head. This sensor was tared for a zero angle corresponding to the driver looking directly at the DAS camera mounted below the rearview mirror. An additional VectorNav VN-100 sensor was mounted in the center of the top of the rear seat to make corrections for measurements when both the car and the driver are in motion. These sensors were found to exhibit considerable drift over time. The data is available in its original form as well as a version where the sensor data was corrected for this drift. This data provides information on the changes in the yaw, pitch and roll of the head during the actions scripted for head pose, check and the short driving segment.

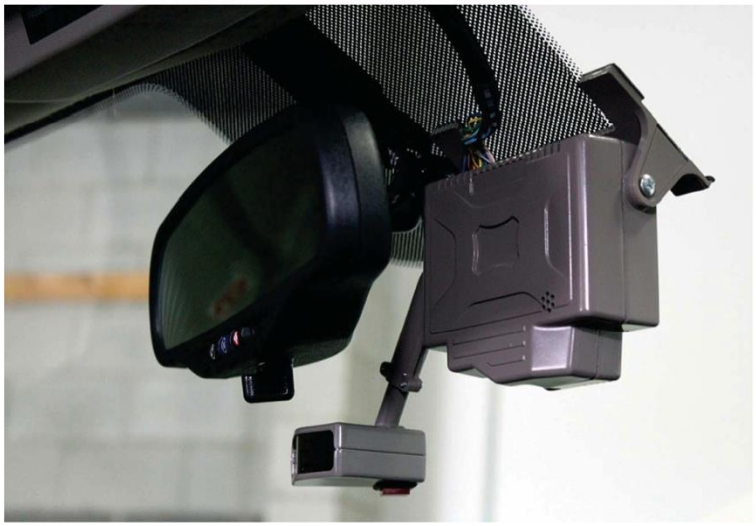


Figure 1. SHRP-2 camera unit mounted behind rearview mirror [CAMPBELL]

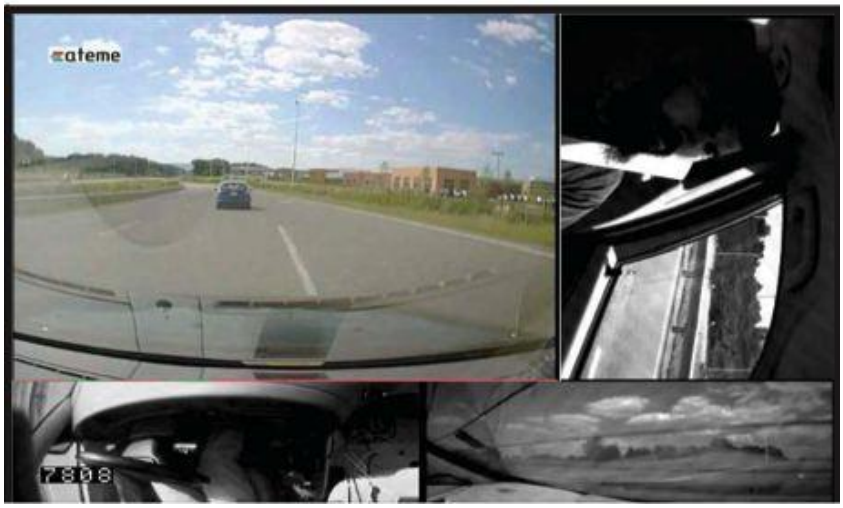


Figure 2. SHRP-2 composite view as recorded by DAS [CAMPBELL]

For most of the tests, the vehicle was parked under a tarp to protect the participants from sun and the tarp was periodically adjusted to reduce the impact of bright sunlight in the cab of the car.

Figure 3. Annotated video view.

3. STUDY DESIGN

With the exception of a brief driving study, all of the participant actions were performed in a parked car. For the longer experiments—HeadPose, Check, ActionsShorter, and GlassAndCell—the participants responded to recorded audio instructions. For the HeadPose and Check experiments, the instructions ran through a sequence of steps five times. For the other experiments, the instructions were completed three times for a series of actions. After conducting experiments with participants T002 and T003, the verbal recordings were amended to run slightly quicker, so the sequence and relative timing of those participants' actions might be slightly different from later trials, but this is reflected in the annotated data.

For each experiment, the following files were generated.

- Face video file—in proper orientation
- Rear video file
- Dash video file (like the hands file for VTTI data sets)
- Front video
- Combined video (SRI-DCODE method of dual hand-face tracking)
- Annotated video (combined video with addition of text showing current annotated action)

- A cropped video file that represents data from the GoPro camera situated directly in front of the driver

- An AlignedGoProDas video file that shows the combined camera views, annotations, and the cropped video from the GoPro camera superimposed on the part of the front image. A csv file showing the corresponding GoPro and DAS frames used for synchronizing this video is included. (It was impractical to modify circuitry to trigger the GoPro and DAS to start at the same time.)

The videos were later reviewed to ensure that the experiments' scripted actions were correctly executed and completed and that a consistent start and stop time for each annotated action was recorded. In particular, for HeadPose, labeled actions such as "P Side Window" were annotated from the time the participant begins to move to complete the action, through actually viewing the described locations or action and then returning to looking straight ahead and simulate driving. If the participant "missed" or got confused about the action to execute, a "-1" was set as the start and stop frame for the action.

4. ANNOTATIONS

The annotations provided with the data set are stored in a MATLAB .MAT format and in a csv format for each table within the .MAT file. The following is a list of tables and the type of information within:

TableName	Description of Contents
ORNLTrialTable	Lists participant identifiers, their gender, whether they typically wear glasses, and references to high resolution still photos of the participants.
ORNLFilesTable	Enumerates the video files included in the release, which video category or experiment the file relates to, and whether the video is from the Dash, Face, Front, or Rear camera or is a composite view. Also includes the number of frames in the video and the frame rate.
ORNLAnnotationsTable	Enumerates the participants, the video category and video source, the start and end DAS frame of the description (described action), and the INQSCRIBE time stamp values.[1] Does have -1 for DASStartFrame and DASEndFrame to indicate actions participants failed to execute properly.
ORNLHeadPoseTable	Provides head pose values from the VectorNav sensors on a frame-by-frame basis for each HeadPose and Check Video for each participant. Note that the VectorNav was tared to zero while facing the DAS camera, not the GoPro camera.
ORNLCorrectedHeadPose Table	If a sensor experienced significant drift over the course of the experiment the sensor's data has been modified to correct for the drift, otherwise it is structured just like the ORNLHeadPoseTable.
ORNLGoProAnnotations	Enumerates the participants, the video category and video source, and the start and end GoPro video (cropped) frame of the actions annotated for the DAS video. The GoPro's frame acquisition rate was roughly two times that of the DAS video, and the start and stop of the videos is not in sync.

5. AUTOGENERATED FACIAL ANNOTATIONS

Accurate location of the face or significant movements of its location can be important indicators of a driver's distraction or actions. To help provide some additional information of potential interest to researchers, an automated face detector was run on all the Face videos and all the GoPro videos. The

[1] INQSCRIBE is a video-annotation tool used to perform the annotation.

results of RetinaFace [DENG] presently one of the most accurate state of the art face detectors, is published on this dataset to be of potential use to researchers.

The annotations provided are in the form of a csv file for each video with each line of the file indicating a face detection and its landmarks. There is a field for the frame, which detected face (not tracked consistently – the detect_id=0 face in one frame may be the detect_id=1 face in another), the confidence score for the detection, the bounding box details, and 5 landmarks corresponding to the eyes, tip of nose, and the corners of the mouth. These annotations are not validated, but those with confidence scores of .95 and greater generally seem to be pretty accurate with this data. Confidence scores below .9 may often be false detections for this dataset.

6. ACKNOWLEDGEMENTS

We acknowledge the support of the Exploratory Advanced Research Program of the Federal Highway Administration. We also acknowledge the assistance of Virginia Tech Transportation Institute.

7. REFERENCES

[HANKEY] Hankey, Jonathan M., Miguel A. Perez, and Julie A. McClafferty. Description of the SHRP 2 naturalistic database and the crash, near-crash, and baseline data sets. Virginia Tech Transportation Institute, 2016.

[VNAV] VectorNav VN100 https://www.vectornav.com/products/VN-100

[CAMPBELL] Campbell, Kenneth L. "The SHRP 2 naturalistic driving study: Addressing driver performance and behavior in traffic safety." Tr News 282 (2012).

[DENG] Deng, Jiankang & Guo, Jia & Zhou, Yuxiang & Yu, Jinke & Kotsia, Irene & Zafeiriou, Stefanos. 2019. RetinaFace: Single-stage Dense Face Localisation in the Wild. https://arxiv.org/abs/1905.00641

[INSIGHT] Virginia Tech Transportation Institute. 2018. Insight Data Access Website. https://insight.shrp2nds.us/.