ORNL/TM-2022/2786

Database to Enable Facial Analysis for Driving Studies (DEFADS)

Thomas P Karnowski
Deniz Aykac
Regina Ferrell
Leanne Thompson
Lauren Torkelson

December 2022

OAK RIDGE National Laboratory

ORNL IS MANAGED BY UT-BATTELLE LLC FOR THE US DEPARTMENT OF ENERGY

ORNL/TM-2022/2786

Cyber Resilience and Intelligence Division

DATABASE TO ENABLE FACIAL ANALYSIS FOR DRIVING STUDIES (DEFADS)

Thomas P Karnowski
Deniz Aykac
Regina Ferrell
Leanne Thompson
Lauren Torkelson

December 2022

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, TN 37831
managed by
UT-BATTELLE LLC
for the
US DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725

CONTENTS

ABSTRACT

Naturalistic Driving Studies (NDS) collect and utilize data on drivers in real-world environments in instrumented vehicles. A common problem with such studies is driver privacy. In this work we collected a dataset of 77 human subjects performing scripted driving-related activities. We used three camera systems for the collection, including two high-resolution webcam devices as well as a third system from an actual NDS (the Second Strategic Highway Research Project, or SHRP2). This report covers the data collection process and summarizes the dataset, which will be made publicly available to researchers under a data usage license.

1. Motivation

Understanding how individuals drive in the 'real world' is a consistent problem with a variety of implications for transportation-related disciplines including energy, safety, human factors, economics, and urban dynamics. Naturalistic Driving Studies (NDS) are one method used to gain insight into driving behaviors. The largest to-date NDS, the Second Strategic Highway Research Project (SHRP2) was undertaken through the auspices of the National Academy of Sciences to gain insight into human driving behavior [SHRP2PROD]. The study was undertaken from 2006-2015 with data collected from roughly 2011-2013. Data consists of a variety of trip data including GPS and real-time sensor data, including video data of the driver's face. The video data allows a considerable amount of insight into driving behaviors, including distraction and fatigue, but due to the sensitive nature of the data and its context, the facial data is very difficult to share in open environments. Furthermore, the volume of data (roughly 3500 people for 2 years) creates a large problem in the costs of finding key data of interest to transportation researchers. Computer vision and machine learning offer opportunities to reduce the data content to more manageable levels, but even modern computer vision algorithms have issues with the relatively poor quality of the SHRP2 data due to compression artifacts, image sizes, and environmental "noise" (such as nonuniform lighting). Thus, there is room for improvement in analysis algorithms for the automated detection of driver features.

There have been efforts to create public training datasets for SHRP2 and other NDS data analysis including:

- ORNL developed the Oak Ridge Naturalistic Driving Study Sample (ONDSS) data set [ONDSS] in 2015. The dataset used 10 participants who performed a variety of activities including a brief driving sample. The dataset was recorded with a SHRP2 DAS as well as a high-resolution color camera (GoPRO Hero 2) as a reference. A VectorNAV Inertial Measurement Unit was utilized as well, mounted on the subjects' heads, to provide ground truth for some activities [VNAV]
- The Virginia Tech Transportation Institute (VTTI) developed the Head Pose Validation (HPV) set, using 48 drivers and a specially modified SHRP2 DAS to study the performance of an early computer-vision solution for head and face characterization. The dataset was collected under less stringent privacy constraints, but is still not openly publishable and is not freely available to researchers. [VTTI-HPV]
- VTTI also developed the "VTTI Training Set" which featured three VTTI employees driving three different vehicles with SHRP2 DAS systems. This dataset is freely available but has very limited ground truth associated with it. [VTTI-TS]

Such datasets are useful for a variety of studies related to SHRP2 analysis [PAONE,HOOGE, BARAGCHISADEH, MEDIAEVAL] but there is still a need for additional data. The Dataset to Enable Facial Analysis for Driving Studies (DEFADS) was created to help provide such additional data, particularly with a larger number of subjects. ORNL began planning the dataset in 2018 but was delayed due to COVID-19 restrictions. The data collection was completed in Fall 2021, with quality check and

data cleaning operations completed in Winter 2022. This document covers some details of the dataset with respect to its unique properties, usability, and general description of its characteristics.

2. Experimental Design

The goal of the DEFADS collection was to create a dataset using a SHRP2 DAS, focusing on facial characteristics and behavior queues, with a relatively large number of participants (targeting 100). The collections were made on the ORNL campus but were limited by COVID-19 restrictions which changed over time, and thus the protocols for the data collection tended to be restrictive as well to emphasize researcher and human subject safety. Due to the relatively large number of subjects targeted, the emphasis was on short collection sequences, and no driving was performed by the subjects; instead, scripted actions were used to simulate driving behaviors.

The data was acquired under IRB protocol ORAU000669. In the consent form, the subjects agreed to consent to share their data publicly; however, ORNL has elected to track usage through a simple data usage license as opposed to simply posting the data on a public website.

The data collection process proceeded by first collecting a portrait-style image of each subject. Then two in-vehicle sequences were conducted for each participant: an "Action" collection and a "HeadPose" collection. In the Action collection, the subject was asked to perform a series of actions that were emulative of driving behavior. In the HeadPose collection, the subject was asked to look at a particular landmark in the vehicle, such as the rearview mirror, console, etc. These positions were marked by tapping a large sign on each position of interest with a letter ranging from A to N to avoid confusion.

Periodically the subject was asked to look at the rearview mirror and "nod", which produced a consistent action across all the cameras to allow time synchronization (note that this action was removed from the final video set).

The actions and head pose locations are listed in Table 1 and Table 2.

Table 1 "Action" sequence activities

Action
Tired
TouchPassengerSeat
Scary Reaction
InteractWithRadio
WaveToLeft
WaveToRight
Dance
TalkToSelf

Table 2 "HeadPose" sequence activities

HeadPose
LeftSideMirror
LeftWindow
LeftRearWindow
LeftBack
RearView
RightFront
RightSideMirror
RightWindow
RightRearWindow
RightBack
Radio
Console
PassengerSeat

Three camera systems were used for the collection. The first, and primary, was a borrowed SHRP2 DAS system mounted to the front window below the rear-view mirror. A magnetic attachment was used to secure the system in place without permanent mounting to the vehicle (a borrowed U.S. government Ford F150 Extended Cab truck) (Figure 3). The DAS system collected data from four camera views, but only the face camera was used (no rear camera was installed, and the hands and the forward camera was defective). A sample composite screen frame is shown in Figure 4. The face camera view was the only retained data from the DAS by design.

Two more modern, high-resolution digital cameras were used for additional data collection. These were simple webcams and were collected using standard USB interfaces to two additional personal computers. The camera dubbed "FrontWeb" (Logitech BRIO 4K webcam with a default field of view of 90 degrees) was mounted directly in front of the driver on the dash, and the other (RVWeb) (a PEGATISAN mini-webcam with a 110-degree field of view) was mounted below DAS unit, with a physical attachment to the bottom of the DAS.

The webcam data was recorded to a pair of Dell Latitude personal computers/tablets using the Windows camera application. The DAS data was acquired using software provided by VTTI at the following resolutions and frame rates:

- RVWeb: 1920 x 1080 at 30 FPS
- FrontWeb: 1920 x 1080 at 30 FPS
- DAS = 360 x 240 at 15 FPS

The truck, instrumentation, and magnetic mount details are shown in Figure 1 through Figure 3.

Figure 1 Truck with cameras

Figure 2 Specific cameras and mounts. 1. FrontWEB camera; 2. DAS camera; 3. RVWeb mounted under DAS; 4: magnetic mount.

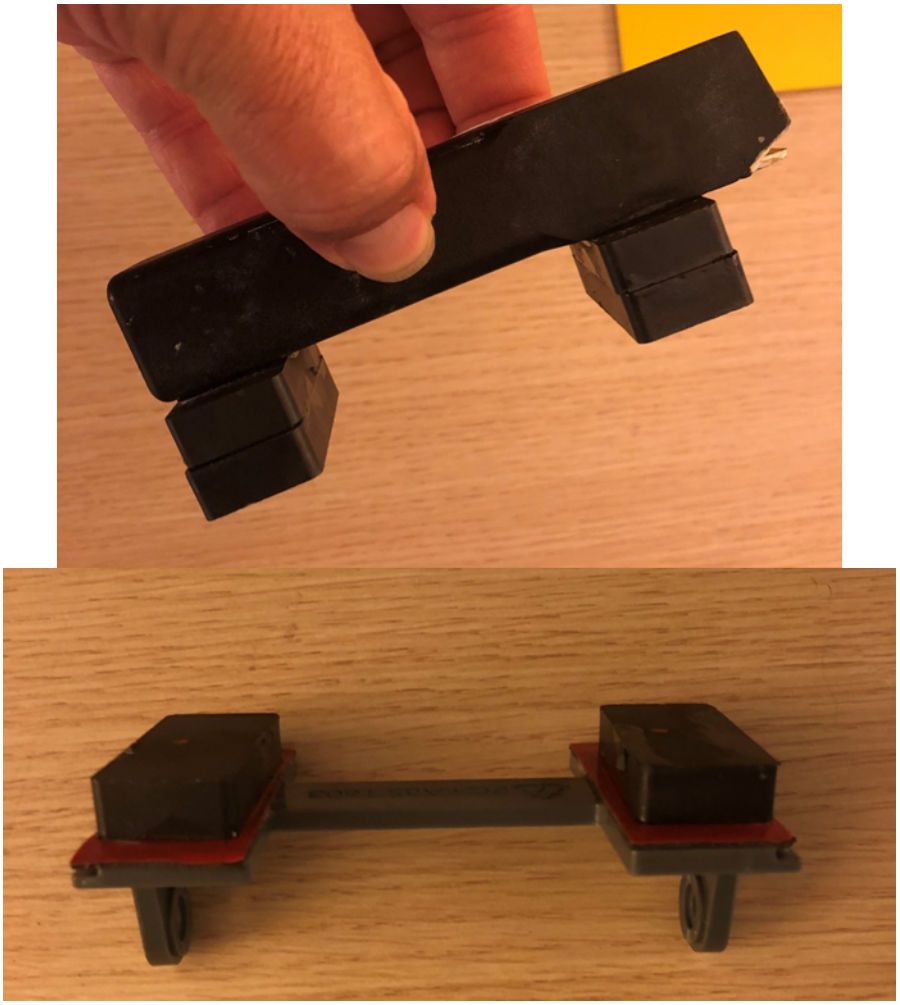

Figure 3 Magnetic units used to mount DAS camera "head unit" inside vehicle. In the top view, the component outside the vehicle is shown, with ceramic magnets show. In the bottom view, the magnets are shown attached to the DAS mounting bracket with magnets over the red adhesive pads which were normally used to affix the mount in the vehicle.

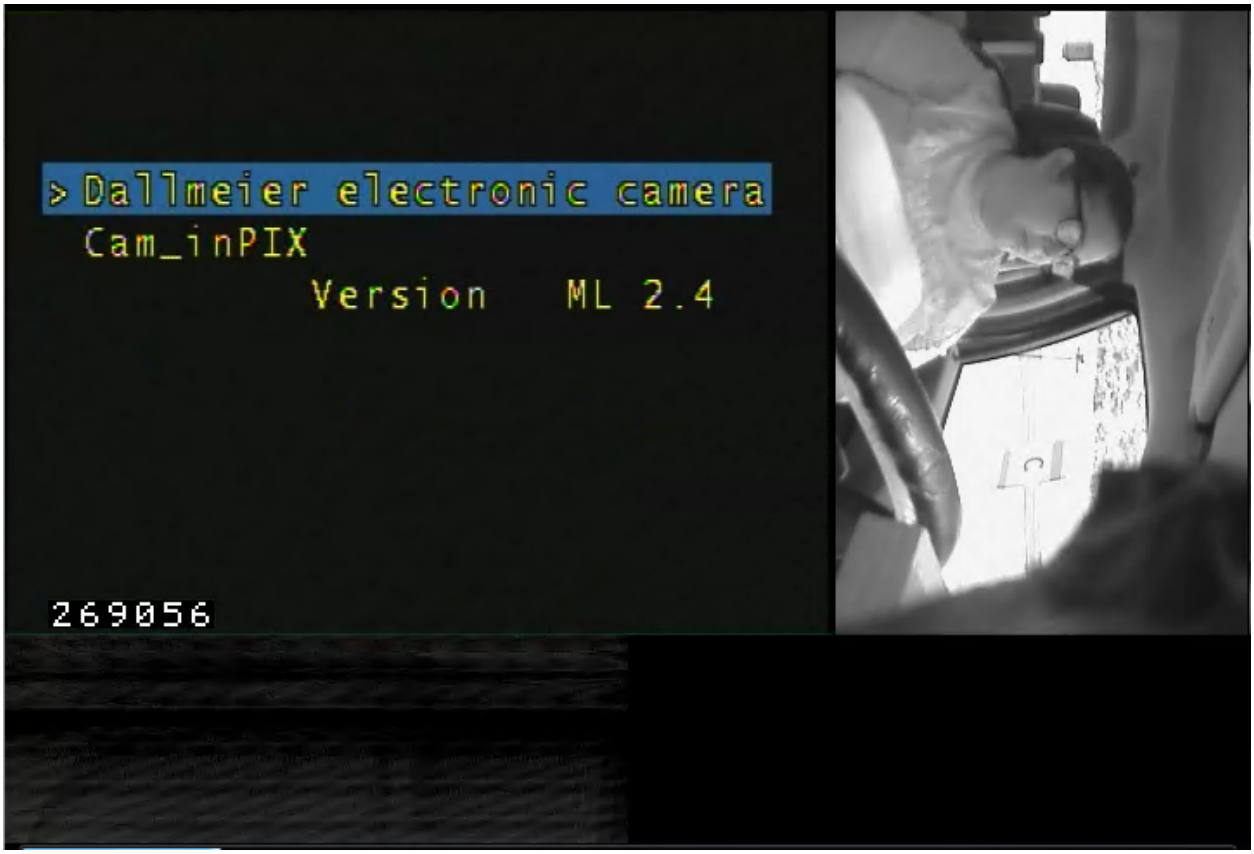

Figure 4 Composite camera view from VTTI DAS. Top left, front camera view (defective front camera); top right, face camera view rotated 90 degrees; bottom left, "hands" camera view (also defective); bottom right, rear camera view (not connected).

The DAS file was cropped, rotated, and converted to an AVI file form for subsequent processing, but the final DAS video (cropped and rotated) is supplied in MP4 format. Sample data frames for all cameras are shown in Figure 5.

Figure 5. Example video frames from (top) RVWEB camera, (middle) FrontWEB camera, and (bottom) DAS camera. Note that the DAS view is not only simply gray scale, but it also has an IR passband filter in place to help assist with night time data collection [SHRP2

Although the scripted sequences followed an audible script, we found that the instructions were often confusing and therefore we resorted to a completely manual quality check. To assist this annotation effort, the three camera views were fused by performing a frame rate conversion with the DAS view, estimating a common start and end point from each video, then fusing the frames together with a python script followed by conversion to an mp4 video file. After the annotations were completed the mp4 files were cropped to the correct starting and ending frames and the DAS frames were recomputed at the lower frame rate. An example of the fused frames is shown in Figure 6.

Figure 6. Example of fused frames at 30 fps for annotation purposes. Note these fused frames are not in the final dataset.

3. Data Organization

In this section we summarize the dataset and describe its organization.

The data is supplied as a root folder DEFADSV1.0_2022_04_03, with a table for the actions and head pose annotations and five folders (Figure 7).

A comma-separated value (CSV) file, "DEFADSMasterTableV1.csv", is provided with one line per activity and video (with three videos from different cameras per subject for the HeadPose sequences and three videos from different cameras per subject for the Action sequence). The CSV file contains labels for each activity as well as the start time, end time, start frame, and end frame (Figure 8).

Each folder contains the following:

- **Video**: 462 video files that are aligned and annotated by the DEFADSMasterTableV1.csv file. There are 77 subjects, each with six videos (three cameras and two sequences).
- **FaceCharacterization**: 462 CSV files with names corresponding to the files in the Video folder. Each file has one entry per frame with the confidence, bounding box, five facial landmarks, and the yaw-pitch-roll of the primary face detected in each frame. Some files may have missing frames if no faces were detected. The data was detected using the InsightFace tool [INSIGHTFACE] implemented with FARO [FARO].
- **Pose_OpenPose25**: 462 CSV files with body joints detected by the OpenPose software [OPENPOSE, OPENPOSEGIT]. Each line is a frame from the video with the landmarks given in x,y coordinates.
- **Pose_OpenPose25b**: Same as Pose_OpenPose25, but uses the 25b model from OpenPose [25B]
- **StillFacesCropped**: Folder with portrait-style images of the subjects; all have been cropped to the facial landmarks using [TCDCN, TCDCNGIT]. We also augmented the dataset with additional synthetic photos [GP] for a total of 108 images.

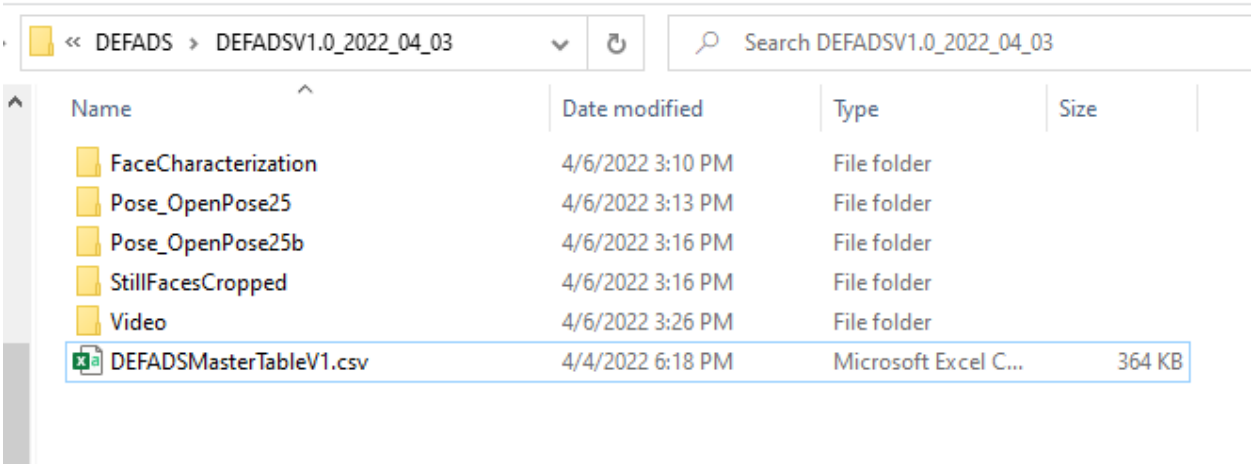

Figure 7 DEFADS Folder structure

Video	VideoTyp	Subject	Camera	Label	StartTime	EndTime	StartFram	EndFrame
P001_DAS_Action	Action	P001	DAS	Tired	00:14.2	00:23.1	217	352
P001_RVWeb_Action	Action	P001	RVWeb	Tired	00:14.2	00:23.1	435	705
P001_FrontWeb_Action	Action	P001	FrontWeb	Tired	00:14.2	00:23.1	435	705
P001_DAS_Action	Action	P001	DAS	TouchPass	00:34.2	00:39.2	521	593
P001_RVWeb_Action	Action	P001	RVWeb	TouchPass	00:34.2	00:39.2	1043	1187
P001_FrontWeb_Action	Action	P001	FrontWeb	TouchPass	00:34.2	00:39.2	1043	1187
P001_DAS_Action	Action	P001	DAS	ScaryReac	00:57.3	00:59.3	868	898
P001_RVWeb_Action	Action	P001	RVWeb	ScaryReac	00:57.3	00:59.3	1736	1796
P001_FrontWeb_Action	Action	P001	FrontWeb	ScaryReac	00:57.3	00:59.3	1736	1796
P001_DAS_Action	Action	P001	DAS	InteractW	01:16.2	01:21.2	1150	1225
P001_RVWeb_Action	Action	P001	RVWeb	InteractW	01:16.2	01:21.2	2301	2450
P001_FrontWeb_Action	Action	P001	FrontWeb	InteractW	01:16.2	01:21.2	2301	2450
P001_DAS_Action	Action	P001	DAS	WaveToLe	01:34.1	01:37.3	1414	1467
P001_RVWeb_Action	Action	P001	RVWeb	WaveToLe	01:34.1	01:37.3	2829	2934
P001_FrontWeb_Action	Action	P001	FrontWeb	WaveToLe	01:34.1	01:37.3	2829	2934
P001_DAS_Action	Action	P001	DAS	WaveToRi	01:51.2	01:55.2	1675	1735
P001_RVWeb_Action	Action	P001	RVWeb	WaveToRi	01:51.2	01:55.2	3352	3471
P001_FrontWeb_Action	Action	P001	FrontWeb	WaveToRi	01:51.2	01:55.2	3352	3471
P001_DAS_Action	Action	P001	DAS	Dance	02:06.2	02:13.2	1898	2003
P001_RVWeb_Action	Action	P001	RVWeb	Dance	02:06.2	02:13.2	3798	4008
P001_FrontWeb_Action	Action	P001	FrontWeb	Dance	02:06.2	02:13.2	3798	4008

Figure 8 Master annotation listing

For access, please contact the data curators at ORNL: karnowskitp@ornl.gov, aykacdb@ornl.gov, and ferrellrk@ornl.gov, and torkelsonlc@ornl.gov. As mentioned, a data usage license will need to be provided to track usage.

4. References

[SHRP2PROD] https://onlinepubs.trb.org/onlinepubs/shrp2/SHRP2ProductsChart.pdf Retrieved 2022-11-22

[VNAV] https://www.vectornav.com/resources/datasheets/vn-100-imu-ahrs Retrieved 2022-11-22

[HOOGE] Hooge, Kimberley D. Orsten, et al. "Evaluating automated face identity-masking methods with human perception and a deep convolutional neural network." ACM Transactions on Applied Perception (TAP) 18.1 (2020): 1-20.

[PAONE] Paone, J., et al. "Baseline face detection, head pose estimation, and coarse direction detection for facial data in the SHRP2 naturalistic driving study." 2015 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2015.

[BARAGCHISADEH] Baragchizadeh, Asal, et al. "Evaluation of automated identity masking method (AIM) in naturalistic driving study (NDS)." 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017). IEEE, 2017.

[MEDIAEVAL] Liu, Alex, et al. Driving Road Safety Forward: Video Data Privacy Task at MediaEval 2021. Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 2021.

[SHRP2] Campbell, Kenneth L. "The SHRP 2 naturalistic driving study: Addressing driver performance and behavior in traffic safety." Tr News 282 (2012).

[VTTI-HPV] Sudweeks, J. et al, “Mask Head Pose Validation Study Dataset”, 2016, DOI 10.15787/VTT1/DAFUH5

[VTTI-TS] https://insight.shrp2nds.us/projectBackground/zip_download/86 Retrieved 2022-11-22

[ONDSS] Ferrell, Regina, et al. A Publicly Available, Annotated Dataset for Naturalistic Driving Study and Computer Vision Algorithm Development. No. ORNL/TM-2019/1106. Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 2021.

[FARO] https://github.com/ORNL/faro Retrieved 2022-11-22

[OPENPOSEGIT] https://github.com/CMU-Perceptual-Computing-Lab/openpose Retrieved 2022-11-22

[OPENPOSE] Cao, Zhe, et al. "Realtime multi-person 2d pose estimation using part affinity fields." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[INSIGHTFACE] https://github.com/deepinsight/insightface Retrieved 2022-11-22

[25B] https://github.com/CMU-Perceptual-Computing-Lab/openpose_train/blob/master/experimental_models/README.md#body_25b-model---option-2-recommended Retrieved 2022-11-22

[TCDCN] Zhang, Zhanpeng, et al. "Facial landmark detection by deep multi-task learning." European conference on computer vision. Springer, Cham, 2014.

[TCDCNGIT] http://mmlab.ie.cuhk.edu.hk/projects/TCDCN/code/TCDCNv0.1.zip Retrieved 2022-11-22

[GP] https://generated.photos/ Retrieved 2022-11-22