Traffic Gesture Dataset Documentation

Nicolai Kern[1], Christian Waldschmidt
Ulm University, Institute of Microwave Engineering, 89081 Ulm, Germany
[1]nicolai.kern@uni-ulm.de

***Abstract* — This document provides the technical documentation for the traffic gesture dataset. The dataset comprises (a) a traffic gesture dataset for the design and validation of gesture recognition algorithms and (b) continuous measurements of gestures and other motions to design and test algorithms for continuous gesture recognition and out-of-distribution detection. All data has been recorded with an incoherent sensor network of three 77 GHz automotive chirp-sequence radar sensors.**

***Keywords* — AutoRad, continuous gesture recognition, radar dataset, gesture recognition, out-of-distribution**

I. Introduction

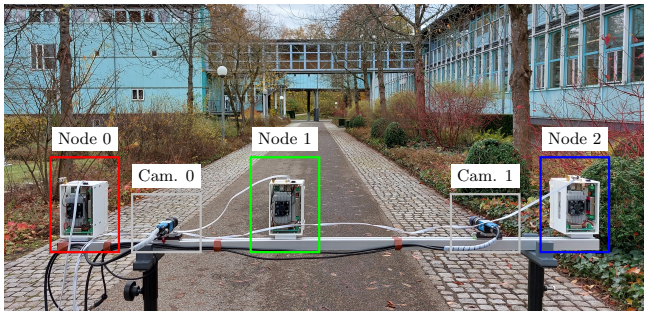
The following documentation gives an overview over the measurement setup and the dataset structure of the available datasets. For further details on the measurement setup, please see our publications [1], [2]. For baseline classification results, please see [1]. Note that a paper with more details on the dataset and further baseline results for classification (with the convolutional neural network (CNN) from [3]), continuous gesture recognition, and out-of-distribution (OOD) detection will be presented at EuRAD 2023.

Section II is a short recap of the measurement setup and signal processing. In Section III, the traffic gesture dataset is introduced, comprising measurements for eight traffic gestures and 35 participants under varying orientations. We provide different data representations (range-Doppler map (RDM), target lists, spectrograms) for the measurements, as well as a ready-to-use dataset with Doppler, range, and angle spectrograms. Finally, Section IV describes the continuous measurements, where sequences of both known and unknown gestures was recorded. These data can be used to test continuous gesture recognition as well as OOD detection, i.e. the distinction between known traffic gestures and unknown motions.

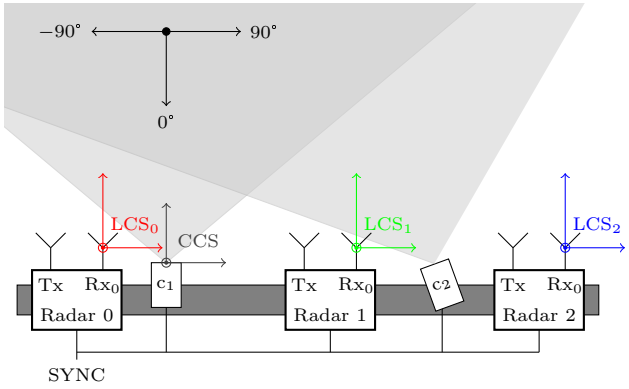
II. Dataset Recording

A. Measurement Setup

Fig. 1 shows a photograph and a top-view sketch of the measurements. The measurement setup consists of three chirp sequence (CS) radar sensors mounted on a rail. The spacing between the sensors is 55 cm between node 0 and node 1, and 140 cm between the outer nodes. The system is augmented by two cameras for stereo depth estimation. The camera data is primarily used to label the radar data and to estimate the positions of the participants. Furthermore, it can be used to resimulate the measurements [4] or for camera-based classification [5][1] The radar parameters of each of the sensors are summarized in Table 1.



(a)



(b)

Fig. 1. (a) Photography and (b) top-view sketch of the measurement setup. The black arrows indicate the orientations of the subjects. Figure taken from [1].

B. Signal Processing

From the raw data, different data representations are derived:

- Sequence of RDMs, where each frame is represented as a single range-Doppler map.
- Sequence of target lists, where each frame is represented by a target list comprising all constant false-alarm rate (CFAR) detections in that frame.
- Doppler, range, and angle spectrograms, which are reconstructed from the target lists.

Please see [1] for an in-depth description of the signal processing. Note that the provided target lists are spatially filtered, i.e. targets more than 2 m away from the participant

[1]The keypoint data and 3D motion data are currently rather unstructed, but are available upon request.

Table 1. Radar parameters

Parameter	Value	Description
f_c	79.0 GHz	center frequency
B	3.36 GHz	bandwidth
T_{RRI}	138.0 µs	ramp repetition interval
N_c	128	number of chirps per sequence
N_s	336	number of samples per chirp
N_{tx}	3	number of transmit antennas
N_{rx}	4	number of receive antennas
ΔR	4.5 cm	range resolution
Δv	10.7 cm/s	velocity resolution
R_{max}	15.0 m	maximum unambiguous range
v_{max}	± 6.87 m/s	maximum unambiguous velocity
θ_{max}	$\pm 60°$	azimuthal field of view

Table 2. Overview over the measurement days.

Date	Location	Participants	#Files
2021-05-07	outdoor	0,1	56
2021-05-10	outdoor	2,3,4,5	104
2021-05-11	outdoor	6,7,8,9,10,12	158
2021-05-17	indoor	13,14,15,16,17	157
2021-05-27	indoor	18,19,20,0,21,22	167
2021-05-28	outdoor	23,24,25,26,27,28,29,30,31	288
2021-06-09	indoor	32,33,34,35,11,23,31	176

position are removed. All data is created for the three radar sensors in the network independently.

III. Traffic Gesture Dataset

A. Measurements

The traffic gesture dataset comprises data from eight gestures (for a visualization check out the website). The gestures were recorded over the course of seven days, and a total of 36 participants took part in the measurements. Measurements were conducted indoors and outdoors, with the scenarios described in [1]. Participants are indexed from 0 to 35. Gestures were recorded under different positions and orientations, with the orientation changes more pronounced. The positions and orientations of the measurements are shown in [1]. Table 2 gives an overview over which participants took part on which day and in which location.
Each measurement took approx. 90 s during which all eight gestures were recorded for approx. 10 s each. Measurements are repeated multiple times for each participant, with different orientations. Each participant was instructed to perform gestures under 0° and 90° (cf. Fig 1 for definition) and at up to two arbitrary orientations (denoted *x1* and *x2*).

B. Per-Gesture Measurements

The full measurements with data from all gestures are split into per-gesture measurements based on the camera time stamps. In the dataset, each extracted per-gesture measurement corresponds to a single h5 file, and all files from one day are compressed to a ZIP file. Each file has a unique identifier (*meas_key*), which is composed of the participant index and the instructed orientation.

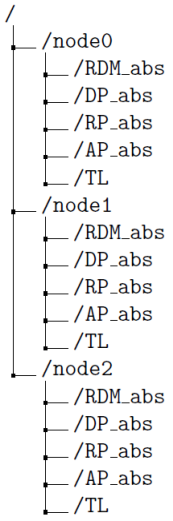


Fig. 2. File structure of the h5 files of the per-gesture measurement files.

Each *h5* file contains data from all three radar sensors, and the file structure of the h5 files is shown in Fig. 2. For each radar sensor node, the sequence of absolute-value range-Doppler maps (*RDM_abs*), the sequence of target lists (*TL*) and the reconstructed absolute-value Doppler, range, and angular profiles (*DP_abs*, *RP_abs*, *AP_abs*) are available. Examples of the data representations are given in Fig. 3.
Further information is provided in each file by the attributes listed in Table 3. The estimated orientation *gesture_oriEst* is useful for the orientation instructions *x1* and *x2*, where the participants were instructed to perform the gestures under an arbitrary orientation. For 0° and 90°, the estimated orientation might slightly deviate from the instructed values, as the instructions were not monitored strictly. The position estimate *gesture_posEst_CCS* specifies the position of the right hip as estimated by the stereo camera in the camera coordinate system (CCS). Moreover, the estimated positions after transformation to the local coordinate systems of the radar sensors are specified as *gesture_posEst_node*{*0,1,2*}. For the transformation, the camera-radar calibration file that is provided in the *misc* folder is used. For further details regarding calibration and transformations, see [1], [2]. The position estimate is used for the spatial filtering, but can also be used e.g. to learn a position normalization (see [1]).

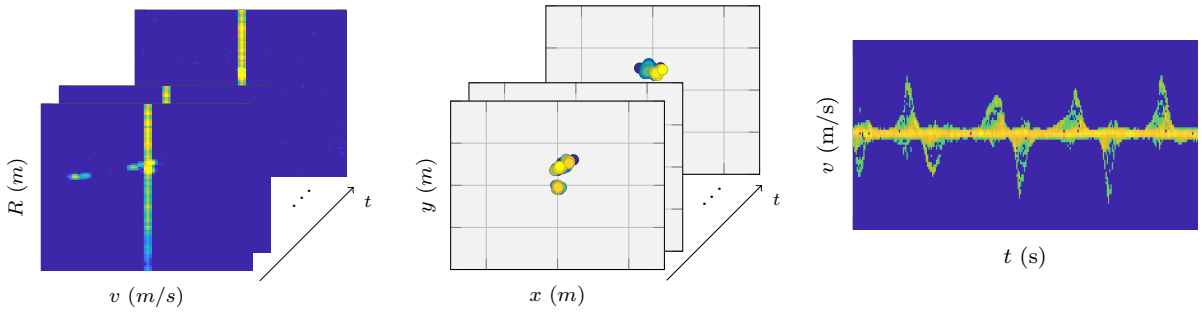
Subsequently, the data representations are described more into detail.

1) RDM

The RDMs are of shape

$$\text{velocity bins} \times \text{range bins} \times \text{frames} = 128 \times 228 \times N_f,$$

and converted to dB. They come with the additional attributes *v_vec* and *r_vec*, defining the velocity and range values of the



(a) (b) (c)

Fig. 3. Data representations: (a) range-Doppler maps, (b) target lists (signal-to-noise ratio (SNR) indicated by color), (c) Doppler spectrogram.

range-Doppler bins. Note that the RDMs are cropped in range dimension to reduce data size.

2) DP, RP, AP

The reconstructed Doppler, range, and angle profiles are of shape

$$(\text{velocity bins}/\text{range bins}/\text{angle bins}) \times \text{frames},$$

and each profile comes with a vector defining its y-axis. Note that the range profiles are cropped to 128 range bins (the range where the gestures take place). Note that the profiles are reconstructed from the target lists after spatial filtering, i.e. only targets closer than 2 m are considered.

3) Target Lists

The target lists contain all targets found by the CFAR algorithm and are of shape

$$num.targets \times \text{num. target parameters} \times \text{frames}\,.$$

If less than 500 targets are found by the CFAR, the list is filled up with zeros which have to be removed. Each target is described by its target parameters, which are listed in an additional attribute *target_params*. The target params are:

1) *t_r*: target range
2) *t_v*: target velocity
3) *t_azi*: azimuth angle of the target
4) *t_x*: x-position of the target (in local radar coordinate system)
5) *t_y*: y-position of the target (in local radar coordinate system)
6) *t_pow*: magnitude of the target's reflection in dB
7) *t_snr_noise*: SNR of the target vs. the estimated noise floor

C. Subsampled Datasets

Besides the full measurements, we also provide training-ready datasets. We typically train on 2 s observation time, i.e. 60 frames at 30 fps. Since the per-gesture measurements are substantially longer, they are subdivided into smaller snippets. In order to increase the number of samples for training, neighboring snippets have an overlap.

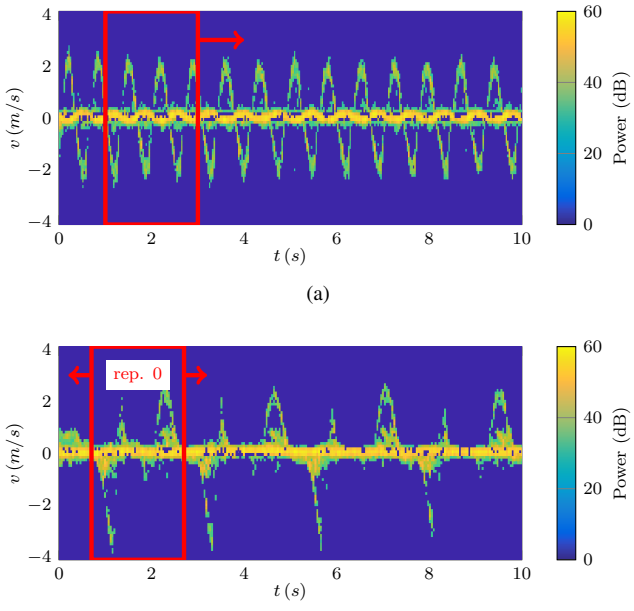


Fig. 4. Subsampling for repetitive and non-repetitive gestures. For repetitive gestures (a), there is no real start and end point, so the 2 s-sliding window is just moved along the measurement with a step width of 0.5 s. For non-repetitive gestures (b), it is ensured that each sample comprises only a single repetition. In the example, the sliding window is shifted to the left and right of the first repetition.

For a more in-depth description of the procedure and the benefits, see [6]. For the subsampling, we distinguish between repetitive gestures and non-repetitive gestures. Repetitive gestures (such as waving someone through) are executed repeatedly and thus take longer than 2 s. As a consequence, the full measurement can be cut into smaller snippets without taking into account the starts and stops of the individual repetitions. In contrast, g_6 ("Stop") and g_7 ("Thank you") are non-repetitive as they are typically executed only once. Hence, the subsampling into smaller snippets should take place around the actual repetitions. The process is illustrated for both types of gestures in Fig. 4.

To start directly with training, one can use the training-ready dataset. Here, each *h5* file corresponds to one sample with

Table 3. Attributes in the h5 files of the full measurements.

Attribute	Value Range	Description
gesture_gesture	0-7	Gesture label: 0: "Fly" 1: "Come Closer" 2: "Slow Down" 3: "Wave" 4: "Push Away" 5: "Wave Through" 6: "Stop" 7: "Thank You"
gesture_user	0-35	participant index
gesture_orientation	{0,90,x1,x2}	Instructed orientation; up to 4 measurements per participant: 0: towards radar 90: side-looking x1: random orientation 1 x2: random orientation 2
gesture_oriEst	$[-180, 180]$	Orientation estimate in degree
gesture_posEst_CCS	-	Stereo camera position estimate (right hip) in camera coordinate system (CCS)
gesture_location	*indoor* or *outdoor*	*indoor*: car park environment *outdoor*: street environment
nr_frames	-	Number of radar frames
meas_date	2021_x_x	measurement date
meas_key	user$X$_Ydeg	unique identifier X: participant Y: instructed orientation
Only for non-repetitive gestures:		
repetition_starts	-	For non-repetitive gestures: Start frames of individual repetitions
repetition_ends	-	For non-repetitive gestures: End frames of individual repetitions

60 frames and comes with a gesture label (*gesture_gesture*) and a participant ID (*gesture_user*). **When using the dataset, make sure to split training/validation/test sets properly by using different users for each set! Splitting the dataset completely randomly results in validation and test set corruption due to the overlap, and results will be unrealistically good!**

Beyond gesture and participant information, each sample contains the original *meas_key* as well as the start and stop frame in the original full measurement file. This can be used to easily create identical datasets for RDMs and TLs without having to consider the gesture time stamps for g_6 and g_7.

IV. Out-of-Distribution Datasets

OOD detection is crucial in automotive applications, as many motions different from the ones in the closed training dataset can be observed. These motion patterns cannot be ruled out by simple thresholding, so OOD detection algorithms have to be designed and tested. To this end, the dataset contains multiple OOD test sets.

Table 4. Single-fold CV accuracies for cross-scenario tests.

Class	Number of samples			
	TGD	OOD Idle	OOD Walk	OOD Gesturing
Fly	1000	-	-	-
Come Closer	1000	-	-	-
Slow Down	1000	-	-	-
Wave	1000	-	-	-
Push Away	1000	-	-	-
Wave Through	1000	-	-	-
Stop	1000	-	-	-
Thank You	1000	-	-	-
Idle	1000	-	-	-
Walk	1000	-	-	-
Gesturing	1000	-	-	-

The OOD datasets are derived from continuous measurements of four different kinds:

1) "backgrnd": Moving and gesturing on the spot; all motions are different from the known traffic gestures.
2) "contIdle": Performing known traffic gestures with idle in between.
3) "cont": Performing known traffic gestures without idle in between.
4) "walk": Walking around without gesturing.

Continuous measurements are recorded for six participants, of which two (with identifier 2 and 31) also took part in the traffic gesture measurements, and four (identifiers 36, 37, 38, 39) didn't. Videos of exemplary measurements are shown on the dataset website. The full measurements are provided in the *cont_measurement* folder. For all continuous measurements, per-frame labels are provided in the *h5* files (attribute named *per_frame_label*). As the continuous measurements contain both known (in-distribution (ID)) and unknown (OOD) motions, ID and OOD test sets can be derived by subsampling the corresponding parts of the full measurements.

Four different ready-to-use datasets are created, which can be found in the *cont_datasets* folder:

1) "backgrnd": OOD test set; contains unknown gesturing motions similar to the ones in the traffic gesture dataset
2) "walk": OOD test set; contains walking motions without gesturing
3) "idle": OOD test set; contains idle samples, where the participant shows little or no motion; very small dataset
4) "cont": ID test set; contains known traffic gestures

The procedure we used to test OOD detection algorithms is shown in Fig. 5. The classifier is trained single-fold on the traffic gesture dataset from the previous chapter, with participants 2 and 31 excluded. The best model is selected and applied to combinations of ID and OOD data. The "cont" dataset serves as ID data and is combined either

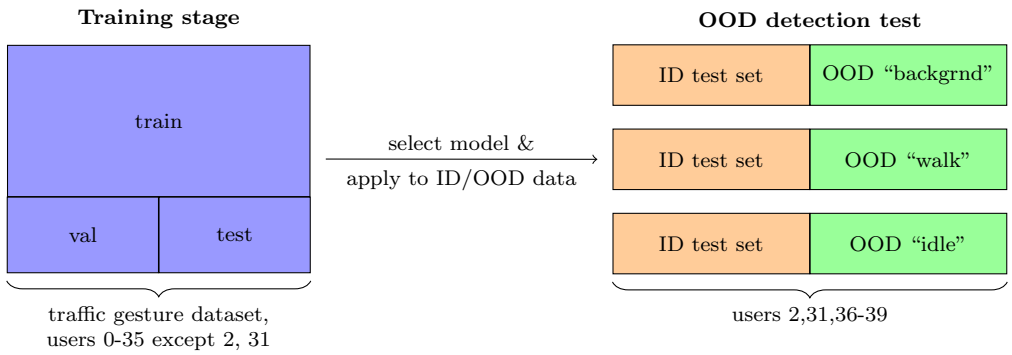


Fig. 5. Procedure to examine the performance of OOD detection algorithms.

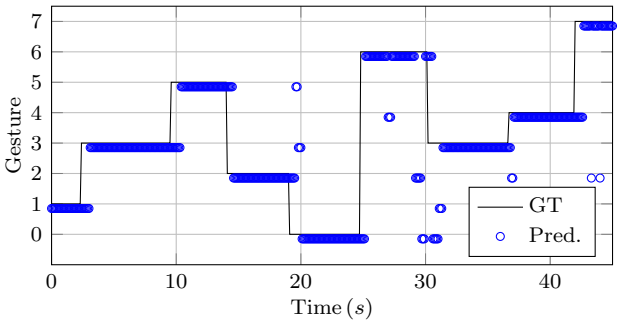


Fig. 6. Continuous classification: Predictions vs. ground truth (GT).

with the “backgrnd”, “walk”, or “idle” dataset to examine OOD detection performance. By following this procedure a proper split between the training stage and the OOD tests is maintained. Baseline results for approaches such as temperature scaling and ODIN can be found in the EuMW paper. For approaches requiring hyperparameter tuning, the different OOD sets can be used as validation and test set.

V. Continuous Classification Dataset

The continuous measurements can also be used to test continuous classification with or without OOD detection. For this, the observation window can be moved along the full sequence frame-by-frame, such that frame-wise predictions are obtained vs. the per-frame labels as shown in Fig. 6. Note that the 2 s samples of the traffic gesture datasets contain no transitions. Hence, additional measures have to be taken to ensure a fast response to gesture transitions, as explained in the EuRAD paper.

VI. Conclusion

This documentation described the main features of the traffic gesture dataset and the accompanying test sets for OOD detection and continuous classification. If you have any questions, please contact us via mail. If you find this dataset helpful and use it for your own research, please cite the EuRAD paper.

References

[1] N. Kern, T. Grebner, and C. Waldschmidt, “Pointnet+lstm for target list-based gesture recognition with incoherent radar networks,” *IEEE Transactions on Aerospace and Electronic Systems*, pp. 1–1, 2022.

[2] N. Kern, A. Holzbock, T. Grebner, V. Belagiannis, K. Dietmayer, and C. Waldschmidt, “A ground truth system for radar measurements of humans,” in *2022 14th German Microwave Conference (GeMiC)*, 2022, pp. 84–87.

[3] N. Kern, M. Steiner, R. Lorenzin, and C. Waldschmidt, “Robust Doppler-based gesture recognition with incoherent automotive radar sensor networks,” *IEEE Sens. Lett.*, vol. 4, no. 11, pp. 1–4, 2020.

[4] N. Kern, J. Aguilar, T. Grebner, B. Meinecke, and C. Waldschmidt, “Learning on multistatic simulation data for radar-based automotive gesture recognition,” *IEEE Trans. Microw. Theory Techn.*, pp. 1–12, 2022.

[5] A. Holzbock, N. Kern, C. Waldschmidt, K. Dietmayer, and V. Belagiannis, “Gesture recognition with keypoint and radar stream fusion for automated vehicles,” in *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part I.* Springer, 2023, pp. 570–584.

[6] N. Kern and C. Waldschmidt, “Data augmentation in time and doppler frequency domain for radar-based gesture recognition,” in *2021 18th European Radar Conference (EuRAD)*, 2022, pp. 33–36.