Automotive Radar Dataset for Deep Learning Based 3D Object Detection

Michael Meyer*, Georg Kuschk*
Astyx GmbH, Germany
{g.kuschk, m.meyer}@astyx.de

***Abstract*— We present a radar-centric automotive dataset based on radar, lidar and camera data for the purpose of 3D object detection. Our main focus is to provide high resolution radar data to the research community, facilitating and stimulating research on algorithms using radar sensor data. To this end, semi-automatically generated and manually refined 3D ground truth data for object detection is provided. We describe the complete process of generating such a dataset, highlight some main features of the corresponding high-resolution radar and demonstrate its usage for level 3-5 autonomous driving applications by showing results of a deep learning based 3D object detection algorithm on this dataset. Our dataset will be available online at: www.astyx.net**

***Keywords*— 3D object detection, deep learning, dataset, multi sensor fusion, radar**

I. Introduction, Related Work, Contributions

For highly-automated and autonomous driving (level 3-5) the corresponding vehicles rely on an accurate and detailled perception of the driving environment - provided by the respective sensor setup, which typically consists of complementary sensors with a focus on camera, lidar and radar. Each of these sensor-families have their corresponding strengths and weaknesses, requiring an intelligent fusion and combination of their data. For decision making w.r.t. the driving strategy, the system relies on an aggregation of the raw sensor data into a more abstract level of scene understanding. For example it needs to know the location and additional attributes of obstacles and other road users, commonly stated as the problem of object detection. While 2D and 3D object detection using highly generalizable deep learning approaches is nowadays a well researched problem on common camera data, yielding increasingly better results, algorithms based on automotive radar data still suffer from a research lag in this area. Exemplary work using deep learning approaches on radar data just recently emerged, for example pointwise semantic classifaction [1] and classification of 2D (Range-Doppler) images [2] . Main reasons for the lack in research activity are the absence of high resolution radar enabling a dense sampling and separation of the environment, a complicated system setup and interpretation of the data, non-disclosed low level sensor data and the missing availability of publicly usable data annotated with ground truth information. Currently introduced new generations of radar (e.g. the radar used in this work, the Astyx 6455 HiRes) now bring the technical capabilities up to a point where the abovementioned drawbacks are mostly diminished to a satisfactory level and the radar sensors even are able to generate spatial 3D measurements (see Fig. 9).

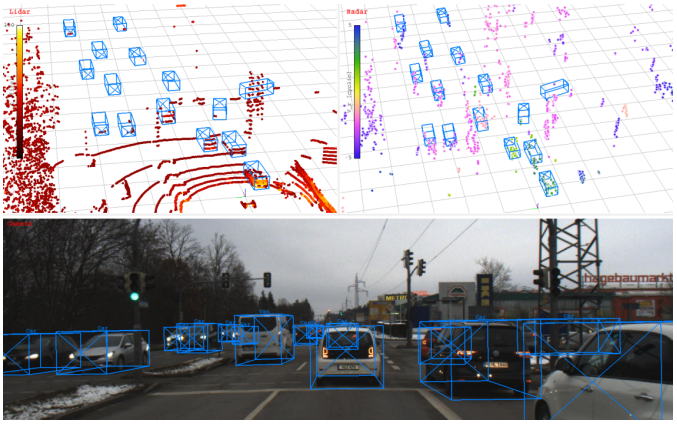


Fig. 1. Exemplary frame of our multi-sensor dataset, containing radar, lidar and camera information plus annotated 3D ground truth objects

Thus, availability of publicly usable data annotated with ground truth information is now a top bottleneck in development of radar based applications - especially for deep learning based machine learning which heavily relies on massive amounts of training data.
History shows that publishing datasets tremendously stimulated research in their respective areas, namely the Middlebury Benchmark [3] in the field of optical stereo reconstruction and optical flow, MNIST [4], COCO [5] and ImageNet [6] in the field of classification, and of course the widespread KITTI benchmark [7] - now the quasi-standard to evaluate on when publishing automotive related research work and making the algorithms systematically comparable.
Other major automotive datasets are [8], [9], [10] - but none of them include high-resolution radar sensor data. The MSTAR dataset in contrast contains radar-based SAR acquisitions of military targets from an airborne platform [11]. To our best knowledge, the only public dataset containing automotive radar data is the recently introduced nuScenes dataset [12]. However, this dataset contains radar data of a different, non-disclosed type of radar sensor with sparsely populated 2D radar information (around $\approx$ 100 2D points compared to $\approx$ 1000 3D points of the Astyx 6455 HiRes).

*Both authors contributed equally to this paper.

We therefore see a strong academic need for publicly available, high-resolution radar data (pointcloud level), including synchronized camera and lidar data and propose to fill this gap with our dataset.

II. Sensor System Setup

The following sensor data mounted on a test vehicle was chosen to be included in the dataset and placed in front-looking direction, maximizing the overlap of the commonly observed area.

- **Radar**: Astyx 6455 HiRes (13 Hz, HFoV/VFoV: $110° \times 10°$, range 100m)
- **Camera**: Point Grey Blackfly (30 Hz, RGB 8-Bit, resolution 2048×618 pixel)
- **Lidar**: Velodyne VLP-16 (10 Hz, 16 laser beams, range 100m)

A. Calibration, Co-Registration

Calibrating the intrinsic parameters of the cameras, the standard approach of [13] is applied, based on sub-pixel accurate corner detection using a well-known physical checkerboard. Internal calibration of the radar sensor was done based on a predefined corner reflector in a controlled measurement chamber. Both intrinsic calibrations were done as a offline preprocessing step. The co-registration process of the three sensors was split into two separate automatic processes, found to yield better accuracy than using one calibration pattern to simultaneously register all sensors in one step. For registering the lidar with the camera we use a modified approach based on [14], based on automatic pointcloud segmentation of the known chessboard in the lidar data and matching this pointcloud to the detected chessboard corners in the camera (2D-3D registration). Registering the radar with the lidar (2D-3D registration) was done with a physical target exhibiting an accurate response in 3D space for both sensors. To further increase accuracy, the 2D-3D (camera-lidar) and 3D-3D (radar-lidar) relative pose estimation, based on least-squares estimation of the 6 unknown parameters per pose, were embedded into a RANSAC framework [15] to reduce the impact of outliers.

III. Ground Truth Generation

Having these co-registered and calibrated sensors, one is now able to generate ground truth data for 3D object detection by using and combining the benefits of the involved radar, lidar and camera sensors. This is typically done at its core by manually drawing 3D bounding box about the objects to annotate:

1) **Lidar**: Positioning and aligning the full-3D orientation is mainly done using the LiDAR sensor, as it is the best sensor to capture detailed and accurate 3D properties.
2) **Camera**: Fine-tuning the object properties is done using the camera information. Especially class information and height of an object is usually insufficiently determinable by a lidar sensor.
3) **Radar**: Due to restricted range of the lidar (mostly due to the angular opening between the different laser beams), far-away objects are usually not covered by any lidar measurment. We nevertheless annotate these objects, as these are typically still visible in both radar and camera (see Fig. 2. However, the certainty of the position and dimension is not as accurate as for nearby objects, hence we also store object attributes for the uncertainty of position and dimension. Additionally we annotated 'invisible' objects as well. This means physical objects which do not have any lidar or camera measurements, but are clearly visible in the radar data (e.g. by multipath reflections propagating below other cars) and could be associated via temporal referencing - becoming visible prior or later on during the data recording.

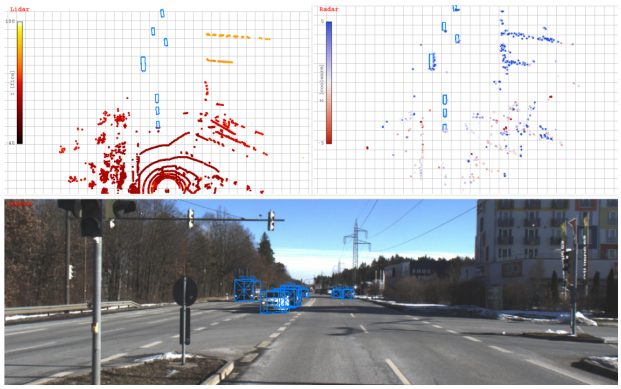


Fig. 2. Sensor output of lidar, radar and camera for one traffic scene. The lidar in the upper left provides dense measurement in proximity of the ego vehicle (which is located in the middle at bottom) while not covering any of the light blue objects. The radar in comparison to the upper right is not well suited for close proximity due to technical reasons, but very well suited for detecting long distance objects.

A. Semi-Automatic Labeling via Active Learning

Since manual labeling is a very tedious, slow and costly task which doesn't scale up to larger datasets, automatic pre-labeling followed by manual refinement is very essential. For this task, we use the work of [16], based on deep learning based 3D object detection and performing multi-sensor fusion on low-level sensor data. To minimize the amount of labels needing manual refinement, we embed this 3D object detection network into an Active Learning approach based on uncertainty sampling using estimated scores as approximation [17], [18] (see Fig. 3). To this end, we draw from the automatically pre-labelled data N frames where the network is most unsure about its decisions, correct these via manual fine-tuning and therefore maximize the information gain for the network in the next training and pre-labeling round.

IV. Results

The resulting dataset which we provide at this stage for free usage consists of 500 synchronized frames (radar, lidar, camera), containing around 3000 very accurately labeled 3D

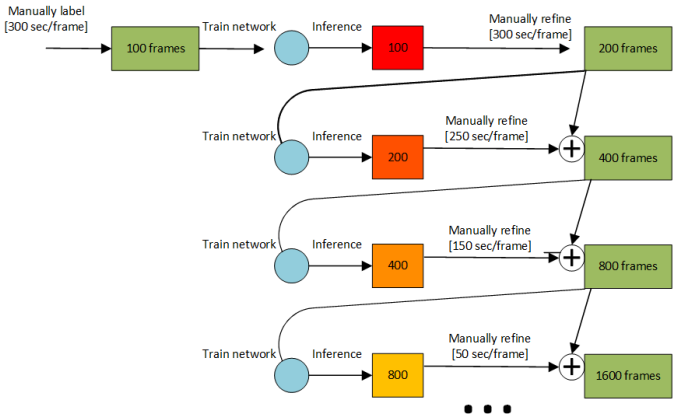


Fig. 3. Workflow of the semi automatic labeling process. The quality of the trained object detection network and of the pre-labeled data gets increasingly better, thus reducing the required time for manual corrections.

object annotations. Whereas the majority of objects are of class 'Car', we also provide a smaller amount of ground truth data for 7 classes in total (Bus, Car, Cyclist, Motorcyclist, Person, Trailer, Truck). The ground truth data annotation contains for each object the following attributes:

- 3D position (x,y,z)
- 3D rotation (rx,ry,rz)
- 3D dimension (w,l,h)
- Class information
- Occlusion indicator
- Uncertainty (position, dimension)

The dataformats are all non-proprietary standard image and pointcloud formats, with the 3D object detection ground truth in text format. We do not host an official benchmark test suite for standardized evaluation and ranking of object detection algorithms as in [7], as this would require withholding test data to avoid an overfitting on the evaluation test set. This poses a conflict of interest - since we're also researching and developing object detection algorithms based on this data and would be the only ones in full control of the ground truth data. A distribution of the point cloud density of both the lidar and radar sensor are shown in Fig. 4 with a distribution of the ground truth objects over the different classes shown in Fig. 5. In Fig. 7 we show the exemplary orientational and spatial distribution of cars in the ground truth data.

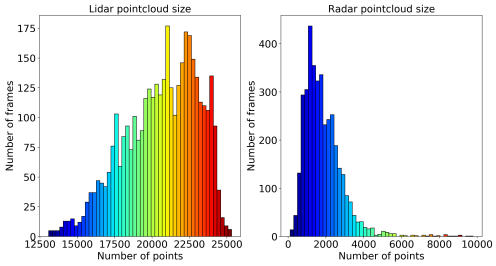
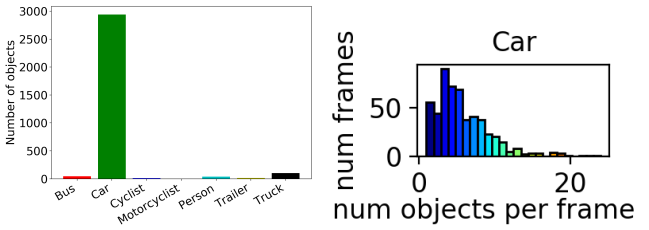


Fig. 4. Sensor data distribution - histogram of point cloud size per frame for lidar and radar.



(a) Groundtruth class distribution (b) Number of class instances per frame of class 'Car'

Fig. 5. Dataset statistics

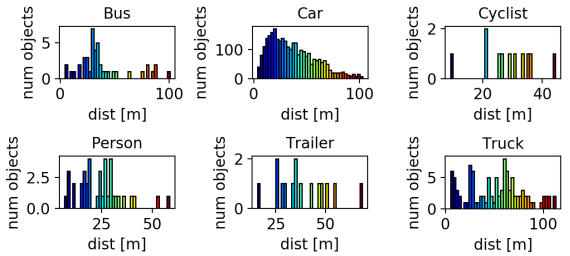
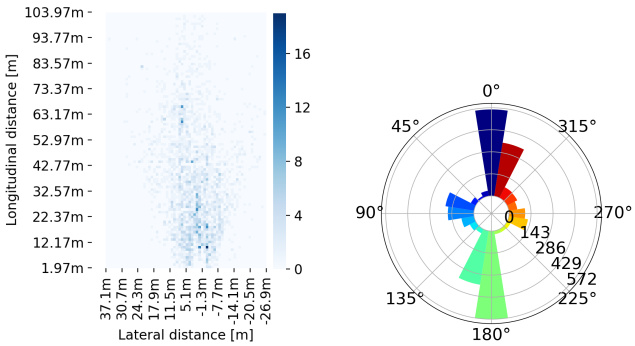


Fig. 6. Distribution of objects over distance. For a complete overview of all statistic see the corresponding dataset.



(a) Spatial distribution of class 'Car' in xy-plane, with the observers ego-vehicle centered at (0,0).

(b) Orientational distribution of class 'Car'

Fig. 7. Dataset statistics

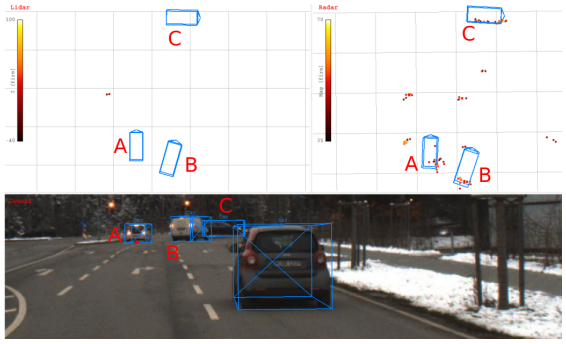


Fig. 8. Radar properties of half partly occluded car in 99m distance (C) and two fully visible cars in about 80m distance (A and B). Color coding of the radar points according to magnitude.

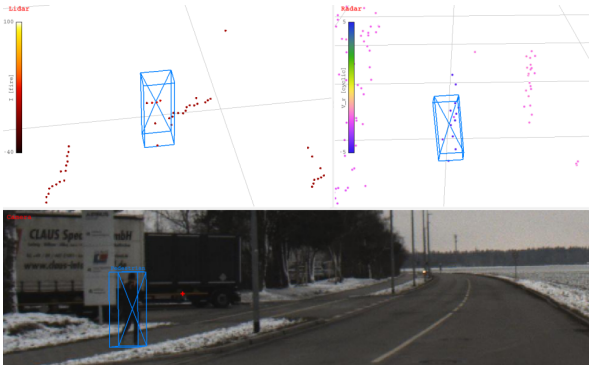
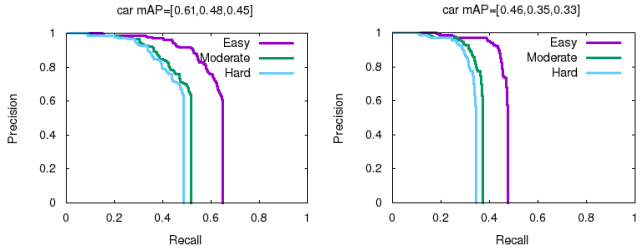


Fig. 9. Lidar and radar properties of a pedestrian in 26m distance.

V. Experimental Evaluation

For evaluation, we applied the work of [16], based on deep learning based 3D object detection, performing radar-camera vs. lidar-camera fusion on low-level sensor data. To this end, we randomly split the dataset into train and test data using a ratio of 4:1, trained two networks (radar-camera and lidar-camera) for 22k iterations using a mini batchsize of 16 and evaluated the results both in terms of classification, localization and orientation accuracy by using an IoU (intersection over union) threshold of 0.5. Despite an immensely small amount of annotated train and test data (in terms of deep learning requirements for automatic feature extraction), the work of [16] achieved an average precision (AP) of $(0.61, 0.48, 0.45)$ for the detection of cars using radar-camera fusion and $(0.46, 0.35, 0.33)$ using lidar-camera fusion (see Fig. 10). Here we differentiate the resulting accuracy between three diffulty categories (easy, moderate, hard), depending on the visibility/occlusion of the objects.



(a) PR-curve for Radar+Camera (b) PR-curve for LiDAR+Camera

Fig. 10. Evaluation of the 3D object detection work of [16] on our dataset, using radar-camera fusion. Easy, Moderate and Hard evaluations differ on the type of test objects they include. Easy only evaluates on objects which are fully visible in all sensors, Moderate also on party occluded objects and *Hard* includes objects which are completely invisible to the camera and lidar and are only visible in the radar measurements.

VI. Conclusion and Future Work

With the proposed public radar-based automotive data we hope to stimulate further research in radar-based object detection, as well as [radar, lidar, camera]-based low-level sensor fusion. We showed that using our dataset, a reasonable accuracy can be reached when training 3D object detection radar-based algorithms. The current main limitation of the dataset is its size, both in terms of the amount of required data as well as the required amount of variety w.r.t. environmental conditions (daylight, season, weather, ..) and scenes (rural, urban, highway, ...). To this end we plan to further extend our dataset, scaling its efficiency up using our semi-automatic labeling approach and invite contributors to join our effort. Another option for the future would be the incorporation into existing benchmark evaluations, allowing for a systematic ranking of radar-based algorithms.

References

[1] O. Schumann, M. Hahn, J. Dickmann, and C. Wöhler, "Semantic segmentation on radar point clouds," in *2018 21st International Conference on Information Fusion*. IEEE, 2018, pp. 2179–2186.

[2] R. Perez, F. Schubert, R. Rasshofer, and E. Biebl, "Single-frame vulnerable road users classification with a 77 ghz fmcw radar sensor and a convolutional neural network," in *19th International Radar Symposium (IRS)*, 2018, pp. 1–10.

[3] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2003.

[4] Y. LeCun, "The mnist database of handwritten digits," *http://yann. lecun. com/exdb/mnist/*, 1998.

[5] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: common objects in context," in *European Conference on Computer Vision*. Springer, 2014, pp. 740–755.

[6] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[7] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *International Journal of Robotics Research (IJRR)*, 2013.

[8] (2016) Oxford robotcar dataset. http://robotcar-dataset.robots.ox.ac.uk/ [Online; Status 11-Feb-2019].

[9] (2017) Udacity annotated driving datasets. https://github.com/udacity/self-driving-car [Online; Status 11-Feb-2019].

[10] (2018) Apollo data open platform. https://data.apollo.auto [Online; Status 11-Feb-2019].

[11] E. R. Keydel, S. W. Lee, and J. T. Moore, "MSTAR extended operating conditions: A tutorial," in *Algorithms for Synthetic Aperture Radar Imagery III*, vol. 2757. International Society for Optics and Photonics, 1996, pp. 228–243.

[12] (2018) nuscenes dataset. https://www.nuscenes.org/ [Online; Status 11-Feb-2019].

[13] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, 2000.

[14] W. Wang, K. Sakurada, and N. Kawaguchi, "Reflectance intensity assisted automatic and accurate extrinsic calibration of 3d lidar and panoramic camera using a printed chessboard," *Remote Sensing*, vol. 9, no. 8, 2017.

[15] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[16] M. Meyer and G. Kuschk, "Deep learning based 3d object detection for automotive radar and camera," in *Manuscript submitted to EuRad 2019*.

[17] D. D. Lewis and W. A. Gale, "A sequential algorithm for training text classifiers," in *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*. Springer-Verlag New York, Inc., 1994, pp. 3–12.

[18] B. Settles, "Active learning literature survey," Tech. Rep., 2010.