DLRAD - A FIRST LOOK ON THE NEW VISION AND MAPPING BENCHMARK DATASET FOR AUTONOMOUS DRIVING

F. Kurz[1,*], D. Waigand[2], P. Pekezou-Fouopi[2], E. Vig[1], C. Henry[1], N. Merkle[1], D. Rosenbaum[1], V. Gstaiger[1], S. Azimi[1], S. Auer[1], P. Reinartz[1], S. Knake-Langhorst[2]

[1] German Aerospace Center (DLR), Remote Sensing Technology Institute, Wessling, Germany
franz.kurz@dlr.de
[2] German Aerospace Center (DLR), Institute of Transportation Systems, Braunschweig, Germany
daniel.waigand@dlr.de

Commission I, WG I/6

KEY WORDS: Airborne camera, Vehicle sensors, Benchmark dataset, Autonomous driving, Sensor fusion

ABSTRACT:

DLRAD - a new vision and mapping benchmark dataset for autonomous driving is under development for the validation of intelligent driving algorithms. Stationary, mobile, and airborne sensors monitored simultaneously the environment around a reference vehicle, which was driving on urban, suburb and rural roads in and around the city of Braunschweig/Germany. Airborne images were acquired with the DLR 4k sensor system mounted on a helicopter. The DLR research car FASCarE is equipped with the latest sensor technology like front/rear radar, ultrasound and laser sensors, optical single and stereo cameras, and GNSS/IMU. Additionally, stationary terrestrial sensors like induction loops, optical mono and stereo cameras, radar and laser scanners monitor defined sections of the path from the ground. Simultaneously, the helicopter with the 4k sensor systems follows the reference car by keeping it all the time in the central nadir view. A next crucial step in the construction of the DLRAD benchmark dataset is the annotation of all objects in the reference dataset.

The DLRAD benchmark dataset enables a huge variety of validation capabilities and opens a wide field of possibilities for the development, training and validation of machine learning algorithms in the context of autonomous driving. In this paper, we will present details of the sensor configurations and the acquisition campaign, which had taken place between the 18th July and 20th July 2017 in Braunschweig/Germany. Also, we show a first analysis of the data including the completeness and geometrical quality. The dataset will be published as soon as the coregistration and annotations are complete.

1. OVERVIEW

Different benchmark datasets for autonomous driving were published in the last year, each with slightly different focus. Datasets dedicated to autonomous driving usually provide environment information captured by various sensors equipped on a vehicle. Color and grayscale images, in either mono, stereo or panorama mode, are considered as the main data source by the vast majority of the state-of-the-art datasets (Geiger et al., 2013, Cordts et al., 2016, Wang et al., 2016, Li et al., 2017, Yu et al., 2018, Huang et al., 2018).

However optical sensors cannot always be trusted, especially in adversarial conditions like bad illumination or weather-related visibility issues, and must be complemented by other sensors. This safety-oriented information redundancy is often achieved with LiDAR and RADAR data as in (Wang et al., 2016) and (Ziegler et al., 2014). In addition, this data can be used to reconstruct the 3D environment around the ego-vehicle and accurately localize static and dynamic objects, augmenting a system’s capacity for situation analysis. The most challenging task that ensues is called sensor data fusion. This is a complex task given the wide data diversity, and deep learning has emerged as the most capable technology in this regard, as the leaderboards of the two most renowned benchmarks prove (Geiger et al., 2013, Cordts et al., 2016). Several tasks are being tackled in these benchmarks, among which semantic and instance segmentation, object and instance detection, and, in the case of the KITTI dataset, instance tracking.

Although achieving excellent performance, deep convolutional neural networks (DCNNs) still have difficulties when encountering cases of occlusion. This ubiquitous issue, where for example a large vehicle might hide a smaller one from the ego-vehicle perspective, is a likely cause of danger for road users. Vision systems must learn to detect cues of such risky situations. This is best addressed using a top-down view of the scene to validate the predictions of a DCNN. Some datasets already provide georeferenced aerial optical and LiDAR data matched in localization with the ground imagery (Mattyus et al., 2016, Wang et al., 2016), the benefits of which are limited to static, permanent objects like buildings and road topology. A simultaneous capture of ground and aerial data must be performed to allow for the matching of dynamic objects like vehicles.

With its Application Platform for Intelligent Mobility (AIM) (DLR, 2018), the German Aerospace Center (DLR) has created a research infrastructure for future intelligent transportation and mobility services. AIM enables DLR scientists and partners to model and systematically study an unprecedented range of topics related to intelligent mobility services, covering both multimodality as well as specific modes of transportation. Based on the variety of data and sensors available from the city of Braunschweig, the AIM test field is an ideal target region for the acqui-

[*] Corresponding author

sition of a new benchmark data set.

For the new benchmark dataset DLRAD (**DLR** – **A**utonomous **D**riving), data from stationary, mobile, and airborne sensors are acquired simultaneously and fused together to provide a complete and geometrically accurate view around a reference vehicle, which is driving through the city of Brunswick. The different sensor views are partly complementary and together the data provide a complete picture of a traffic scene. Details of the surroundings are captured from different angles. The view from the helicopter gives a detailed overview of the overall situation, while from the vehicle's perspective one has a rather limited view of the overall situation (see Figure 1), for example, with the airborne view distant and partially concealed objects can be better observed and most of all, they can be geometrically precisely located.

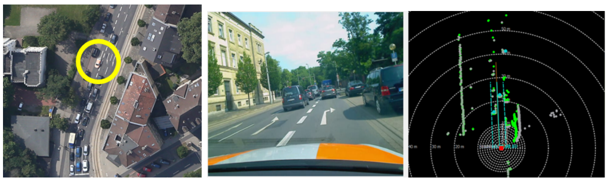


Figure 1. Different views at the same time

As the reference vehicle FAScarE, a Volkswagen eGolf, travels along the test track and scans the environment with the vehicle-based sensors, it is recorded by the optical camera system, the 4k sensor system, which is installed on the DLR Bo105 helicopter. At some positions on the route, the vehicle is also recorded by terrestrial sensors. The route of the reference vehicle was planned to cover different scenarios in the benchmark data set: city areas, motorways, rural roads, industrial and suburbs areas (see Figure 2). The planned benchmark path has a total length of $156km$, with $34km$ urban and suburbs roads, $50km$ rural roads, $26km$ roads in industrial areas and $46km$ motorways, but some parts were optionally planned and some parts of the path were canceled during campaign. A section of about $100km$ was driven in total.

In this paper, a first look on the new DLRAD benchmark dataset is provided by listing all available sensors with a description of their configuration and properties. Potential applications in the field of machine vision are addressed by showing examples of data annotations. Focus will lay on the classification of non-static objects like vehicles and persons, as well as the classification of static objects like lane-markings. The data set will be published when necessary pre-processing steps like coregistration and annotations are completed.

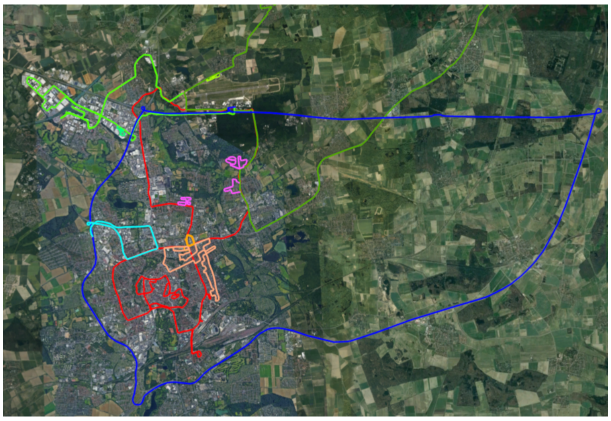
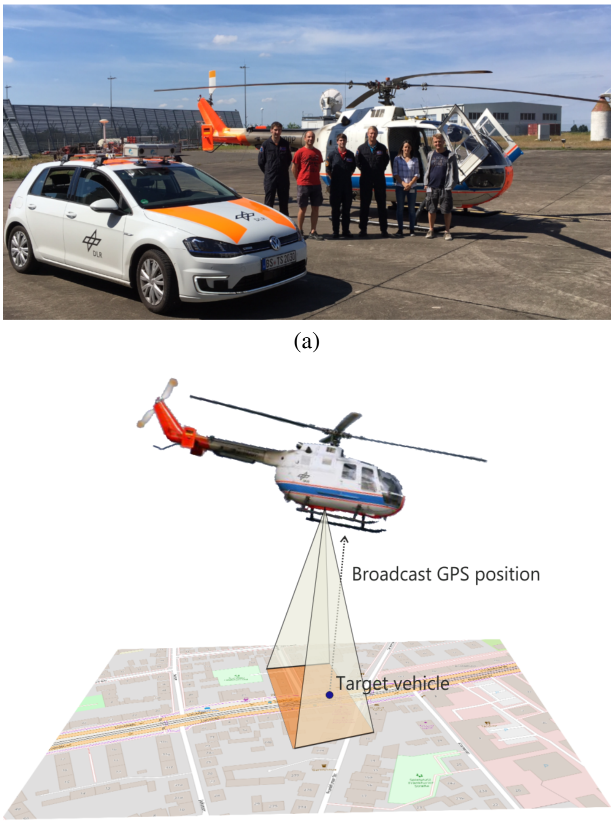


Figure 2. Planned routes of the reference car in and around Brunswick/Germany; blue: motorway, dark green: rural roads, bright green: industrial area, red: city area, orange/magenta: suburbs (from GoogleEarth)



(a)

(b)

Figure 3. FASCarE and BO105 helicopter; (a) team during campaign days from 18th till 20th July 2017, (b) flight setup

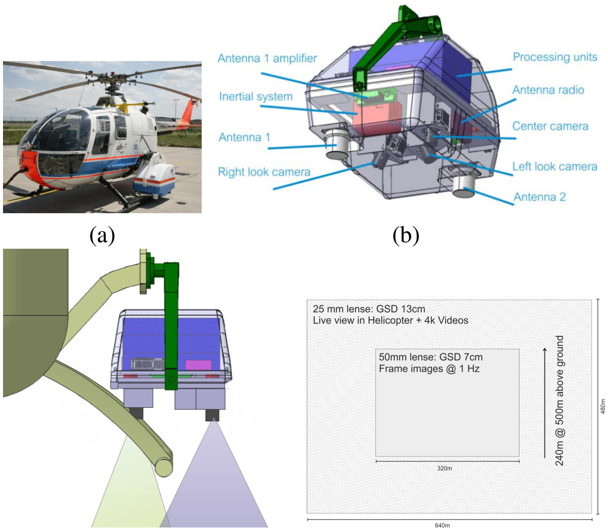
2. SENSOR DATA

With the 4k sensor (Kurz et al., 2014) aboard the DLR helicopter Bo105, the reference vehicle FASCarE (see Figure 3 a)) was recorded while driving the previously defined routes. It was difficult for the helicopter pilots to hover vertically above the moving reference vehicle at all times, as the flight altitude did not allow a direct line of sight to the reference vehicle. Therefore, the current position of the reference vehicle was sent to the helicopter via microwave data link and displayed on a screen (showing a scene map) so that the operator could provide the pilots with instructions (see Figure 3 b)).

2.1 Airborne sensors

The 4k system is designed weight-optimized, small, and relatively low-cost, but equipped with a full real-time image processing chain including a high-capacity data down-link to the ground station. Figures 4 a) and b) show the composition of the 4k system, which consists of three full frame non-metric off-the-shelf cameras, a microwave data-link system including two antennas, three processing units and a GNSS/IMU system (IGI IId). For the benchmark acquisition, two cameras with different focal lengths, $50mm$ and $100mm$, and with looking direction in nadir are used (Figure 4 c)). The footprints of the images cover the area around the reference car staggered according to the distance from the reference car with different GSDs of $7cm$ resp. $14cm$. With a focal length of $50mm$ an area of $320m \times 240m$ is covered assuming a flight height of $500m$ above ground. Assuming that the reference vehicle is at perfect nadir position below the helicopter, the environment around the vehicle is mapped $120m$ resp. $240m$ in forward and backward direction depending on the camera‘s focal length. The image repetition rate was set to 1 Hz during the whole campaign.

Another important database is the 3d surface model along the routes, which are necessary to locate every object in the 3d space. It is possible to generate the 3d surface models (DSM) directly from the acquired aerial images based on structure from motion and using semi-global matching (Hirschmüller and Bucher, 2008, dAngelo and Reinartz, 2011). Based on frame rates around 1 Hz, it will be possible to create a 3D reference map and database with the positions of all moving and non-moving objects around the reference car including pedestrians, cyclists and all kinds of vehicles. Gaps in the 3d surface model caused by occlusions and not-moving helicopter can be filled with a HD surface models derived from prior airborne acquisitions with the 4k camera system.

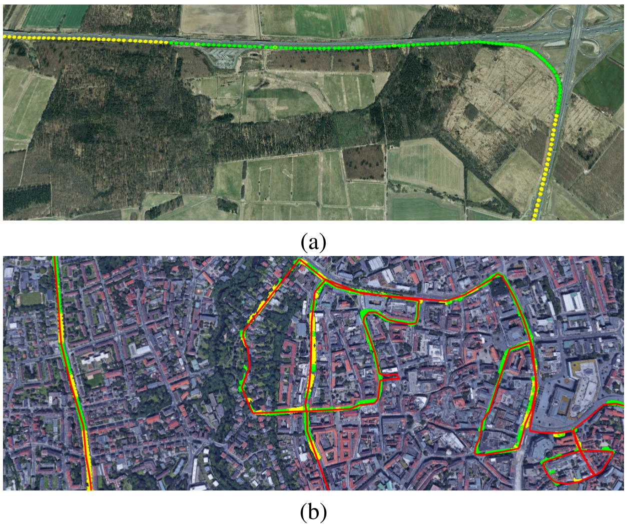


(a) (b)

(c) (d)

Figure 4. 4k sensor on Bo105 helicopter; (a) 4k sensor mounted on multi-purpose carrier, (b) sensor components, (c) viewing geometry and (d) footprints of the two cameras with focal length 50mm and 25mm

The completeness of airborne data acquisition plays an important role. The data sets in which the reference vehicle is located outside the cover of the two cameras cannot be used any longer. Ideally, the reference vehicle is located exactly vertically below the helicopter so that the environment can be mapped with the highest resolution. In Figure 5, the achieved completeness with regard to the coverage of the reference vehicle from the aerial images is illustrated for two scenarios. The colored points in green, yellow and red mean the reference vehicle is mapped in optimal resolution and in the center nadir direction (green), outside the footprint of the long focal length, but still mapped (yellow), and not mapped at all (red).



(a)

(b)

Figure 5. Completeness of image acquisition (examples) on motorways (a) and over the city (b); green and yellow dots indicate the reference vehicle within the footprint of the camera with 50mm resp. 25mm lense; gaps (in red) indicate the reference vehicle outside the footprints

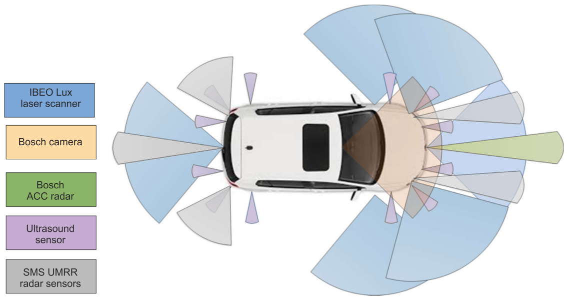


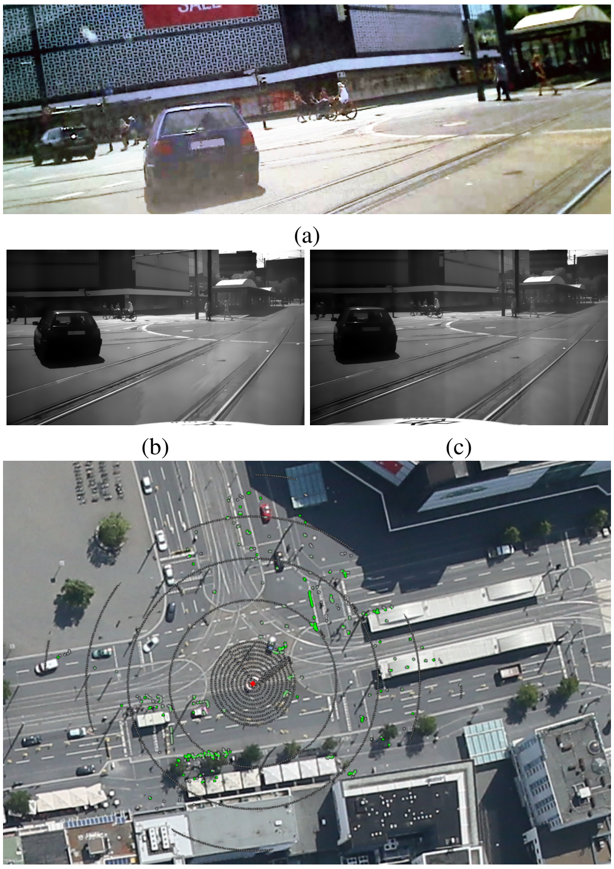
Figure 6. Position and field of view of vehicle based sensors

On motorways the completeness was relatively high (61% green, 21% yellow, 18% red) (see Figure 5 a), as the reference vehicle was traveling with more or less constant speed in a manageable environment. At rural roads the completeness (43% green, 27% yellow, 30% red) was low, as the helicopter pilots had to cope with strong wind and low altitude cloud cover on the one hand, as well as with temporary problems with the aboard map for guiding the pilots. Similar situation in the urban scenarios, even though the completeness (45% green, 44% yellow, 11% red) is relatively high, the reference vehicle is often mapped off nadir. The reasons are, that the pilots had problems to follow the reference vehicle due to the complex terrain in cities, long waiting times at traffic lights and necessary manoeuvers to avoid overheating of the motor in the helicopter.

2.2 Vehicle sensors

The reference vehicle FASCarE is a Volkswagen eGolf. It is equipped with different range detectors as follows: four rear and six front ultrasound sensors for the close range detection smaller than 5m, three (five planned) front and one rear IBEO laser scanner with range up to 70m, each two front and three rear SMS radar, and one front Bosch radar (see Figure 6). Also a stereo camera system is installed at the car roof for 3D and object detection purposes. For scene overview, there is a webcam behind the front window.

The calibrated stereo camera system acquires 1380×800 pixel gray value images with a frame rate of 10 Hz and it is synchronized with the a board GNSS/Inertialsystem (see Figure 7). The setup can be used to calculate a depth map or to use the single frame images for classification purposes.



(a)

(b) (c)

(d)

Figure 7. Screen shot from video (a) and image examples from stereo cameras left (b) resp. right (c) acquired from the FASCar reference vehicle. (d) shows the corresponding aerial image overlaid with LIDAR reflectance points (green dots) from the reference vehicle.

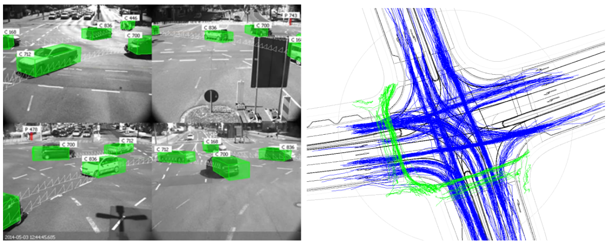
Other vehicle based data are read out from the car bus, e.g. steer angles, pedal actions and gear shift. A smartcam from Mobileye and from eGolf detects traffic lanes and moving obstacles like pedestrians and cyclists.

2.3 Stationary sensors

This database will be augmented with the data from the stationary sensors at the AIM research intersection to have a more detailed view at defined sections. The stationary sensors are installed on gantries at a main crossing. There are various viewing angles that cover the entirety of the inside area of the intersection and also partly the roads that lead to the intersection. The system uses mono and stereo camera systems and radar sensors to detect and classify all moving and stationary traffic participants (cars, trucks, buses, pedestrians, bicyclists) in real-time. The trajectories (containing position, velocity, heading, size, object type) of all objects are recorded. Also available is a low-resolution scene video stream for every viewing angle. Furthermore, data from traffic signals at this particular intersection is available. Figure 8 shows examples of detected objects and trajectories at the research intersection.

2.4 Other data

Auxiliary data set from other sources are quite useful as additional reference. A set of geodetic SAR points in and around Brunswick (Gisinger et al., 2017), which are derived from TerraSAR-X satellite, can be used as very accurate ($< 10cm$) absolute reference points. The points can be used to further improve the geolocalization of the aerial images.



(a) (b)

Figure 8. (a) Examples of detected objects and (b) trajectories of vehicles (blue) and passengers (green) at the research intersection

2.5 Summary of sensor data

In Table 1, the specifications of airborne and vehicle based sensors are listed including a small product description. From a geometrical view point, the ranges of the vehicle based sensors are covered by the aerial images, and more important, the geolocalization accuracy of the aerial maps outperforms any vehicle based sensor. The optimal accuracy of around $10cm$ can be reached, if geodetic SAR points are within the image footprints, in regions without these points, absolute accuracies around $30cm$ can be reached (Fischer et al., 2017). Thus, the airborne imagery could be used as reference for the validation of any vehicle based sensor data classification in terms of a geometrical and semantic quality analysis. Nevertheless, every classification method requires geometrically accurate coregistered data from aerial and vehicle based sensors. Thus, the coregistering of all data and semantic annotations are the next important processing steps before publication of the benchmark data set.

3. ANNOTATIONS

Annotations are planned for the aerial images as well as for vehicle based sensors in pixel-wise format in which each pixel is assigned to a specific category. All non-static and most of the static object will be annotated as illustrated in Figure 9 for aerial images. In this example 22 classes are labeled: eight vehicle classes, eight road classes, lane markings, buildings, two vegetation classes, impervious surface and clutter.

Annotations of the vehicle based sensors comprise non-static objects like vehicles, pedestrians as well as static object like lane-markings and road areas. In addition, sidewalks, bike path, and parking places can be annotated containing distinguished paved and non-paved areas like in Figure 10. This example with 28 classes was semi-automatically generated using a pre-trained FCN (Long et al., 2015) and data from (Neuhold et al., 2017).

Based on these annotations, the development of automatic classification methods will be possible separately and in combination for air and vehicle-based sensors.

	Data	**Specifications**
4k sensor	HD 2D orthophoto map HD 3D model of static objects	True ortho; $\sigma_{XY} < 0.1m^*$; $GSD = 0.7cm^{**}$ @1 Hz SGM derived DSM $\sigma_{XYZ} < 0.1m^*$; $GSD = 20cm$
Stereo camera	Depth map Front view frame images	$b = 0.45m \rightarrow @25m : \sigma_{dist} = 1.1m$ @10 Hz 1360 × 800 px (cropped, 8bit, gray value) @10 Hz
Lidar 4× IBEO Lux 4L	Cloud of measurement points	4 layers, range $< 50m$ @10$\%permission$ $\sigma_{dist} = 10cm, FOV = 110^\circ_h/3.2^\circ_v$ @25 Hz
Radar 5× SMS UMRR ***	Object distance, angle and speed	$dist < 70m, \sigma_{dist} < 0.5m\ or\ 1\%$ $FOV = 130^\circ_h/15^\circ_v$ @20 Hz
Radar 1× Bosch ACC	Object distance, angle and speed	
Ultrasound 10× ***	Object distances in the near range	$< 5m$
GNSS/IMU Novatel SPAN CPT	Reference vehicle position and speed	corrected by SAPOS
Video Bosch Smart-camera	Lane detection, detection of traffic signs and other objects	
Video Mobileye Smart-camera	Lane detection, detection of traffic signs and other objects	
Video Scene camera	front view videostream	VGA 640 × 480 video stream @22 Hz
Vehicle Data	Data from vehicle bus like stearing wheel angle, pedal activity, indicator lights	up to 100Hz
Research Intersection	Video streams from 8 mono cameras and 7 stereo cameras; trajectory data from all objects inside detection area	25Hz

* using geodetic SAR points
** based on focal length of 50mm
*** not present in dataset

Table 1. All sensors (first row is airborne, the others are vehicle based) with their specifications

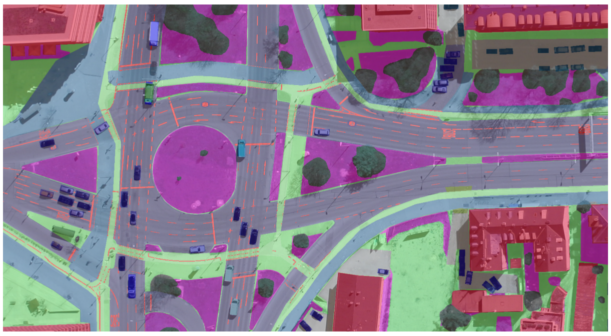


Figure 9. Example of pixel-wise annotations with 22 classes containing both static and non-static objects.

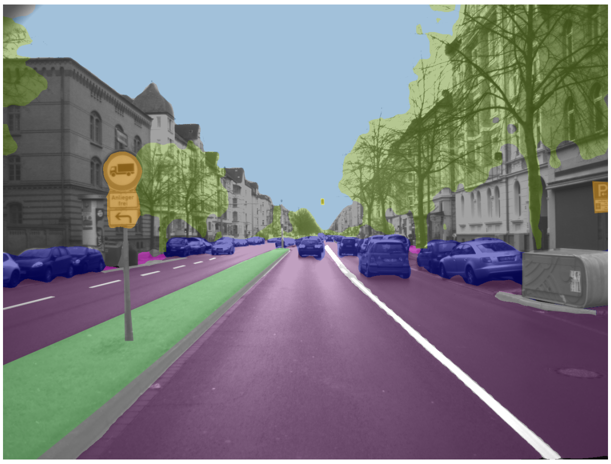


Figure 10. Example of pixel-wise annotations with 28 classes containing both static and non-static objects.

4. FUTURE WORK/APPLICATIONS

Before the data is published, the following processing steps are required:

- Geometrical accurate coregistering of all data in one mapping frame, which is prerequisite for any further developments
- Annotations of aerial and vehicle based sensor data

Possible applications and research priorities are outlined below:

- Geometrical validation of vehicle based GNSS/IMU systems by airborne images and geodetic SAR points
- Validation of vehicle based multi-sensor navigation/localization approaches by airborne imagery
- Development of combined airborne and terrestrial mobile mapping approaches
- Joint utilization of aerial and ground sensor data to perform SLAM, road topology estimation and road segmentation, and urban zoning classification (residential, commercial, etc.)

ACKNOWLEDGEMENTS

The authors would like to thank the helicopter pilots and operators Sebastian Soffner, Uwe Göhmann, Markus Bernhardt and Timo Kausche from DLR flight experiments, who had to brave the partly adverse weather conditions like wind and low clouds and during long hovering flights with overheating of the engine.

REFERENCES

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S. and Schiele, B., 2016. The Cityscapes Dataset for Semantic Urban Scene Understanding. In: *CVPR*, Las Vegas.

DLR, 2018. Application platform for intelligent mobility, http://www.dlr.de/ts/en/desktopdefault.aspx/tabid-6422/. [Online; accessed 20-June-2018].

dAngelo, P. and Reinartz, P., 2011. Semiglobal matching results on the isprs stereo matching benchmark. *High-Resolution Earth Imaging for Geospatial Information*.

Fischer, P., Plaß, B., Kurz, F., Krauss, T. and Runge, H., 2017. Validation of hd maps for autonomous driving. *International Conference on Intelligent Transport Systems in Theory and Practice, mobil.TUM*.

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

Gisinger, C., Willberg, M., Balss, U., Klügel, T., Mähler, S., Pail, R. and Eineder, M., 2017. Differential geodetic stereo sar with terrasar-x by exploiting small multi-directional radar reflectors. *J Geod*.

Hirschmüller, H. and Bucher, T., 2008. Stereo processing by semiglobal matching and mutual information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(2), pp. 328–341.

Huang, X., Cheng, X., Geng, Q., Cao, B., Zhou, D., Wang, P., Lin, Y. and Yang, R., 2018. The apolloscape dataset for autonomous driving. *arXiv: 1803.06184*.

Kurz, F., Rosenbaum, D., Meynberg, O., Mattyus, G. and Reinartz, P., 2014. Performance of a real-time sensor and processing system on a helicopter. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Denver, USA, pp. 189–193.

Li, Y., Wang, J., Xing, T., Liu, T., Li, C. and Su, K., 2017. Tad16k: An enhanced benchmark for autonomous driving. In: *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 2344–2348.

Long, J., Shelhamer, E. and Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: *CVPR*.

Mattyus, G., Wang, S., Fidler, S. and Urtasun, R., 2016. HD Maps: Fine-Grained Road Segmentation by Parsing Ground and Aerial Images. In: *CVPR*, Las Vegas.

Neuhold, G., Ollmann, T., Bulò, S. and Kontschieder, P., 2017. The mapillary vistas dataset for semantic understanding of street scenes. In: *ICCV*, Venice.

Wang, S., Bai, M., Mattyus, G., Chu, H., Luo, W., Yang, B., Liang, J., Cheverie, J., Fidler, S. and Urtasun, R., 2016. TorontoCity: Seeing the World with a Million Eyes.

Yu, F., Xian, W., Chen, Y., Liu, F., Liao, M., Madhavan, V. and Darrell, T., 2018. BDD100K: A Diverse Driving Video Database with Scalable Annotation Tooling. *CoRR*.

Ziegler, J., Bender, P., Schreiber, M., Lategahn, H., Strauss, T., Stiller, C., Thao Dang, Franke, U., Appenrodt, N., Keller, C. G., Kaus, E., Herrtwich, R. G., Rabe, C., Pfeiffer, D., Lindner, F., Stein, F., Erbs, F., Enzweiler, M., Knoppel, C., Hipp, J., Haueis, M., Trepte, M., Brenk, C., Tamke, A., Ghanaat, M., Braun, M., Joos, A., Fritz, H., Mock, H., Hein, M. and Zeeb, E., 2014. Making Bertha Drive - An Autonomous Journey on a Historic Route. *IEEE Intelligent Transportation Systems Magazine* 6, pp. 8–20.