Investigation of Lidar Data for Autonomous Driving with an Electric Bus

Electric mobility and automated driving are merging together. Bertrandt investigates the measurement data of a laser scanner (lidar technology), which recorded the environment of an electric city bus 4-km route through Regensburg (Germany). On the basis of these data, two methods of object detection and classification were examined. One of the challenges was the low vertical resolution compared to the frequently used 32 to 128 lines.

TESTS WITH AN ELECTRIC CITY BUS

From May to September 2018, Bertrandt had the opportunity to use its Emil electric city bus from Regensburg in live operation as a research platform. The aim was to show the challenges for future autonomous driving in an urban environment and to collect the corresponding data. Since the beginning of the project, an interdisciplinary team has been working intensively on the topics of localization, data transfer, lidar image processing, and object recognition using machine learning.

After around 1800 test drives, the data pool amounts to 14 TB, which was collected over 120 days while driving a total distance of 5800 km. Before the e-bus of the Italian company Rampini Carlo S. p. A. began recording data, additional sensorics were installed in the bus. These include essentially a Quanergy M8 lidar and a GNSS receiver with an integrated

AUTHORS

Dipl.-Ing. Christian Feller
is Development Engineer for Information Technology at Bertrandt in Regensburg (Germany).

Dipl.-Ing. Ulrich Haböck
is Leader for Strategy and Innovation at Bertrandt in Regensburg (Germany).

Stefan Maier, MCSE, MCITP
is IT Architect and Connectivity Specialist at Bertrandt in Regensburg (Germany).

Dipl.-Ing. Jochen Schwenninger
is Chief Engineer and Function Architect for ADAS Functions and Systems at Bertrandt in Regensburg (Germany).

Inertial Measuring Unit (IMU) from Hexagon. A processor unit (Intel i7 kernel and Nvida GPU) was responsible for controlling the sensors, recording the data, and in some cases transferring the data in real time via LTE to the backend (Microsoft Azure Cloud). The completely independent system without connection to the vehicle bus system or the driver was powered by the 24-V vehicle electrical system. The sensor data were recorded and processed using the ROS software framework [1] in the "Kinetic Kame" version. **FIGURE 1** shows the entire architecture of the experimental setup with built-in hardware.

CHOOSING THE SENSOR

A lidar system, when used in sensor fusion for autonomous driving, offers an interesting mixture of spatial resolution and depth information, and therefore supplements the more traditional sensors such as radar and camera. The lidar sensor, which has a range of 150 m at 80 % reflectivity, made it possible to repeatedly record the surroundings that the bus drives through several times a day on its approximately 4 km long route through Regensburg's historical city center. The result is a grayscale image (reflectivity) with eight lines and up to 10,000 points per line as well as the corresponding depth information (distance).

As the bus itself was already very tall, the sensor was not mounted on the roof but on the front of the vehicle at a height of around 3 m. For that reason, a visibility of only 180° is usable and not the full 360°. On the basis of these data, two methods of object detection and classification were examined. One of the challenges was the low vertical resolution compared to the frequently used 32 to 128 lines that are provided by the significantly more expensive reference sensors used in automotive applications.

OBJECT DETECTION VIA TWO ALGORITHMS

The main intention was to detect and classify relevant objects in the bus's surroundings. The classes "Vehicle" and "Pedestrian" were considered. Due to the high variability of the data, approaches from machine learning were applied. Training required data which, in addition to the actual sensor data, also contain the desired result of the object detection process, so-called labels. These are usually given in the form of bounding boxes which completely enclose the object. The Kitti 3-D object detection dataset was used as the source for the training data [2], **FIGURE 2**.

In order to measure the quality of the process, the labels generated by the respective algorithm were compared to manual reference annotations. If there is a sufficiently high agreement of the bounding boxes, the prediction is recognized as true-positive; otherwise, the object is registered as an error. Processes in the Kitti environment are normally evaluated using the Mean Average Precision (MAP).

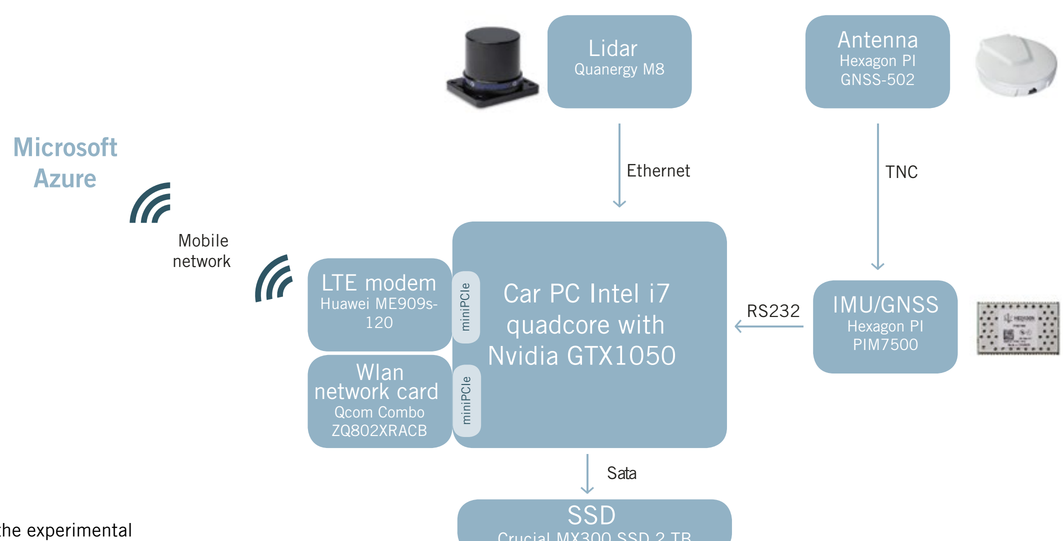


FIGURE 1 Entire architecture of the experimental setup with built-in hardware (© Bertrandt)

Two algorithms in a direct comparison were used in order to detect the objects. One approach is based on Yolo 2-D object detection [3]. The second approach implements and adapts the ideas from VoxelNet for the lidar sensor used [4].

The fundamental approach of the Yolo-based algorithm is to generate a 2-D representation from the 3-D point cloud and then to use an established approach from image processing to detect the desired objects in this representation. First of all, the frontal view was implemented. Following the 2-D detection, the missing coordinates from the underlying point cloud were determined in order to give a genuine 3-D object. The advantage of this approach is that intermediate results, for example the 2-D projection or the 2-D objects, can easily be manually checked and, when selecting suitable neural networks, it was possible to make use of the already extensively evaluated architectures from classical image processing as neural networks.

For the VoxelNet-based approach, an occupancy grid is generated from the point cloud. The grid divides the 3-D area around the sensor into cubes and, with the aid of neural networks, determines different properties or features for each cube (dimensions, for example 0.4 × 0.2 × 0.2 m), such as the number of points in the cube or their typical distance from each other. This grid, combined with the properties, forms the input data for a further neural network (Region Proposal Network), which is trained with typical contours and surroundings. At the end of this processing, both the class of a detected object and its 3-D bounding box are made available.

For both processes, the data for training and evaluation from Kitti were adapted to the sensor used. This includes, among others, the selection of the eight most similar beams from the 64 available as well as the masking of the objects in the environment not visible as a result and the correction of the sensor installation height, **FIGURE 3**. Therefore, due to the reduced information, a deterioration of the detection results can be expected compared to a process based on the complete data.

At the end of the project, the two implemented algorithms were compared with the best results in the official benchmark for the detection of vehicles in a bird's eye view. The UberTAG-HDNet [5] of the University of Toronto, which uses additional training data in order to learn typical maps and therefore to enable better masking of the static environment, generates a MAP of 89 % for completely visible vehicles (Kitti Easy).

The Bertrandt innovation project, which had a duration of only six months, achieved an MAP of 59 % for the same vehicles using the VoxelNet approach with eight height layers. The alternative approach with Yolo still delivered a respectable 34 % MAP without determining the angle of rotation of the objects in the XY plane. This trend can also be observed in the task of pedestrian detection. The University of Waterloo (AVOD-FPN [6]) achieved the best results so far for this task with Kitti data (lidar and camera): 59 % of the simple data were correctly classified. In the project presented, the VoxelNet approach achieved 41 % MAP with the data reduced to eight height layers, while Yolo achieves an accuracy of 19 % without estimating the angle of rotation.

FUSION OF LIDAR INFORMATION

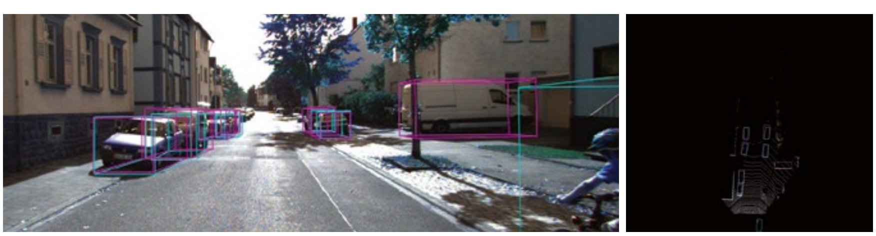
With the aid of the data collected, object detection can be optimized even further. New approaches can be evaluated without the need for recording new test drives. In particular, the time sequence of the lidar frames can be used to obtain a more accurate image of the surroundings and to check the plausibility of potentially detected objects or to reject them. The fusion of lidar information with other sensor data will also be necessary for future automated driving.

LOCALIZATION IN URBAN ENVIRONMENT

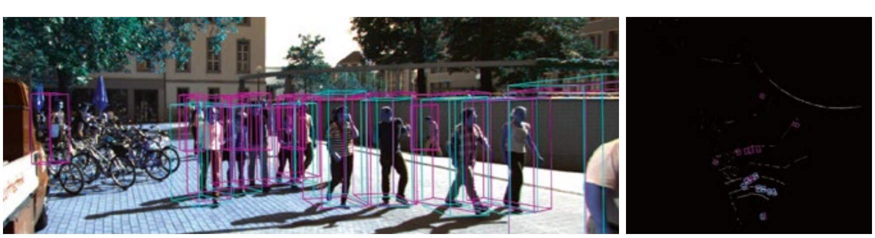
A further focus of the project was on localization. Satellite-based navigation systems such as with GPS or Galileo have a positioning accuracy of around 10 m without using correction data. In urban environments, streets with high buildings pose a problem, as it is not always possible to have a continuous clear view to the sky, with the result that there is an increase in typical positioning errors. To counteract this, the vehicle's motion can be tracked with the aid of an IMU and this data can be used together with the GNSS position data. Unfortunately, even minor errors in the measured values result in a drift in the position, which means that this approach does not provide a long-term stable position.

By contrast, simultaneous localization and mapping integrate a second absolute source of information with the static environment in order to determine the position. These systems use sensor information not only to determine the position of the vehicle but also to generate a map of the surroundings, **FIGURE 4**. The basic idea is that changes in the distances from walls or corners are a sign of motion and are therefore an indication of a change in position. Various implementations, for example for mobile robots, are already in use.

Kitti training data



Cars are reliably recognized



Pedestrians nearby can be recognized

FIGURE 2 Examples for the lidar data in the Kitti benchmark: The reference labels are displayed in magenta, the automatically generated labels in turquoise (© Bertrandt)

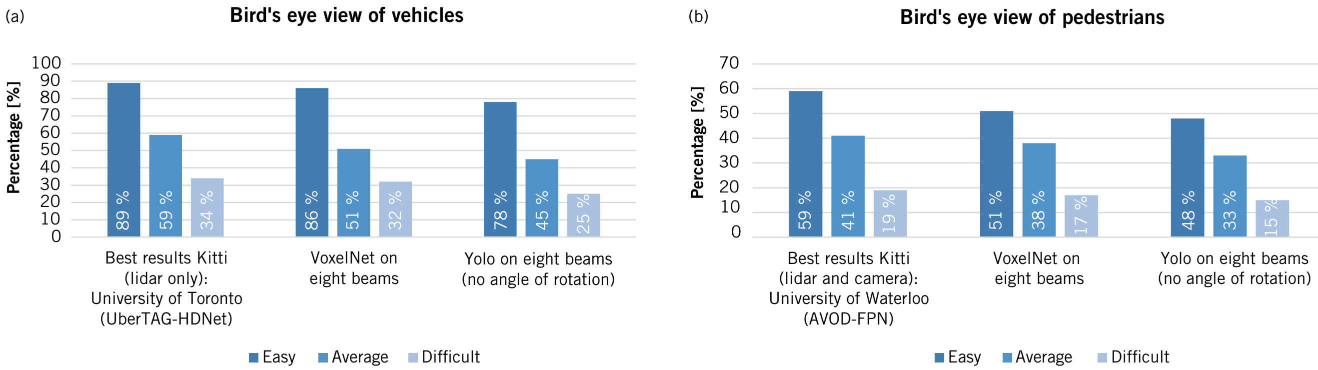


FIGURE 3 Results for the detection of vehicles (a) and pedestrians (b): The degree of difficulty depends, among other things, on the occlusion and size of the objects; the reference algorithms work on all 64 beams; VoxelNet and Yolo use only eight beams (© Bertrandt)

Several methods were evaluated during the project duration, but as yet these have not been capable of dealing with the relatively dynamic environment of an inner-city area. Therefore, an image of the surroundings was generated by projecting the lidar impressions on the respective GNSS/IMU position and then aggregating them over a complete test drive. This representation can then be further refined and made more precise over several test drives until an accurate map of the route has ultimately been generated.

CHALLENGE OF DATA TRANSFER

A further key point that became apparent during the project was that the network connection in Regensburg is basically stable, but reproducibly deteriorates at the same places each time. In these situations, latencies of more than 1000 ms were observed, while at the same time the usable data rate fell to below 4000 kbit/s, **FIGURE 5**. The reasons for this were, among other things, overloaded network cells at busy locations and narrow streets with tall buildings in Regensburg's historic city center.

FIGURE 4 Automatically generated representation of the vehicle environment in real time on a journey (therefore reduced resolution); in the background a satellite image for comparison purposes (© Bertrandt)

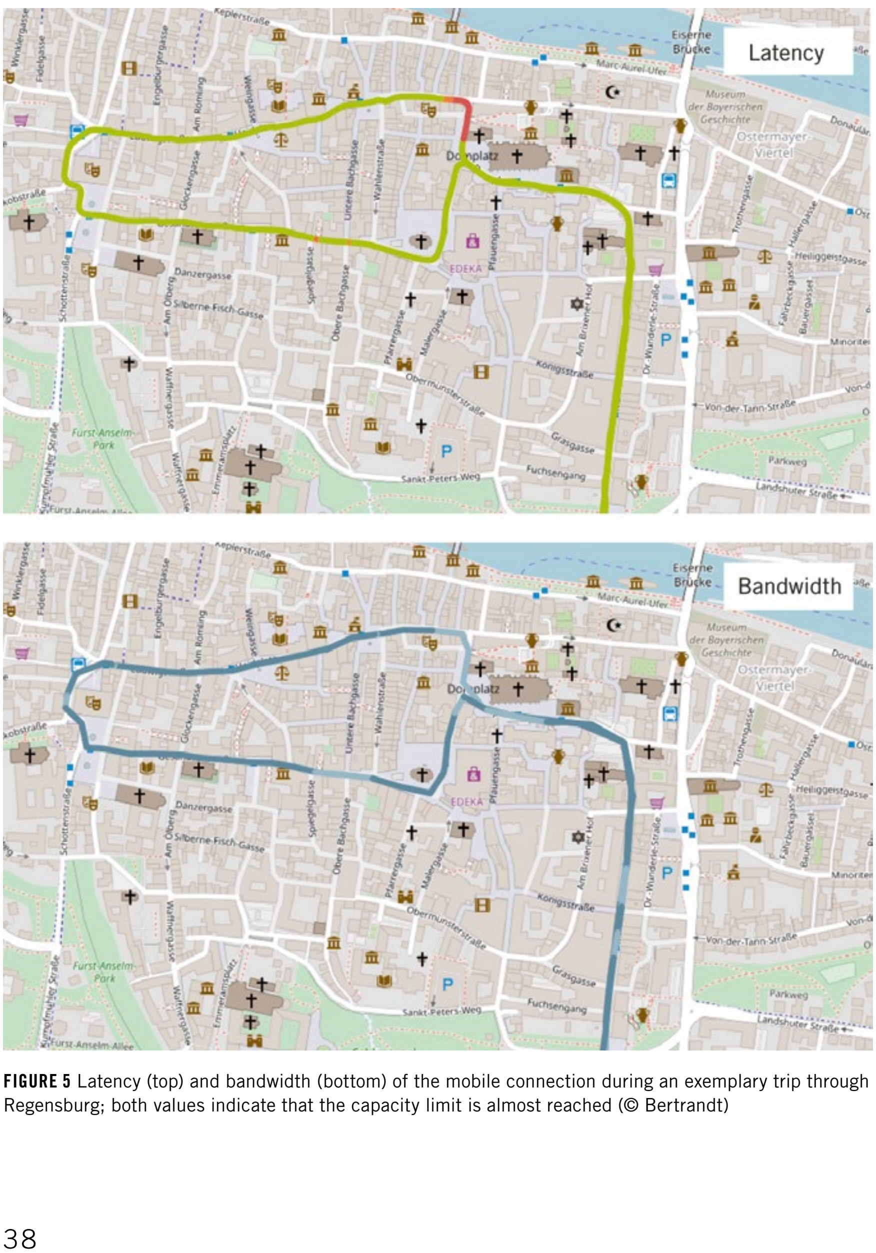


FIGURE 5 Latency (top) and bandwidth (bottom) of the mobile connection during an exemplary trip through Regensburg; both values indicate that the capacity limit is almost reached (© Bertrandt)

VEHICLES AS PRODUCERS OF DATA

If we take a look at the future of data transfer, a major advantage of (semi-) autonomous vehicles will be the fact that they can share large amounts of data about their sensor impressions and planned maneuvers with other vehicles in the vicinity and with backend systems. Even today, vehicles are increasingly becoming producers of data.

However, the current 4G network infrastructure is not powerful enough to transfer this data reliably for a large number of vehicles. For that reason, the focus is increasingly on the upcoming 5G standard, which not only offers greater bandwidths but also the possibility to broadcast information. In addition to car-to-X connectivity, another application case is the distribution of correction data for satellite-based positioning solutions.

The raw data acquired in the innovation project were already collected and evaluated by internal project groups during ongoing operation with the aim of establishing greater expertise. In keeping with its Open Innovation approach, Bertrandt intends not only to share the findings with customers but also to further develop them with cooperation partners.

REFERENCES

[1] OSR Foundation: Software framework of the Robot Operating System (ROS). Online: https://www.ros.org/, access: November 9, 2018

[2] Geiger, A.; Lenz, P.; Urtasun, R.: Are we Ready for Autonomous Driving? The Kitti Vision Benchmark Suite. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence (USA), 2016

[3] Redmon, J.; Diwala, S.; Girshick, R.; Farhadi, A.: You Only Look Once (Yolo): Unified, Real-Time Object Detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas (USA), 2016

[4] Zhou, Y.; Tuzel, O.: VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City (USA), 2018

[5] Yang, B.; Liang, M.; Urtasun, R.: HDNet: Exploiting HD Maps for 3D Object Detection. In: 2nd Conference on Robot Learning (CoRL), Zurich (Switzerland), 2018

[6] Ku, J.; Mozifian, M.; Lee, J.; Harakeh, A.; Waslander, S.: Joint 3D Proposal Generation and Object Detection from View Aggregation. In: International Conference on Intelligent Robots and Systems (Iros), Madrid (Spain), 2018