ACTIVE SHAPE MODEL PRECISION ANALYSIS OF VEHICLE DETECTION IN 3D LIDAR POINT CLOUDS

S. Busch[1]

[1] Institute of Cartography and GeoInformatics, Leibniz Universität Hannover, Germany - busch@ikg.uni-hannover.de

Commission I, WG I/5

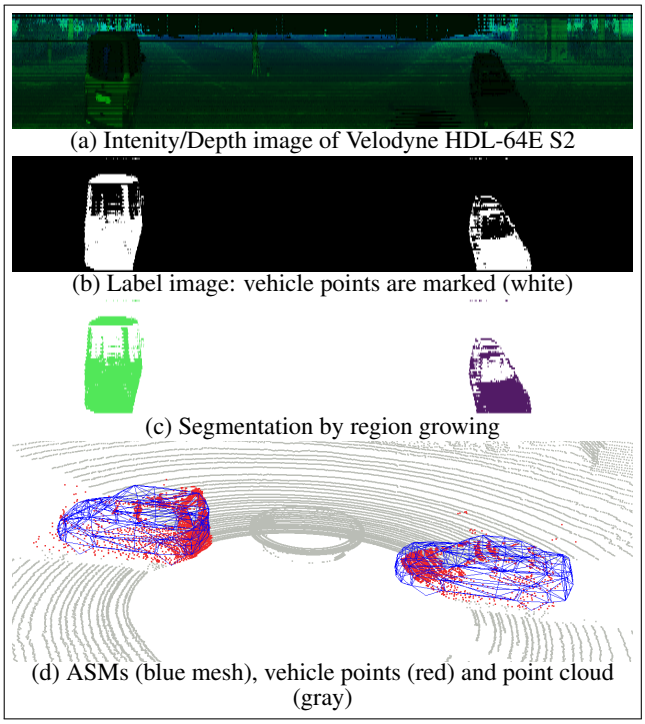
KEY WORDS: Active Shape Model, LiDAR, Vehicle Detection, Point Cloud, Pose Estimation, Segmentation

ABSTRACT:

LiDAR systems are frequently used for driver assistance systems. The minimal distance to other objects and the exact pose of a vehicle is important for ego movement prediction. Therefore, in this work, we extract the poses of vehicles from LiDAR point clouds. To this end, we measure them with LiDAR, segment the vehicle points and extract the pose. Further, we analyze the influence of LiDAR resolutions on the pose extraction by active shape models (ASM) and by the center of bounding boxes combined with the principal component analysis (BC-PCA).

1. INTRODUCTION

The accurate prediction of traffic participant behavior is essential for avoiding accidents. Driver assistance systems and especially, autonomous driving use a variety of sensors like cameras and LiDAR sensors to detect traffic participants and predict their behavior. The prediction strongly depends on the observation accuracy and variance. Nowadays, more and more LiDAR systems are used for driver assistance, because of their high distance accuracy. The 3D point information is used to improve the camera based object detection. An object enclosing bounding box could be used for deriving a precise pose for the object. We analyze the accuracy of pose estimation approaches for these objects by comparing the extracted poses of vehicles to highly accurate total station references. We validate the pose accuracy to the effect of distance, viewing angle, different resolutions and vehicle shapes. Therefore, we use a Velodyne HDL-64E S2 and Velodyne VLP-16 scanner, with 64 and 16 vertical beams and a vertical resolution of 0.4°and 2°. The vehicle shapes have different effects on the detected bounding boxes, due to the count of measured points and occlusion. We analyze the impact of vehicle shapes on the bounding box by comparing the pose accuracy of a sedan type car to a van. In addition, we compare the accuracy of bounding box estimation approaches by BC-PCA and ASMs. An ASM (Cootes et al., 2000) estimates the position by the geometric center of a deformable vehicle model and uses its orientation. The BC-PCA uses the center of the enclosing bounding box for the position and the main component of the PCA as orientation. In more detail, we extract the vehicle points from the scans by using a region of interest and removing the ground by subtracting the ground plane estimated via random sampling consensus (RANSAC). The remaining points are clustered to vehicle points by region growing. For each vehicle, two poses are estimated by bounding box centers and orientations derived from BC-PCA and ASM, see Figure 1. To overcome the restriction of the region of interest and plane estimation in future works we present a neural network. We used the proposed segmentation technique to generate 200.000 training samples from 6 different junctions in the city of Hannover, Germany. For the accuracy analysis, we scan two vehicles with two LiDARs simultaneously from 24 different poses and build a global coordinate frame, where the relative poses of vehicles,



(a) Intenity/Depth image of Velodyne HDL-64E S2

(b) Label image: vehicle points are marked (white)

(c) Segmentation by region growing

(d) ASMs (blue mesh), vehicle points (red) and point cloud (gray)

Figure 1. Data processing: from intensity/depth image via label image to ASM

scanners and total stations are known. The paper is structured as follows: first, section 2 gives an overview over object detection approaches. Secondly, we explain our pose extraction process in section 3 and our experiment in section 4, before we present the accuracy validation in section 5. Finally, we end with a conclusion in section 6.

2. RELATED WORK

In this section we give an overview of related work for object detection in LiDAR point clouds by use of objects segmentation and object pose estimation approaches. Some point

cloud segmentation approaches remove the ground plane by local plane estimation and cluster the remaining points by occupancy grid segmentation (Himmelsbach et al., 2010; Douillard et al., 2011). Other approaches calculate the normal vector for each point and use region growing (Rabbani et al., 2006) or graph cut (Moosmann et al., 2009) methods. All of these geometric feature based segmentation approaches need certain thresholds which must be tuned for optimal results. However, these results create a database for the training of neural networks (NN) (Li et al., 2016; Zhou , Tuzel; Qi et al., 2017). NN segmentation for object detection approaches often combine LiDAR data with camera images (Qi et al., 2018; Barea et al., 2018; Xu et al., 2018; He , Soatto). After the objects are segmented, their pose can be estimated by a bounding box. He (Soatto) and Chen et al. (2018) estimate the bounding box directly using a NN. Nevertheless, a part of the bounding box is often occluded from the object itself. The use of ASMs (Coenen et al., 2018; Ferryman et al., 1998; Menze et al., 2015; Zia et al., 2011, 2015) is a suitable approach to overcome the occlusion issue and can improve the accuracy of a vehicle bounding box. Many approaches validate the algorithm with a public reference data set based on camera labeled images. Up to now and to the best of our knowledge, there is no validation of LiDAR detected vehicle poses by a reference with superior accuracy. Our work addresses this validation gap by determining the accuracy of the pose estimation by comparing pose detections to a total station.

3. DATA PROCESSING

First we pre-process the recorded point clouds from the Velodyne HDL-64 and VLP-16 to intensity/depth images with a resolution of 64×1500 and 16×1500 pixels, respectively, see Figure 1 (a). The information of each beam is stored in a row. We label the pixels in the images as vehicle pixels and segment single vehicles. Finally, we estimate their poses.

3.1 Classification

We use two approaches to classify the scan points and corresponding pixels of the intensity/depth images as vehicle points. On the one hand, we use information about lanes to filter background pixels and filter the road points/pixels by ground subtraction. On the other hand, we use a neural network to label vehicle points without information about lanes and ground plane. For the *lane and ground filtering* all points of the different perspectives are transformed into the frame of the total station to use a manually determined lane for point filtering. We approximate a lane by picking points along a line through the vehicle positions. This line represents a middle axis of a lane with the width of 3 m. We estimate the ground plane of this lane by RANSAC with a 10 cm threshold. All points within the lane and 10 cm above this plane are marked as vehicle points, see figure 1 (b). We trained the neural network for vehicle labeling to overcome the dependence on lane accurate maps during the labelling process. The network is inspired by the VGG16 (Simonyan , Zisserman) structure. For a pixel-wise classification we add deconvolution layers, three with a *rectified linear unit* activation function and one final deconvolution with a *linear* activation function. The training data is generated by label measurements of six junctions in Hannover via the mentioned *lane and ground filtering*. In addition we use the scan and label data provided by the KITTI benchmark (Geiger et al., 2013).

3.2 Segmentation

We use the pixel information of the intensity/depth images, as well as the 3D-coordinates to assign points/pixels to vehicles. A region growing algorithm uses a 5×5 pixel neighborhood to overcome measurements errors. Pixels which are marked as vehicle points are added to a set of seed points. One seed point is selected randomly and each seed point within its pixel neighbourhood with a distance below 1 m is added to its region and removed from the seed point set. The region growing will stop if there are no more seed points, see Figure 1 (c).

3.3 Pose detection

The main focus of the work is an accurate pose estimation of detected vehicles. In contrast to the BC-PCA approach, which uses the scanned vehicle points, the ASM approach uses the derived shape points. Both approaches use the min and max values of the x-, y- and z-coordinates of the points to determine a bounding box and estimate the position of the vehicle by the center of the bounding box. The BC-PCA estimates the heading Θ of the vehicle by using the eigenvector e with the highest eigenvalue:

$$\Theta = \text{atan2}\,(e_y, e_x) \tag{1}$$

The heading of the ASMs, blue triangles Figure 1 d, is calculated by the ASM optimization (Coenen et al., 2018). We use the database from Coenen et al. (2018) with 30 cars and 2 vans as well as their particle optimization. The shape is optimized by four eigenvalues $\sigma_1 - _4$ and the pose by a 2D-transformation. The z-coordinate is fixed at the lowest z-value of the scan points and the pitch and roll are not considered by assuming vehicles driving on the ground. The particle optimization (Coenen et al., 2018) changes the shape and pose of the model for each particle and keeps the n-th best particle at each iteration. We calculate the score for each particle by the log-likelihood (the squared mean error) of a detected pose to the nearest triangle and use an occupancy voxel grid to punish free voxels inside the ASM bounding box. Voxels will be marked as free if the corresponding pixel of the back projected voxel center has a bigger distance than the voxel.

4. EXPERIMENT

We used a sedan type car and a van for our experiment. The bounding box of the sedan type car is 4.5×1.78×1.52 m, in contrast to the van, which is 4.9×1.9×1.9 m. Figure 2 shows

Figure 2. The setting for the measurements: the total station (left), both cars (in the back) and the scanners at the dynamic rack (center)

the rack with the two LiDAR Velodyne HDL-64E S2 and VLP-16 between the vehicles (measurement scene 0) and the total station, statically placed 10 m in front of the vehicles. The vehicle positions were determined by measuring the middle axis

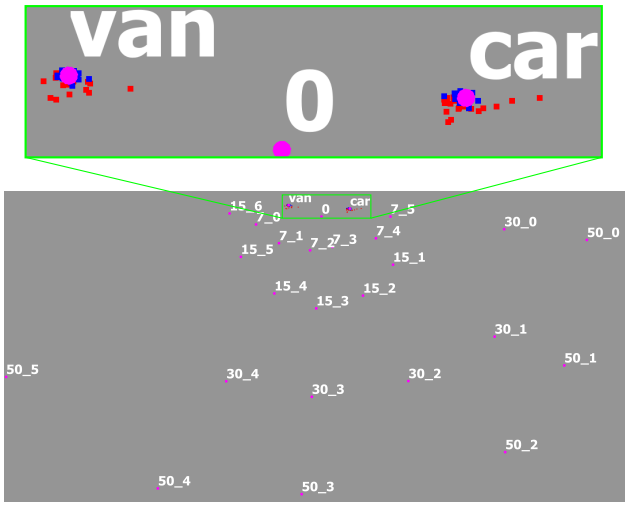


Figure 3. Bottom: the different locations of the scanner and car/van positions during the measurements (magenta). Top: the detections using the BC-PCA (blue) and ASM (red) in the HDL-64 scans

and a front point at the brand sign of the vehicle. The middle axis of the van was calculated by the measurement of roof rack fix points and the sedan type car middle axis by measuring the position of the antenna. We assumed a semicircle center between the vehicles and spread 25 measurement poses in total, 5-6 poses at each semicircle with approximated radii of 7 m ,15 m, 30 m and 50 m, see Figure 3. At each pose we measured 4 points at the rack with the total station to determine the pose of the scanner.

5. EVALUATION

We first validated our neural network by labeling a completely different junction scenario. Secondly, we calculated the Euclidean distance between the detected positions and the reference positions. We showed how the detections by ASMs slightly outperform the detection by BC-PCA. For this, we analyzed the mean pose accuracy, the viewing angle influence on the improvements for the different car models (van and sedan type) and the impact of different resolutions. For the evaluation of the vehicle heading we analyzed only $\pm 90^\circ$ difference, because the BC-PCA does not distinguish between the front and the back of a vehicle and the ASM also has difficulties to distinguish between the two opposite orientations, because of the geometric symmetries of vehicles (Coenen et al., 2018). However, the pose derived from the center of the bounding boxes is not effected by the 180°rotation.

5.1 Neural Network

We used 2000 scans from another junction to calculate an accuracy of 93%, a precision of 40%, a recall of 90% and a f1-score of 62%. Figure 4 shows the difficulties of our network. There are some erroneous detections in the background and obstacles in the front which occlude vehicles and are labeled as vehicles. Furthermore, the network slightly inflates the objects. Gaps in the training labels lead to false negatives and indicate a need for improvement in the trained labels.

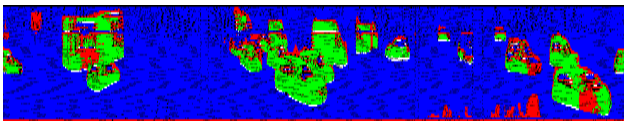
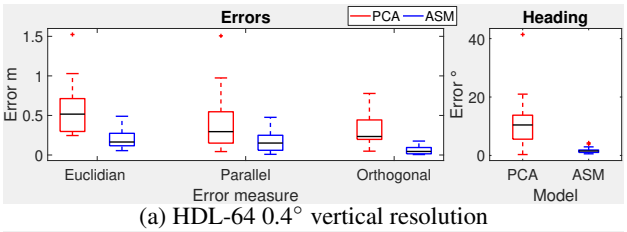


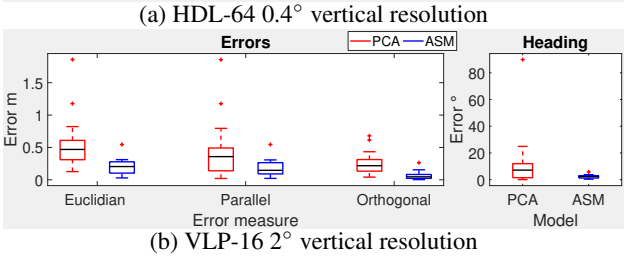
Figure 4. The validation of the neural network: true positives (green), true negatives (blue), false positives (red), false negatives (white)

5.2 Sedan type car

The height of the sedan type car is below our scan height of 1.8 m, which has the benefit of having measured points on the roof of the car in many scenarios. However, it is also partly occluded by the van in other scenarios, because of the same reason. By using ASMs we improve the mean position accuracy by 0.34 m (0.54 m to 0.2 m) in Velodyne HDL-64 scans and by 0.37 m (0.57 m to 0.2 m) in Veldoyne VLP-16 scans for a sedan car type, see Figure 5. Whereby for the VLP-16 analysis less scan positions (16/23) are considered due to the low amount of detected points. The comparison of the mean errors relatively to the car axis shows a balancing error parallel and orthogonal to the axis for the BC-PCA, because of the viewing angle depending position shift, see enlarged image part in Figure 3 (red points). The ASM position error orthogonal to the car axis is smaller compared to the error parallel to this axis, because of an more adequate model width and heading estimation. The improvement of the mean heading error is approximately $5^\circ$for both resolutions. Figure 6 shows the improvement by using ASMs compared to the BC-PCA by the error differences for each considered scene in more detail. Independent of the resolution the poses of ASMs show a continuous improvement of the position and the heading accuracy. The improvements in the HDL-64 scans (Figure 6 a) show no significant deteriorations in all scenarios (except scene 50_5), whereas the VLP-16 (Figure 6 b) indicates 30 m as maximum distance for adequate car detection in scans with a 2°resolution. There are small deteriorations in both resolutions for different scenes, because the ASM is not restricted by missing scan points in the car shadow and thus the shape is not adequately restricted, especially the length of the model. In total there is a maximum improvement of 1.74/1.37 m towards a maximum decline of -0.08/-0.11 m. Figure 7 shows the strengths and weaknesses of the two pose determination approaches by their total error in the HDL-64 scans for different

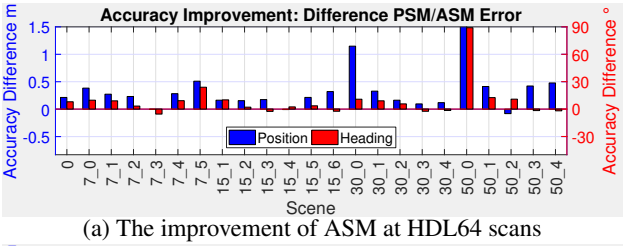


(a) HDL-64 0.4° vertical resolution

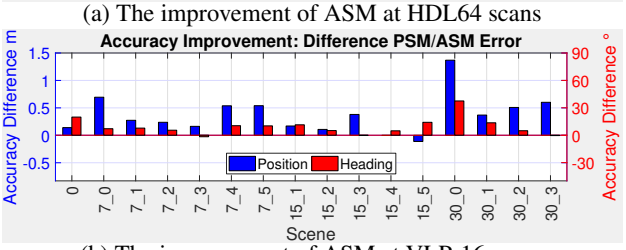


(b) VLP-16 2° vertical resolution

Figure 5. The mean error of the models for the sedan type car. Split in the total Euclidean error, the error along the car axis (left) and the heading error (right)



(a) The improvement of ASM at HDL64 scans



(b) The improvement of ASM at VLP-16 scans

Figure 6. The improvement of the accuracy using the ASM compared to the BC-PCA for the sedan type car

scenarios. It shows the underestimation of the car length by the BC-PCA because of the missing scan points at the end of the car by the error parallel (green) and the systematic error orthogonal to the car axis (yellow). The systematic shift towards the scanner is also visible in the enlarged image part at the red points in Figure 3. The underestimations of the length are mitigated by the ASMs (Figure 3 blue points), especially the error orthogonal to the car axis is reduced. The peaks in the total HDL-64

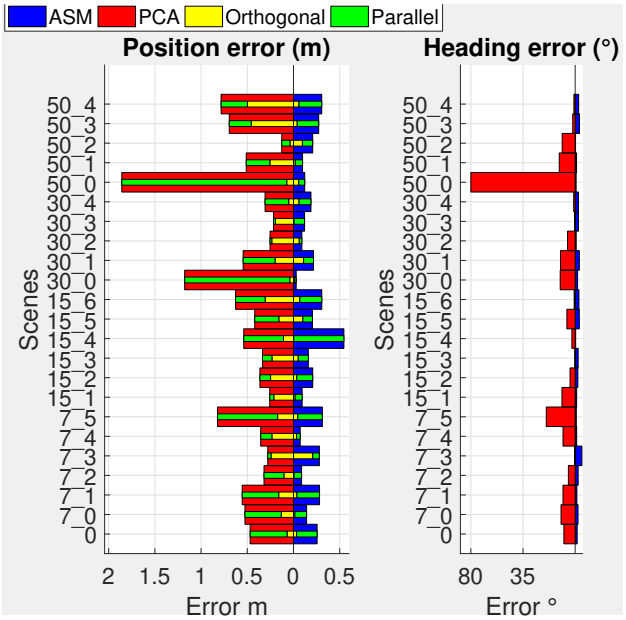
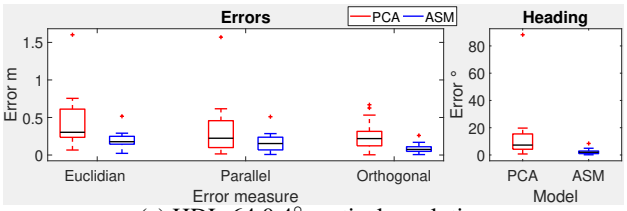


Figure 7. The absolute errors of the detections of the sedan type car in the HDL-64 scans. The position error (left) includes the error ratio orthogonal (yellow) and parallel (green) to the car axis

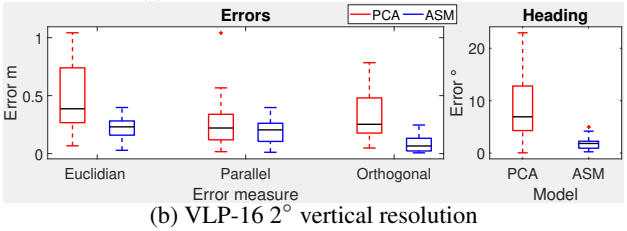
error diagram (Figure 7) in scene 7_5, 30_0 and 50_0 identify the weakness of the BC-PCA approach for only having scan points at the front or the back of the car. In these scenes, the pose accuracy can be highly improved by the ASM. The ASM significantly reduces the error along the car axis by assuming a ordinary vehicle length. In the other scenes the car is well visible as an L-shape (edge and front/back or roof). Therefore, both approaches work well and the differences are relatively small, compare Figures 6 and 7.

5.3 Van

The mean accuracy of the van's pose estimation is improved by using ASMs by 0.25 m (0.44 m to 0.19 m) in Velodyne HDL-64 scans and 0.26 m (0.48 m 0.22 m) in Veldoyne VLP-16 scans, see Figure 8. Looking at the point clouds, it can be observed that, in contrast to the previous car, the scan points on the roof of the van are missing. This is explainable by the experimental setup, in which the scanner height is below the car's roof. The missing points lead to a negative effect on the pose estimation by the BC-PCA, which determines the van width more inaccurate in contrast to the sedan type car. However, compared to this car, the van is detected in more scenarios (16/18 in VLP-16, compare Figure 6 and 9) because of its bigger size. In addition to that and looking at scene 50_5, the missing detection of the car in the HDL-64 scans (in contrast to the van, which is visible) indicates a maximum distance of around 50 m for a reliable detection in HDL-64 scans. Figure 9 shows the detailed error



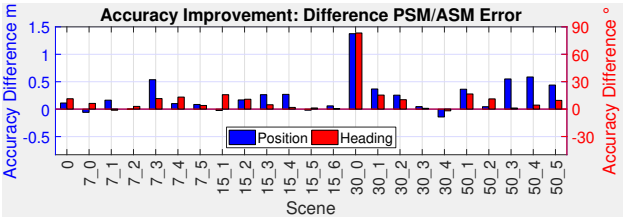
(a) HDL-64 0.4° vertical resolution



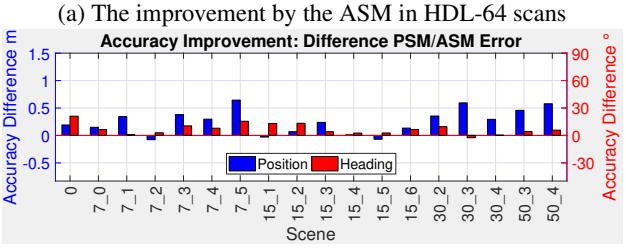
(b) VLP-16 2° vertical resolution

Figure 8. The mean error of the models for the van. Split in the total Euclidean error, the error along the van axis (left) and the heading error (right)

difference for each considered scene. Independent of the resolution using ASMs improves the detected poses in almost all scenarios. However, for the HDL-64/VLP-16 there are maximal improvements of 1.38/0.64 m toward a maximal decline of -0.14/-0.08 m. The impact of the different vehicle shapes is presented clearly by comparing the error of the van and the car in the HDL-64 scan, Figure 10 and 7. The different effects of the shape on the ASM and BC-PCA are visible by comparing the relation of the error parallel and orthogonal to the vehicle axis. The BC-PCA orthogonal error for the van is higher compared to the car. Thus, the center shifts to the left front because of the too short bounding box and only one visible edge. The ASM completes the missing edge and compensates the orthogonal error very well, but also often underestimates the length of the van. In contrast to the van pose, the sedan type car's pose can be estimated more accurately in the orthogonal direction of the vehicle axis by the BC-PCA because parts of the roof are visible. However, the occluded last part of both vehicles leads to an underestimation of their lengths. The ASM estimates the vehicle length and width more accurately and thus reduces the errors along both vehicle axes by adding missing bounding box edges assuming common width, height and length ratios.



(a) The improvement by the ASM in HDL-64 scans



(b) The improvement of the ASM in VLP-16 scans

Figure 9. The improvement of the accuracy by using ASMs in contrast to the BC-PCA for the van

6. CONCLUSION AND OUTLOOK

In summary, we used a geometric *lane and ground filtering* approach to generate training data for a simple neural network. We determined the mean accuracy of vehicle poses extracted from LiDAR point clouds generated by 3D scanners with 64 beams at 0.4°and 16 beams at 2°vertical resolution. We accurately detected cars, with at least 20 scan points, in distances of up to 50 m in Velodyne HDL-64 scans and up to 30 m in VLP-16 scans. We calculated a mean accuracy for a BC-PCA approach of around 0.48 m with a variance of around 0.12 m. We showed that ASM (Coenen et al., 2018) could be used to improve this pose accuracy by around 58% in comparison to the BC-PCA. We reached a mean accuracy of 0.2 m with a variance of 0.01 m by using ASMs. Further, we showed that the vertical resolution of the scanner has a negligible influence on the accuracy, but it affects the detection robustness due to the higher amount of scan points at the vehicle. The same applies for the vehicle size. The current ASM data set includes only two vans. We propose to first classify the vehicle type in order to use a more proper ASM for different vehicle types like cars, vans or buses and trucks. In future work, we will relabel our training data set by optimizing the lane and ground filtering segmentation and also distinguishing between cars, van, trucks/buses, pedestrians and cyclists. We will train other neural networks to come up with a robust traffic participant detection. In addition, the classification of cars, vans and buses/trucks will be used for training separate ASMs, which might lead to more specific ASMs and thus to a better pose accuracy. We will use the ASMs to track vehicles in point clouds and improve the detection accuracy by integrating the scan points from different time steps in the ASM estimation.

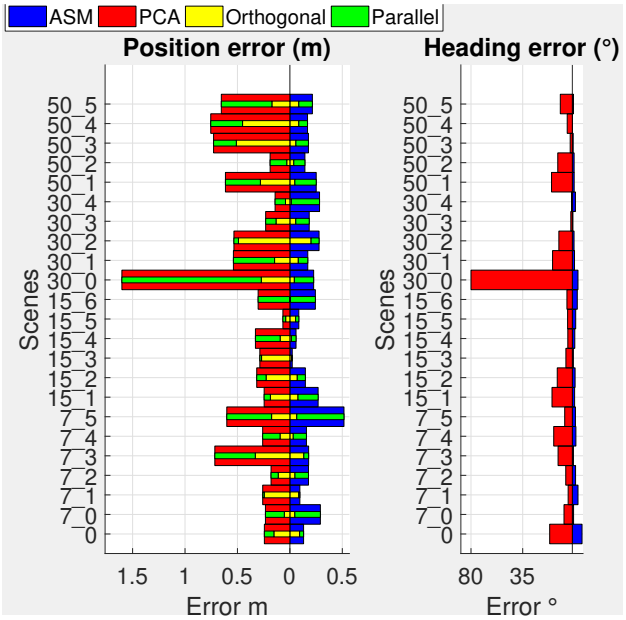


Figure 10. The absolute errors of the detections for the van in the HDL-64 scans. The position error (left) includes the error ratio orthogonal (yellow) and parallel (green) to the van axis

ACKNOWLEDGEMENT

This work was funded by the German Science Foundation DFG within the priority programme SPP 1835, “Cooperative Interacting Automobiles”.

References

Barea, Rafael, Pérez, Carlos, Bergasa, Luis M, López-Guillén, Elena, Romera, Eduardo, Molinos, Eduardo, Ocana, Manuel, López, Joaquín, 2018. Vehicle detection and localization using 3d lidar point cloud and image semantic segmentation. *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 3481–3486.

Chen, Xiaozhi, Kundu, Kaustav, Zhu, Yukun, Ma, Huimin, Fidler, Sanja, Urtasun, Raquel, 2018. 3d object proposals using stereo imagery for accurate object class detection. *IEEE transactions on pattern analysis and machine intelligence*, 40, 1259–1272.

Coenen, Max, Rottensteiner, Franz, Heipke, Christian, 2018. RECOVERING THE 3D POSE AND SHAPE OF VEHICLES FROM STEREO IMAGES. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 4.

Cootes, Tim, Baldock, ER, Graham, J, 2000. An introduction to active shape models. *Image processing and analysis*, 223–248.

Douillard, Bertrand, Underwood, James, Kuntz, Noah, Vlaskine, Vsevolod, Quadros, Alastair, Morton, Peter, Frenkel, Alon, 2011. On the segmentation of 3d lidar point clouds. *2011 IEEE International Conference on Robotics and Automation*, IEEE, 2798–2805.

Ferryman, James M, Worrall, Anthony D, Maybank, Stephen J, 1998. Learning enhanced 3d models for vehicle tracking. *BMVC*, 1–10.

Geiger, Andreas, Lenz, Philip, Stiller, Christoph, Urtasun, Raquel, 2013. Vision meets Robotics: The KITTI Dataset. *International Journal of Robotics Research (IJRR)*.

He, Tong, Soatto, Stefano, 2019. Mono3D++: Monocular 3D Vehicle Detection with Two-Scale 3D Hypotheses and Task Priors. *arXiv preprint arXiv:1901.03446*.

Himmelsbach, M., v. Hundelshausen, F., Wuensche, H. ., 2010. Fast segmentation of 3d point clouds for ground vehicles. *2010 IEEE Intelligent Vehicles Symposium*, 560–565.

Li, Bo, Zhang, Tianlei, Xia, Tian, 2016. Vehicle detection from 3d lidar using fully convolutional network. *arXiv preprint arXiv:1608.07916*.

Menze, Moritz, Heipke, Christian, Geiger, Andreas, 2015. Joint 3d estimation of vehicles and scene flow. *ISPRS Workshop on Image Sequence Analysis (ISA)*, 8.

Moosmann, Frank, Pink, Oliver, Stiller, Christoph, 2009. Segmentation of 3d lidar data in non-flat urban environments using a local convexity criterion. *2009 IEEE Intelligent Vehicles Symposium*, IEEE, 215–220.

Qi, Charles R, Liu, Wei, Wu, Chenxia, Su, Hao, Guibas, Leonidas J, 2018. Frustum pointnets for 3d object detection from rgb-d data. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 918–927.

Qi, Charles R, Su, Hao, Mo, Kaichun, Guibas, Leonidas J, 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 652–660.

Rabbani, Tahir, Van Den Heuvel, Frank, Vosselmann, George, 2006. Segmentation of point clouds using smoothness constraint. *International archives of photogrammetry, remote sensing and spatial information sciences*, 36, 248–253.

Simonyan, Karen, Zisserman, Andrew, 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Xu, Danfei, Anguelov, Dragomir, Jain, Ashesh, 2018. Pointfusion: Deep sensor fusion for 3d bounding box estimation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 244–253.

Zhou, Yin, Tuzel, Oncel, 2018. Voxelnet: End-to-end learning for point cloud based 3d object detection. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Zia, M Zeeshan, Stark, Michael, Schiele, Bernt, Schindler, Konrad, 2011. Revisiting 3d geometric models for accurate object shape and pose. *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, IEEE, 569–576.

Zia, M Zeeshan, Stark, Michael, Schindler, Konrad, 2015. Towards scene understanding with detailed 3d object representations. *International Journal of Computer Vision*, 112, 188–203.