

Reviewer 2

The manuscript introduces the ExistenceMap-PointPillars method and how it improves object recognition for safe autonomous driving.

In my opinion, the manuscript needs to be carefully revised and improved in the following aspects:

I would like to thank the reviewer for careful reading and valuable comments. We revised the paper based on the comments.

Comments 1: *It is recommended that the content of the manuscript should not exceed the layout.*

Response 1: We apologize for the wrong formatting and thank you for the mention. When writing the manuscript, it seems to be appropriate format. However, after submitted, the format is changed into the wrong format. We fixed the manuscript to the correct format.

Comments 2: *The basis of sensor fusion is spatio-temporal matching. How do the methods described in the manuscript achieve consistency in time and space?*

Response 2: In the manuscript, the method does not consider the time error between sensors, so this is to be the future work. We proposed a framework that could model some errors and the manuscript demonstrates modeling the error of sensor calibrations and the position error of detected objects in fused sensor. Modeling an error is not difficult, so it is easy to model time errors. We added the description to the explanation of pseudo 2D map generation in section 2.2 and the conclusion section.

Comments 3: *It is recommended to explain the interconversion theory of 2D and 3D data. Does this operation cause loss of accuracy?*

Response 3: As in the section 3.2.1, when the maximum of existence probability value is low, true positive could not be increased. The maximum value is defined by the region of the error ellipse, and the maximum value of the big error ellipse could be decreased. Therefore, setting the high variance cause loss of accuracy. In addition, when the center of the error ellipse is far from where the object is truly located, the same thing could be happening. In short, the manuscript shows fusing camera and tracking objects is effective to detect near objects. We added above consideration in the evaluation section 3.2.1.

Comments 4: *The experiment used the author's own data set. How to determine the data set label?*

Response 4: We offered professional company to label the dataset that contains LiDAR point cloud and omnidirectional camera images. In the dataset, the labels are tracking objects.

Comments 5: *It is recommended to compare and analyze the following articles separately in a targeted manner.*

- *Object Classification using CNN-Based Fusion of Vision and LIDAR in Autonomous Vehicle Environment. DOI: 10.1109/TII.2018.2822828.*
- *A Structure Constraint Matrix Factorization Framework for Human Behavior Segmentation. DOI: 10.1109/TCYB.2021.3095357.*
- *An Interacting Multiple Model for Trajectory Prediction of Intelligent Vehicles in Typical Road Traffic Scenario. DOI: 10.1109/TNNLS.2021.3136866.*
- *Robust Lateral Trajectory Following Control of Unmanned Vehicle Based on Model Predictive Control. DOI: 10.1109/TMECH.2021.3087605.*

Response 5: We appreciate for the recommendation of the articles. We added the explanation of the article about object classification to related work section as another example of LiDAR-camera fusion model. We covered the other articles in our introduction section.