Assessment of the master thesis by Rakshith Madhavan

Single View Depth Completion of Sparse 3D Reconstructions

doc. Ing. Tomas Pajdla, Ph.D.
thesis supervisor

The goal of the thesis was to investigate how single view depth reconstruction using machine learning can be combined with sparse 3D reconstruction by photogrammetry and visual odometry to obtain metrically accurate and dense 3D depth information.

Reconstructing reliable 3D models of indoor environment for robot localization and motion planning is a difficult task when no active 3D sensing is available. For the classical passive structure from motion and SLAM methods, many challenges in indoor scenes are (i) lack of features, (ii) repetitiveness, and (iii) changing environment and illumination. Altogether, 3D reconstructions build by SLAM are almost always sparse and noisy and often with severe distortions. Recent developments in single view depth estimation suggests to try combining machine learning with classical methods. The main problem of machine learning methods is that they often provide only qualitative depth information that is not metrically accurate. Thus, it suggests itself to combine sparse but metrically correct SLAM reconstructions with dense but less correct single view reconstructions to obtain more accurate and dense 3D models. The topic of this thesis is to investigate how to combine SLAM with learned single view depth estimation and whether it would be possible to replace active sensors with passive ones on ARI robot system. This is an interesting and non-trivial research topic.

The thesis of Rakshith Madhavan brings several contributions.

First, a detailed review of monocular depth estimation methods is provided including training and evaluation data sets and their evaluation metrics. The review covers supervised as well as unsupervised depth estimation and completion methods and provides a very useful comparison of a number of methods including the access to their implementations. Particular attention is paid to the single view reconstruction technique of Wong et al. [1], which is explained very well in great detail. The review of SLAM techniques nicely connects SLAM to various machine learning approaches to depth completion and reconstruction but also very explains in detail the principles of SLAM itself.

Secondly, the thesis proposes a new method how to densify sparse point cloud data from a modified ORB SLAM algorithm [2] using a modified Wong et al. depth completion network [1]. The modifications introduce new approaches based on an overfitting training paradigm, and back-projection to the original point cloud. This is a new modification and new approach to combining a depth completion network with a SLAM sparse reconstruction.

Finally, the thesis evaluates the results on existing and on a new dataset. The new data set consists of images from ARI robot taken in four different rooms spanning two different environments. Quantitative evaluation on the VOID dataset shows RMSE errors around 10cm, which is close to acceptable for general purpose robotics and definitely acceptable for some tasks related to global ARI robot navigation. Qualitative results on the new data sets also show promising results. It is important to see that using ORB SLAM instead of PoseNet produces lower errors. I believe that the method suggested in the thesis has potential to reach even smaller errors and to provide dense 3D maps useful for general 3D mapping.

Rakshith was an excellent student. He worked very actively and independently. He learned all the background in machine learning, suggested an approach, implemented it, designed experiments and demonstrated the performance of his method. He fulfilled all the goals set in the assignment and presented a new approach to solving a challenging problem in depth sensing for robotics.

The master thesis is excellent, and thus I grade it as excellent (A).

Prague, 5 June 2022
doc. Ing. Tomas Pajdla, Ph.D.
thesis supervisor