УДК 004.896

ENVIRONMENT MAPPING USING MONOCULAR DEPTH ESTIMATION FROM RGB IMAGES

Бархум М (Университет ИТМО),
Научный руководитель – профессор, Пыркин А.А.
(Университет ИТМО)

Abstract. This work is directed towards developing a mapping system that relies only on a simple RGB camera as a sensor, where a depth map will be estimated from the image provided by the camera, then it will be refined, and reprojected as a dense 3-D point cloud to the world frame as a pseudo laser sensor, which will be used to map the environment using a Simultaneous Localization and Mapping algorithm.

Introduction. For navigation in mobile robots a mapping system is a necessity, since it provides the robot with information about the surrounding world, perceived by onboard sensors the robot is equipped with. But some advanced mobile robots are so small in size that it is almost impossible for it to be equipped with multiple sensors that are essential for the mapping system. Artificial intelligence can prove to be a capable alternative to these sensors, and the recent works in monocular depth estimation show promising efforts towards robust depth estimation that may in fact in the near future replace expensive sensors in some areas.

Main part. Monocular depth estimation is considered a challenge in artificial intelligence, due to the lack of depth cues in a monocular image. A common practice in literature is to estimate the disparity or the inverse depth from monocular images, to simplify encoding points at infinity as well as having a gaussian distribution approximation of the noise, but those disparity maps values are relative and not metric, thus the need for a rescaling step or refining where these disparity maps are refined to a metric map representation and inversed to get depth values.

Taking that into account the work will go through the following steps:

- Estimating disparity maps or inverse depth maps from monocular images.
- Refining the resulting disparity to acquire metric depth from relative depth estimated by the network (network are trained on various datasets with different camera characteristics).
- Reprojecting 3-D points from the depth map to the world frame as a dense point cloud.
- Taking a set of these points as a pseudo laser scanner and projecting them to 2-D.
- Mapping using the pseudo laser scanner using a Simultaneous Localization and Mapping algorithm.

Conclusion. The work is implemented as a simulation on Robotic Operating System (2.0), where the results will show implementation of monocular depth estimation and reprojection of point clouds in real time, comparison between the reprojected point cloud from the depth predicted by the neural network and the ground truth point cloud projected by the RGBD sensor, and how well does the mapping system behave using the pseudo laser scanner created from reprojected point clouds.

Бархум М (автор) Подпись

Пыркин А.А (научный руководитель) Подпись