

Semantic Probabilistic Traversable Map Generation For Robot Path Planning

Yimo Zhao, Peilin Liu, Wuyang Xue, Ruihang Miao, Zheng Gong, Rendong Ying

Abstract—Probabilistic traversable map plays a critical role for mobile robot in safe and reliable navigation. Different from structured environment, traversable region in unstructured environment such as grass and sidewalk is relatively more complex. Traditional elevation-based traversable map cannot represent such complex environment well. Thus, it may cause navigation failure. To address this limitation, this paper proposes a novel semantic-elevation mapping approach for navigation task. We first build a multi-layer semantic map from continuous semantic segmentation images. Then, this multi-layer semantic map is fused and converted into a probabilistic map by a distance transform approach. Generated semantic probabilistic map is then fused to an elevation map at path planning stage. The proposed approach is tested on an Unmanned Ground Vehicle (UGV) platform. The results show that our semantic-elevation mapping approach works more reliably and safely than that only with elevation-based approach.

Index Terms—probabilistic traversable map, semantic segmentation, LiDAR pointcloud, grid map, sensor fusion

I. INTRODUCTION

Navigation and obstacle avoidance are essential to autonomous mobile robot and vehicle. One key challenge is understanding the environment and further generating a traversable map. This challenge is more demanding in unstructured environment, as the environment can become complicated and unpredictable. Furthermore, traversable regions such as grass and sidewalk require passing with more caution. Fig. 1 shows the unstructured terrain area and the illustrated probabilistic traversable map.

Early research on traversable map generation for navigation is mainly based on elevation map and relative geometric information, such as slopes, orientation, gradients [1]–[3]. Region with large height difference could be filtered out using elevation based map. Thanks to the development of ranging sensors such as Light Detection And Ranging (LiDAR), point-cloud based 3D geometric traversable map method is also proposed [4]. However, these maps cannot distinguish between different traversable region types with similar spatial structure. This may cause unsafe movement, such as going through a rough muddy meadowland.

Vision based traversable map generation also becomes a research hotspot. In [5] the authors use disparity from stereo camera and select the ground region as the traversable map. Recently, learning based methods especially Deep Neural

The authors are with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: S-carlet@sjtu.edu.cn; liupeilin@sjtu.edu.cn; icecreamxwy@sjtu.edu.cn; m-rhcat@sjtu.edu.cn; gongzheng@sjtu.edu.cn; rdying@sjtu.edu.cn).

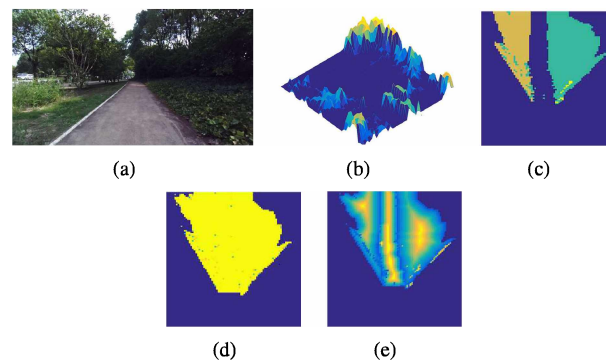


Fig. 1. An example of unstructured environment and illustrated probabilistic traversable map by our approach. (a) unstructured environment scene which contains different road types and structures, (b) illustrated elevation map (c) semantic label of illustrated semantic map, (d) label confidence of illustrated semantic map (e) final traversable probability of illustrated semantic map

Network (DNN) gain popularity [6], which could separate different regions by semantic information. However, vision based traversable map cannot obtain accurate geometric structure. Besides, data driven approach is not stable enough compared to the elevation-based approach.

For improving the detection accuracy and stability, sensor fusion becomes the first choice in [7]. Bayesian rules are also used in [8] for better information fusion. Moreover, In [9] the authors consider to use Dempster’s conjunctive rule to fuse sensor information.

In this paper we propose a framework that could detect the traversable region and generate a probabilistic traversable map for navigation combining camera and LiDAR. The main contributions of this paper are:

- 1) A novel semantic probabilistic traversable map generation approach using semantic information from DNN model.
- 2) A fusion approach of generated semantic probabilistic traversable map and elevation-based map at path planning stage.

The paper is organized as follows: Section II gives an overview of related work in traversable map generation, sensor fusion and DNN based road detection. Section III gives the detailed description of traversable probability map generation approach from camera and LiDAR, and the path planning approach based on our traversable map. Section IV presents the experiment settings and corresponding experimental results. At last, final conclusion is given in V.

II. RELATED WORK

A. Probability Traversable Map Generation

Probabilistic traversable map, or drivable map, represents the traversable probability value of a position. In [10] the author fuses neural network, fuzzy-logic and rule base terrain classification up, use these all together to define traversability. A method that uses a stored metric to represent traversability of given space is proposed in [11], which could provide a global solution for mobile robot. In [12] the authors propose an evidential mapping method using stereo vision. An enhanced mapping model is generated by fusing U-disparity image and V-disparity image. 3D traversable map is also generated based on pointcloud in [4], which uses voxel to represent the traversability in 3D space.

In this paper, semantic information together with geometric information are used to represent the traversability. Furthermore, the safety of robot movement is considered in our proposed approach. According to the traversability region type and the geometric structure, we transform the map into a probabilistic map of which each grid stores the probability of traversability level.

B. Sensor Fusion

In the past, there are multiple sensor fusion schemes used for detecting the road. Early methods detect the road with different sensors in a sequential order. [13] first determines drivable region using LiDAR sensor, then the result region is further refined using image data. In this way, the whole system will collapse if the result from the previous sensor is not correct. Later, some researchers fuse the image and LiDAR data in feature stage. Features are directly extracted from image and LiDAR data and sent into a Support Vector Machine (SVM) model for classifying in [14]. Some researchers choose asynchronous methods to fuse the separate temporal result from different sensors. For example, [9] uses evidential theory to fuse the evidential grid map generated from image and LiDAR scans.

Sensors are fused at the path planning stage in our proposed approach. After generating semantic grid map and elevation grid mapping asynchronously, we use these two grid maps to calculate the path cost during the path planning procedure.

C. Deep Neural Network based Road Detection

With the development of DNN model, 2D image road detection has achieved state-of-art performance on several datasets. In [15] a method using DNN model combining color line model prior in a CRF model for detecting unstructured road is proposed. Current road detection are mainly based on semantic segmentation models which only concentrate on single road class.

The drawbacks of current road detection models still exist. Lack of geometric information and single detection class let DNN based methods cannot directly be used for navigation tasks.

In this article we proposed a semantic grid mapping approach to transform 2D segmentation results to a semantic probabilistic grid map.

III. FRAMEWORK ILLUSTRATION AND APPROACH DESCRIPTION

A. Mapping System framework Overview

We build the probabilistic traversable grid map based on grid map library [16], which is a multi-layer grid map library. The overall semantic-elevation mapping framework is shown as Fig. 2.

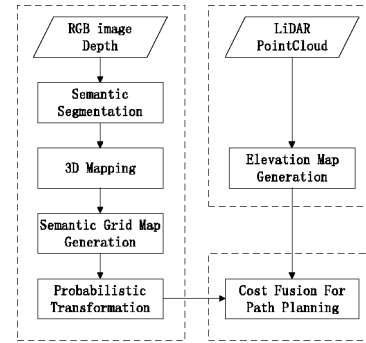


Fig. 2. The overall structure of our proposed mapping system.

As shown in the workflow, semantic grid mapping and geometric grid mapping works asynchronously. Semantic grid mapping module takes RGB and dense depth images as input, and outputs semantic probabilistic traversable grid map. Since the depth measured by stereo camera is not as accurate as that measured by LiDAR and could not handle the illuminance variance. The geometric map needs more stable and accurate spatial information to avoid tough non-traversable condition. Thus, geometric mapping module takes LiDAR pointcloud as input, and outputs an 2.5D elevation grid map. After generating both grid maps, fusion is performed during path planning.

B. Geometric Grid Mapping From LiDAR Pointcloud

LiDAR can provide accurate geometric information of the environment. Height value is mainly used in our geometric grid mapping, which is especially essential for navigation.

For raw LiDAR pointcloud data, we choose to use elevation mapping library [17] to generate 2.5D elevation grid map. Each grid cell stores the updated height value. The elevation grid map considers the sensor noise and motion noise, and it could also remove moving objects.

An example of elevation grid mapping procedure is shown as Fig. 3.

C. Semantic Grid Mapping From Images

Although LiDAR provides accurate geometric information of the environment, the specific traversable region types are unknown. As for the UGV platform, it needs to be careful while crossing tough areas or going through different

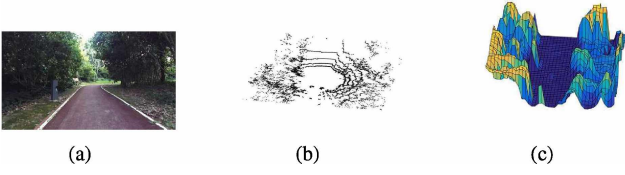


Fig. 3. An illustration of elevation grid map generation approach. The colors in the elevation map reflect the height value of each cell. (a) unstructured environment with complex spatial structures (b) input LiDAR pointcloud data (c) generated elevation grid map, light color represents high height value.

traversable region types. Semantic grid map helps to solve this situation when the environment contains consistent geometric information but with different semantic meanings. This could help the path planning module to give a more reliable and safe path. The generation of semantic grid map contains 4 main steps: 2D semantic segmentation, 3D mapping (3D re-projection), grid map generation and traversable probability transformation.

For 2D semantic segmentation, we use DNN based segmentation model to provide semantic label for our map. Deeplab v3+ [18] is considered to be one of the most accurate 2D semantic segmentation model. After getting the semantic segmentation result from Deeplab v3+ model, we re-project the dense segmentation label results back to 3D space using dense estimated depth image from ZED stereo camera. Denotes homogeneous coordinate of image pixel as $p = (u, v, 1)$ and the corresponding 3D homogeneous coordinate as $x = (X, Y, Z, 1)$, the relationship between p and x is given by

$$p = K [R|t] x, \quad (1)$$

where $K \in \mathbb{R}^{3 \times 3}$ is the intrinsic matrix of the camera. T (composed of rotation $R \in \mathbb{R}^{3 \times 3}$ and translation $t \in \mathbb{R}^3$) denotes the transformation matrix between camera coordinate system and world coordinate system. In our implementation, T is given by the ORB_SLAM2 [19]. From (1) we can transform the image coordinates to world coordinates by combining the depth image.

After re-projecting the dense semantic label result into world coordinate system, grid mapping is used to transform labeled pointcloud to a multi-layer 2D grid map. Each cell at l -th layer stores the confidence of being label l . For updating the confidence of each grid cell, we follow the Bayes' update formula, which is similar to the octomap [20]. Let $p(x_t|z_{1:t})$ denote the segmentation confidence of grid cells at time t . When there is a new observation z_t , the confidence $p(x_t|z_{1:t})$ is updated based on Bayes' rule

$$p(x_t|z_{1:t}) = \frac{p(x_t|z_t)p(z_t)}{p(x_t)} \cdot \frac{p(x_{t-1}|z_{1:t-1})}{p(z_{t-1}|z_{1:t-1})}. \quad (2)$$

The initial prior label probability is set to be equal. Depart from 0/1 occupancy problem, here we face a multi-label classification problem. The strategy as "one-vs-all" is used. Let c denote the confidence of being segmented to label i , the number of classes is n_i . The segmentation label of a point

falling into cell is denoted by $l \in \{1, 2, 3 \dots L\}$. Then, the update prior probability can be defined as

$$p(x_t|z_t) = \begin{cases} c, & l = i \\ \frac{1-c}{n_i-1}, & else \end{cases}. \quad (3)$$

As for the prior update probability c for label l , we choose to use the Intersection over Union (IOU) of each label I_l to represent the update probability of correctly segmenting this label. Suppose an area of label l in ground truth is X , and its test result is Y . The overlap area O_{XY} and the union area of X and Y is U_{XY} can be shown as Fig. 4.

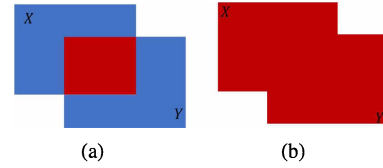


Fig. 4. Overlap and union of label i , region with red color represents the calculated region. (a) overlap area of X and Y (b) union area of X and Y

Then, the IOU of semantic segmentation I_l can be calculated as

$$I_l = \frac{O_{XY}}{U_{XY}}. \quad (4)$$

To get the final segmentation label for each grid cell, we calculate the maximum posterior probability through all label layers according to the calculated segmentation confidence. To maximize the steering reliability, the UGV platform is kept away from the boundaries, or avoid crossing different type of terrain. To achieve this goal, we use distance transform to define the distance between non-zero grid area to the zero grid area. Given the traversable region Tr and the non-traversable region Nt in a generated semantic grid map M_s , the transform D can be shown as

$$D(p) = \min(|m - n|^2), m \in Tr, n \in Nt. \quad (5)$$

From the transformed distance of semantic grid map, we can get a high score at the center of the traversable area, and low score at the boundaries. Finally, we transform the calculated distance to a probability representation p_D using Gaussian distribution with mean μ and variance σ as

$$\begin{aligned} \mu &= \max(D), \\ \sigma^2 &= \mathbb{E} \left[\frac{1}{n} \sum (d - \mu)^2 \right], d \in D, \\ p_D &\sim N(\mu, \sigma). \end{aligned} \quad (6)$$

Where n is the number of D . Then we assign a prefixed prior traversable probability p_l to each different label with different distance factors to get the final semantic traversable probability p_t as

$$p_t = p_D \cdot p_l. \quad (7)$$

The semantic grid mapping procedure can be shown as Fig. 5.

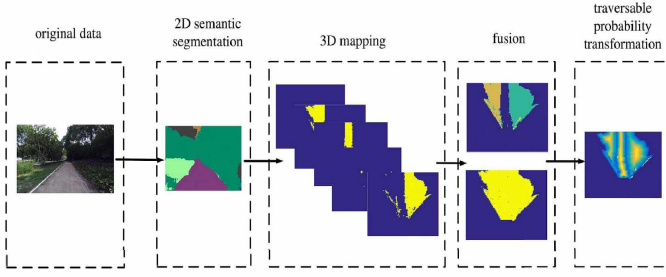


Fig. 5. An illustration that shows the whole procedure of semantic grid mapping with intermediate result at each stage.

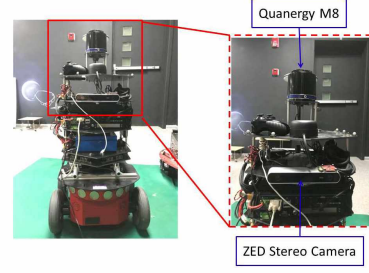


Fig. 6. UGV test platform with LiDAR and ZED stereo camera. The relevant sensors are annotated in the figure.

D. Fusion For Path Planning and navigation

In this paper, we also provide a path planning approach for navigation task. It fuses the information from our generated probabilistic traversable map. We choose Dynamic Window Approach (DWA) [21] for our path planning. It selects several rays with different angles and curvatures as candidates, and calculates the cost along each ray. Final path planning result will be the ray with lowest cost.

The cost calculated by the semantic and elevation grid map is shown as

$$\begin{aligned} C_e &= \delta h + R, \\ C_s &= c_i \cdot (2 - p_t), \end{aligned} \quad (8)$$

where C_e refers to the cost derived from geometric grid map. δh denotes the height difference between two neighborhood grid cells. R is the resolution of grid map. C_s refers to the cost derived from semantic grid map.

Except the cost from semantic map and elevation map, the distance from current position to the destination d and the steps number that need to pass through S are also considered. Let k denote the num of the candidate ray pass through grid cells in the grid map. The final fusion cost function C_f is shown as

$$C_f = \alpha_s \cdot S \cdot R - \alpha_d \cdot d - \sum_k C_s \cdot C_e. \quad (9)$$

Costs can be adjusted by the weight factor α_s and α_d .

IV. EXPERIMENTAL RESULTS

A. Experiments Setup

We evaluate our grid mapping approach via navigation task in real environment in our campus, the performance of the semantic mapping module and probability transformation module are first evaluated. Then, the performance of the whole mapping system is evaluated.

Our UGV platform is equipped with quanergy M8 LiDAR, a stereo camera which provides stereo RGB images and depth image with resolution 1280×720 . We resize the input image to 480×320 . The test UGV platform [22] is shown as Fig. 6.

B. Performance of Semantic grid Mapping on Path Planning

In this section, we conduct two comparative experiments to show the effectiveness of our proposed semantic grid mapping for path planning. The chosen test environment as in Fig. 7 on page 7 contains consistent geometric information but different semantic information. Under this environment settings, the comparative experiments are conducted by performing path planning with and without semantic grid mapping module. A joystick is used to only give a direction order to the UGV platform, such as "move forward", and the path planning results are recorded shown as Fig. 7 on page 7.

Fig. 7 shows that, path planning without semantic grid map can not distinguish between safe traversable and non-safe traversable region when the geometric structure is consistent. Thus, the UGV platform moves out of the road and cause the navigation failure. On the contrary, with semantic grid map, our mapping approach could differ road from grass or sidewalk, and avoid passing through different type of traversable regions. As we can see from the results, the map with semantic mapping module drives the UGV platform to turn right earlier and successfully move along the road.

C. Performance of Traversable Probability Transformation on Path Planning

In this section, the performance of traversable probability transformation is evaluated via path planning. The experiment is set as following step:

- 1) Turn off the traversable probability transformation module, use the mapping result for path planning at chosen test environment.
- 2) Turn on the traversable probability transformation module, use the mapping result for path planning at the exact same initial position.
- 3) Record the result path along the movement step by step.

Here we set the UGV platform to be at one side of the road and face towards the boundaries of other side. We use joystick to set the move order as "move forward", the path planning results in the given test environment are shown as Fig. 8 on page 7.

The difference between two path planning results can be clearly seen from the test result. The mapping system without traversable probability transformation treats all the position of

the traversable area as equal and moves towards boundaries. When the UGV platform reaches the boundaries with different geometric structure, then it will turn. By considering the distance to the boundaries, our mapping framework assign a higher probability value to the center of the traversable region, which drives the UGV platform to walk along the center of the traversable region and turn earlier and gives a better performance on planning strategy. Fig. 9 also shows difference performance via the trajectory during the experiment.

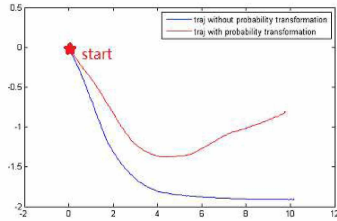


Fig. 9. The trajectory of the experiment with and without probability transformation. The coordinates of x-axis and y-axis is meter(m).

D. Performance of whole grid mapping system on navigation

At last, two navigation task experiments with our whole semantic-elevation grid mapping system in real environment are performed. The test environment is shown as Fig. 10.



Fig. 10. Scene examples of two test environments for navigation tasks. (a) scene for navigation experiment in complex environment. (b) scene for middle-range navigation experiment.

In the first scene, the performance of our grid mapping approach is tested via navigation in a complex environment, which contains road, grass and sidewalk with similar height. But the fitness trail climbs upwards. We let the UGV platform stay at the middle of the fitness trail and the sidewalk. The navigation result is shown as Fig. 11. At first, The area in front is with similar height. Without semantic mapping, the UGV platform will move forward until meet the gap. On the contrary, with semantic-elevation map, the UGV platform manage to choose one side(the fitness trail in our experiment) with higher traversable value, and move back to the center of the trail successfully. The entire trajectory of this experiment is shown as Fig. 12.

For the middle-range navigation, a path about 200 m is chosen as the second test scene. The UGV platform also autonomously moves from the front door to the back door of the building according to the given direction order. The trajectory during the entire navigation is shown as Fig. 13.

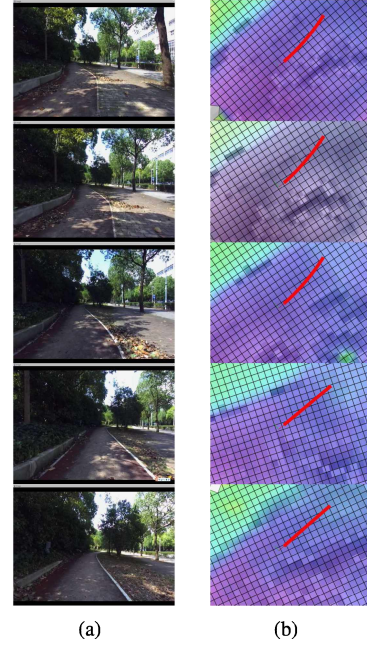


Fig. 11. Experimental results for navigation in a complex environment. (a) left image from ZED stereo camera. The UGV platform manages to differ road from other area and drives away from the boundary (b) the generated grid map and calculated path. Red line represents the selected path candidate. Purple color represents lower height value in the map, and green/yellow represents high height value. Grid with different grayscale value represents transformed probability value



Fig. 12. The trajectory of the navigation in the complex test environment.

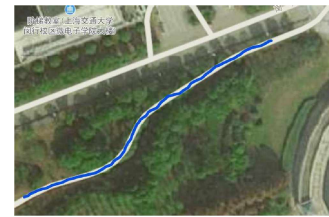


Fig. 13. The trajectory of the middle-range navigation.

The results show that our grid mapping approach can not only handle the complex environment, choose a better traversable path and move along the road, but also manage to successfully navigation through middle-range path.

V. CONCLUSION

In this paper, we propose a semantic-elevation probabilistic grid map for mobile robot path planning. The generated

semantic and elevation grid map are fused up at path planning stage. Several experiments are conducted on the UGV platform to evaluate our mapping approach, results are also compared with the baseline mapping method. The results of the experiments show that: semantic mapping could make the whole system more robust and reliable by assisting the elevation map, drive the UGV platform to choose more reliable and safer traversable region while the geometric structure is similar; Traversable probability transformation could keep the UGV platform away from the boundaries, which also makes the path planning safer.

REFERENCES

- [1] M. R. Blas, S. Riisgaard, O. Ravn, N. A. Andersen, M. Blanke, and J. C. Andersen, "Terrain classification for outdoor autonomous robots using single 2d laser scans - robot perception for dirt road navigation," *Integrated Computer-Aided Engineering*, vol. 13, no. 3, pp. 223–232, 2005.
- [2] C. Ye and J. Borenstein, "A method for mobile robot navigation on rough terrain," in *IEEE International Conference on Robotics and Automation*, 2004.
- [3] A. Garcia, A. B. Cruz, A. Medina, P. Colmenarejo, L. Mollinedo, and C. Rossi, "3d path planning using a fuzzy logic navigational map for planetary surface rovers," 2011.
- [4] B. G. Seo and M. J. Chung, "Traversable ground detection based on geometric-featured voxel map," in *The 19th Korea-Japan Joint Workshop on Frontiers of Computer Vision*, Jan 2013, pp. 31–35.
- [5] Y. Xie, S. Zeng, Y. Zhang, and L. Chen, "A cascaded framework for robust traversable region estimation using stereo vision," in *2017 Chinese Automation Congress (CAC)*, Oct 2017, pp. 3075–3080.
- [6] Y. Lyu, L. Bai, and X. Huang, "Road segmentation using cnn and distributed lstm," *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5, 2018.
- [7] H. P. Moravec, *Sensor Fusion in Certainty Grids for Mobile Robots*, 1989.
- [8] J. Sock, J. Kim, J. Min, and K. Kwak, "Probabilistic traversability map generation using 3d-lidar and camera," in *IEEE International Conference on Robotics and Automation*, 2016.
- [9] E. Capellier, F. Davoine, V. Fremont, J. Ibanez-Guzman, and Y. Li, "Evidential grid mapping, from asynchronous lidar scans and rgb images, for autonomous driving," 11 2018, pp. 2595–2602.
- [10] A. Shirkhodaie, R. Amrani, and E. Tunstel, "Soft computing for visual terrain perception and traversability assessment by planetary robotic systems," in *IEEE International Conference on Systems*, 2006.
- [11] H. Seraji, "New traversability indices and traversability grid for integrated sensor/mapbased navigation," *Journal of Robotic Systems*, vol. 20, no. 3, pp. 121–134, 2003.
- [12] C. Yu, V. Cherfaoui, and P. Bonnifait, "Evidential occupancy grid mapping with stereo-vision," 2016.
- [13] H. Dahlkamp, A. Kaehler, D. Stavens, S. Thrun, and G. R. Bradski, "Self-supervised monocular road detection in desert terrain," in *Robotics: Science and Systems II, August, University of Pennsylvania, Philadelphia, Pennsylvania, Usa*, 2006.
- [14] F. Schilling, X. Chen, J. Folkesson, and P. Jensfelt, "Geometric and visual terrain classification for autonomous mobile navigation," 09 2017, pp. 2678–2684.
- [15] S. Yadav, S. Patra, C. Arora, and S. Banerjee, "Deep cnn with color lines model for unmarked road segmentation," in *2017 IEEE International Conference on Image Processing (ICIP)*, Sep. 2017, pp. 585–589.
- [16] P. Fankhauser and M. Hutter, "A Universal Grid Map Library: Implementation and Use Case for Rough Terrain Navigation," in *Robot Operating System (ROS) The Complete Reference (Volume I)*, A. Koubaa, Ed. Springer, 2016, ch. 5. [Online]. Available: <http://www.springer.com/de/book/9783319260525>
- [17] P. Fankhauser, M. Bloesch, and M. Hutter, "Probabilistic terrain mapping for mobile robots with uncertain localization," *IEEE Robotics and Automation Letters (RA-L)*, vol. 3, no. 4, pp. 3019–3026, 2018.
- [18] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *ECCV*, 2018.
- [19] R. Mur-Artal and J. D. Tardós, "ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.
- [20] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "OctoMap: An efficient probabilistic 3D mapping framework based on octrees," *Autonomous Robots*, 2013, software available at <http://octomap.github.com>. [Online]. Available: <http://octomap.github.com>
- [21] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," *Robotics and Automation Magazine, IEEE*, vol. 4, pp. 23 – 33, 04 1997.
- [22] Z. Gong, W. Xue, Z. Liu, Y. Zhao, R. Miao, R. Ying, and P. Liu, "Design of a reconfigurable multi-sensor testbed for autonomous vehicles and ground robots," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2019, pp. 1–5.

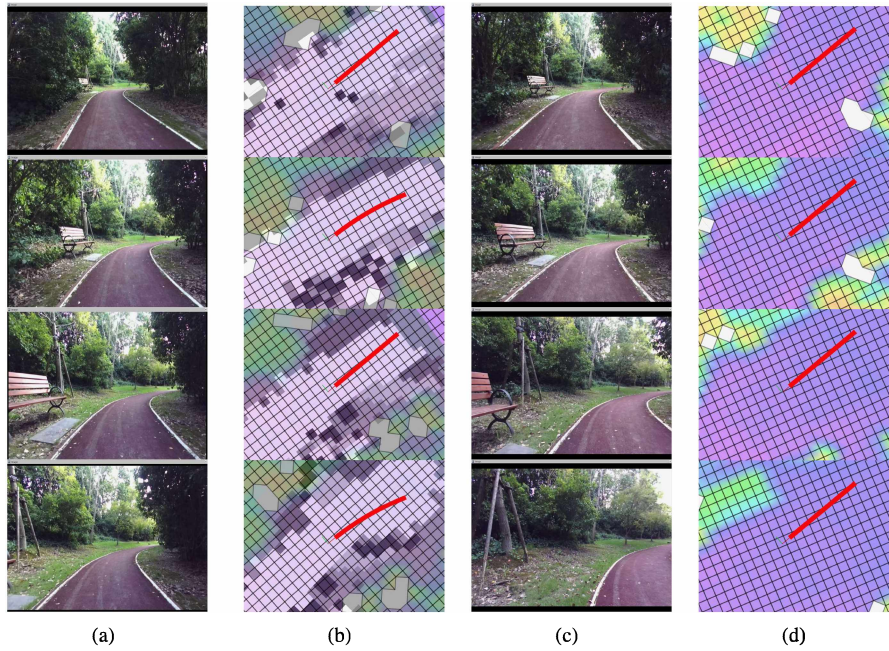


Fig. 7. Experimental result of path planning with and without semantic grid mapping. (a) and (c) are the left image from ZED. (b) and (d) are the generated grid map and the path planning results. Red line represents the selected path candidate. Purple color represents lower height value in the map, and green/yellow represents high height value. Grid with different grayscale value represents different label. (a) With semantic grid mapping, the UGV platform successfully pass through the turn (b) The right curved path candidate leads the UGV platform (c) Without semantic grid mapping, the navigation fails (d) The path candidate drive the UGV platform always move forward until reach a clear geometric boundary.

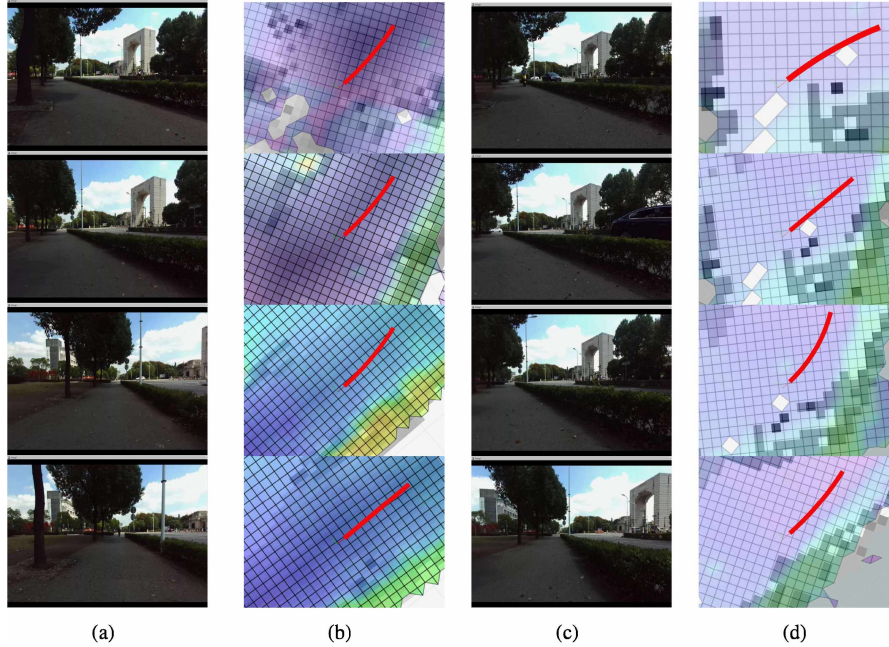


Fig. 8. Experimental result of path planning with and without probability transformation. (a) and (c) are the left image from ZED. (b) and (d) are the generated grid map and the path planning results. Red line represents the selected path candidate. Purple color represents lower height value in the map, and green/yellow represents high height value. Grid with different grayscale value in (b) represents transformed probability value, grid with different grayscale value in (d) represents different label. (a) With probability transformation, the UGV platform turns earlier to keep itself near the center (b) The higher probability at center drives the UGV to turn earlier (c) Without probability transformation, the UGV turns until reach the boundary (d) The path candidates are straight forward at the beginning