A Novel Ground Plane Detection Method Using an RGB-D Sensor

Li Chen\textsuperscript{1,}\textsuperscript{*}, Jun Zhou\textsuperscript{1} and Xuepeng Chu\textsuperscript{1}

\textsuperscript{1}School of Electrical and Mechanical Engineering, HoHai University, Changzhou, China

\textsuperscript{*}hhuc_cl@hhu.edu.cn

Abstract. A novel ground plane detection method is put forward based on three-dimensional (3D) point cloud data obtained using an RGB-D sensor. It consists of three stages: data pre-processing, occupancy grid map construction and ground plane segmentation. In order to obtain more accurate 3D point cloud data, the weight median filter (WM) is applied to recover the invalid depth pixel in the depth image. Different from the traditional approaches which process the 3D point cloud data directly, our algorithm transforms the 3D point cloud data into an occupancy grid map. Considering that the occupancy in the occupancy grid map and the distance from the point to the ground plane are distinct between ground points and other points, we will get a part of the 3D points that is definitely in the ground. The points selected are back-projected to the pixel coordinate system to get the result of the ground plane detection. In terms of experiment, the sensor is mounted on the mobile robot, Turtlebot2. The proposed method can detect more than 95 percent of the ground point. It produces accurate ground plane detection in different scenes.

1. Introduction

It is crucial for 3D navigation and environment perception in robotics fields. In order to perceive the 3D world, the robots are equipped with 3D sensors such as laser radars and stereo cameras. Lidars are relatively expensive and big consumers of energy. It is complex to build imaging systems with stereo cameras. Imaging systems with stereo cameras is difficult to deal with textureless regions. In recent years, Microsoft Kinect sensors (Kinect) have created many opportunities for multimedia computing. Kinect's\cite{1} impact has extended far beyond the gaming industry. With its wide availability and low cost, a quantity of researchers and practitioners in computer science and robotics adopt it in the field of 3D modelling and human-computer interaction. Nevertheless, it doesn’t work well in many cases, such as strong light exposure, transparent object, reflective or IR absorbing surfaces etc.

Several previous works approach to the ground plane detection problem using a \textit{v}-disparity representation\cite{2}. Hough transform\cite{3}, is a relatively mature algorithm that can be employed not only for the detection of lines and circles on 2D images, but also for 3D space for plane detection. Another common approach is RANSAC algorithm\cite{4,5,6}. Hahnel et al.\cite{7} randomly selected 3D points. Then they found the point set with the most points by region growth method and fit the plane parameters. Hoang Vu et al.\cite{8} combines gradient threshold segmentation and mean height evaluation to segment ground plane.

RANSAC algorithm realizes plane extraction even the image includes other planes. Whereas, the result of detection with the RANSAC algorithm is wrong in the scene where other large planes appear.
These large planes contain more 3D points than the ground plane. Hadsell et al. [9] propose a stereo-based solution to segmentation in which plane fitting on point cloud data is also used. The disadvantage is that the solution requires one to specify a-priori the number of planes to extract, and thus is not suitable for multi-plane environments.

In this paper, we present a novel and simple algorithm to detect the ground plane using occupancy grid map. It is not new to use occupancy grid map to detect the ground plane [10], but our approach gives a different method to calculate the occupancy. The paper is organized as follows: Section 2 includes some work about data pre-processing. Then our proposed algorithm is shown in Section 3. Section 4 includes experiments of our proposed algorithm. Finally, the paper is concluded in Section 5.

2. Data pre-processing
The depth image obtained by Kinect contains noise and black holes, so it is necessary to pre-process the depth image. Color and depth images must be converted to 3D point cloud data before.

2.1 Image pre-processing
A filter algorithm which is called the weight median filter (WM) is proposed. The WM filter involves two steps, row filling (see in figure 1) and weight median filling.

\[
W = e^{- \frac{((u-u_0)^2+(v-v_0)^2)}{\sigma_{\text{wnd}}} - \frac{(depth-depth_0)^2}{\sigma_{\text{dis}}}}
\]

where \(\sigma_{\text{wnd}}\) represents half the size of the window and \(\sigma_{\text{dis}}\) is a constant. \((u_0,v_0)\) is the center of the window.

2.2 Coordinate System Conversion
In coordinate transformation, the Kinect is considered as the center of the world coordinate system. In (2), we assume a pinhole camera having no skew.

\[
\begin{pmatrix}
x_k \\
y_k \\
z_k
\end{pmatrix} = \begin{pmatrix}
\frac{i_k - c_x}{f_x} z_k \\
\frac{j_k - c_y}{f_y} z_k \\
z_k
\end{pmatrix}
\]

\(Z_k\) represents the value of Z-axis coordinate with the corresponding 2D pixel \((i_k, j_k)\). \((i_k, j_k)^T\) is the pixel coordinates in the kth depth image. \((c_x, c_y)^T\) is the camera’s principal point, in pixels. \(f_x\) and \(f_y\)
represent focal length. In our experiments, the parameters of RGB-camera were retrieved from the datasets or from the calibration procedure of the device.

Synchronization of RGB-camera and Depth-camera is guaranteed by the timestamp from Microsoft Software Development Kit (SDK). On the basis of Eq.2, we can obtain the 3D point cloud data of the scene.

3. The Proposed Method
A new approach to detect ground plane (see in figure 2) is proposed which includes occupancy grid map construction, ground plane fitting and ground plane segmentation.

![Figure 2. Overview of the proposed ground plane detection algorithm](image)

3.1 Occupancy Grid Map Construction
It is the key step in the entire algorithm. The point in the occupancy grid map is either occupied or free. We can obtain more accurate data about obstacle in the occupancy grid map. This paper modifies the process of occupancy calculation to adapt to our algorithm. The paper updates the occupancy with binary Bayes filter.

![Figure 3. some points in the same grid](image)

The classic method of occupancy calculation is an estimation of the grid occupancy in the two dimensions, time and distance. We find that some points in the ground are mapped to the same grid (see in figure 3.). If you follow the classic method of occupancy calculation, the occupancy of points in the ground will be big. This is not suitable for our algorithm. In the paper, the points at the same location with different height are considered to construct occupancy grid map. Our method to calculate the occupancy is shown in figure 4.

![Figure 4. our method to construct the occupancy grid map](image)

In general, the lowest height point is definitely in the ground. Each point in the 3D point cloud data is compared to this point in height. The result of the comparison will guide to update the occupancy.

3.2 Ground Plane Segmentation
If a pixel is in the ground plane, the occupancy of the pixel is below zero. Some points on the ground will be selected based on the occupancy. Then these filtered 3D point cloud data are used for ground
plane fitting. This paper employs the way of Jann Poppinga [11] to obtain the normal vector of the ground plane.

Suppose we have a set of 3D points \( r_i = (x_i, y_i, z_i)^T, i = 1...k \) in the ground plane and we want to get the normal vector of these data. It can be shown that this is the eigenvector problem under the assumption of the isotropic Gaussian noise. The gravity center of the given data is defined as:

\[
r_{gc} = \frac{1}{k} \sum_{i=1}^{k} r_i
\]

Using this, the following matrix \( M_k \) is defined:

\[
M_k = \sum_{i=1}^{k} (r_i - r_{gc})(r_i - r_{gc})^T
\]

Then the normal vector of the ground plane is equal to the eigenvector \( n \) which corresponds to the smallest eigenvalue of the matrix . Suppose the resulting vector is normalized. The parameters of the plane equation are solved by the normal vector and the gravity center. On the basis of Eq.5, the distance from the point to the plane is obtained. If the distance is beyond the threshold, the points will be removed as outliers. The remaining points are back-projected to the pixel coordinate system to get the result of the ground plane detection.

\[
dis = \frac{|ax_i + by_i + cz_i + d|}{\sqrt{a^2 + b^2 + c^2}}
\]

4. Experiments
Color and depth images are converted to 3D point cloud data, since the input of the system are 3D point cloud data. The resolution of all input images is 640×480. In the following figure 5,6, we show the results of experiment on a group of depth images. They were obtained with a Kinect on the TurtleBot2. In the image pre-processing, the WM filter is proposed. The image in figure 5 is the original depth image. There are many black holes in the image, especially around the object. After the WM filter method, the number of holes in the image in figure 6 decreases significantly.

The results of the detection of the three scenes are shown in figure 7,8,9. Some points are in the ground, but the occupancy of them is beyond zero. A total of two reasons lead to the problem. The first reason is that the ground is not an ideal plane and the second reason is that we can’t guarantee that the Kinect is placed in a horizontal position. There are a lot of holes in the detection image without WM filter (see in figure 9). The number of points in the ground selected by occupancy (detect1) is less than the number of points actually in the ground (origin) (see in figure 10). Nevertheless, it guarantees the accuracy of the ground plane fitting. The result of the ground plane detection is correct. The result of ground plane detection does not cover all the ground in the image. The reason is that the limitation of the sensor's range.
Figure 7. result in simple scenes

Figure 8. result in complex scenes

Figure 9. result without WM filter

Figure 10. the number of points in the ground in three scenarios

Figure 11. the rate of successful detection in three scenarios

*detect1* represents the process of filtering points based on occupancy. *detect2* represents the process of filtering points based on distance. *origin* represents the original point cloud data only includes the points in the ground.

There is a significant increase in the number of points in the ground from *detect1* to *detect2*, which means that the calculation about the distance from the point to the plane plays a good role in ground plane segmentation. The Rate of successful detection describes the proportion of the ground points detected in the all ground points. The algorithm proposed can detect more than 95% of the points in the ground (see in figure 11).

Figure 12. the result of RANSAC in multi-plane environment

Figure 13. the result of RANSAC in simple scenes

Figure 14. the result of our method in multi-plane environment

Figure 15. the result of our method in simple scenes

The red points in figure 12,13 are the ground point detected by [4]. This detection algorithm is effective in the scene where there are no other large planes. Nevertheless, the result of detection is
incorrect in the scene where a large plane appears. It is not suitable for multi-plane environments. The green points in figure 14,15 are the ground points detected by the proposed method. The result of ground detection is correct in the scene where a large plane appears. In terms of data pre-processing, [4] adopted Voxelization in Point Cloud Library (PCL [12]), which lead to poor real-time performance.

5. Conclusion
We put forward a new ground plane detection algorithm, which consists of three stages: data pre-processing, occupancy grid map construction and ground plane segmentation. The algorithm of ground plane detection is computationally efficient, since the 3D point cloud data are reduced to two-dimensional occupancy grid map for processing. Moreover, the effect of the weight median filter is quite obvious. The algorithm can detect more than 95 percent of the ground point and the result of the ground plane detection is accurate. We also find that the results of the algorithm are correct in different scenes, even if the sensor has a change in pitch and roll angles. These contributions come from the method to calculate the occupancy we introduced above. The algorithm proposed in this paper is not only computational efficiency but also robust.

Acknowledgments
This work was supported by the Postgraduate Research & Practice Innovation Program of Jiangsu Province no.KYCX18_0540 under grant agreement no.2018B733X14.

References
[1] Chen W, Yue H, Wu X, et al.: 'Real-time obstacle detection for legged robots using the Kinect sensor'. Advanced Robotics, 2014, 28(20), pp. 1375-1387
[2] Harakeh, Ali , D. Asmar , and E. Shammas . "Ground Segmentation and Occupancy Grid Generation Using Probability Fields." International Conference on Intelligent Robots & Systems 2015.
[3] Vera E, Lucio D, Fernandes L A F, et al.: 'Hough Transform for Real-Time Plane Detection in Depth Images', Pattern Recognition Letters, 2018
[4] Ramy Ashraf Zeineldin, Nawal El-Fishawy: 'Fast and accurate ground plane detection for the visually impaired from 3D organized point clouds Conference'. 2016 SAI Computing Conference (SAI), London, United Kingdom, July, 2016, pp. 373-379
[5] R?Fer, Thomas , et al.: 'Real-Time Plane Segmentation Using RGB-D Cameras', in 'Computer Science' (2012), pp. 306-317
[6] Hong Liu, Yongqing Jin, Chenyang Zhao.: 'Real-time trust region ground plane segmentation for monoclar mobile robots'. 2017 IEEE International Conference on Robotics and Biomimetics (ROBIO), Macao, China, December, 2017, pp. 952-958
[7] H?Hnel D, Burgard W, Thrun S.: 'Learning compact 3D models of indoor and outdoor environments with a mobile robot', Robotics and Autonomous Systems, 2003, 44(1), pp. 15-27
[8] Vu, Hoang , et al. "A Ground Segmentation Method Based on Gradient Fields for 3D Point Clouds.", 2018.
[9] Vijayanagar, K.R., Loghman, M., Joohee Kim.: 'Refinement of depth maps generated by low-cost depth sensors'. SoC Design Conference (ISOCC), Jeju Island, South Korea, November, 2012
[10] C. Braillon, C. Pradalier, K. Usher, J. Crowley, and C. Laugier,“Occupancy grids from stereo and optical flow data,” in ISER, 2006.
[11] Jann Poppinga, Narunas Vaskevicius, et al.: 'Fast Plane Detection and Polygonalization in Noisy 3D Range Images'. IEEE/RSJ International Conference on Intelligent Robots and Systems, Nice, France, Sept, 2008, pp. 22-26
[12] Hadsef R, Sermanet P, Scoffier M, et al.: 'Learning Long-Range Vision for Autonomous Off-Road Driving'. Journal of Field Robotics, 2009, pp. 120-144