

Ground Plane Detection Using an RGB-D Sensor

Doğan Kırçalı and F. Boray Tek

Abstract Ground plane detection is essential for successful navigation of vision based mobile robots. We introduce a very simple but robust ground plane detection method based on depth information obtained using an RGB-Depth sensor. We present two different variations of the method: the simplest one is robust in setups where the sensor pitch angle is fixed and has no roll, whereas the second one can handle changes in pitch and roll angles. Our comparisons show that our approach performs better than the vertical disparity approach. It produces accurate ground plane-obstacle segmentation for difficult scenes, which include many obstacles, different floor surfaces, stairs, and narrow corridors.

Keywords Ground plane detection · Kinect · Depth-map · RGB-D · Autonomous robot navigation · Obstacle detection · V-disparity

1 Introduction

Ground plane detection and obstacle detection are essential tasks to determine passable regions for autonomous navigation. To detect the ground plane in a scene the most common approach is to utilize depth information (i.e., depth map). Recent introduction of RGB-D sensors (Red-Green-Blue-Depth) allowed affordable and easy computation of depth maps. Microsoft Kinect— is the pioneer of such sensors— integrates an infrared (IR) projector, a RGB camera, a monochrome IR camera, a tilt motor and a microphone array to provide a 640×480 pixels depth map and RGB video stream at a rate of 30 fps.

Kinect uses an IR laser projector to cast a structured light pattern over the scene. Simultaneously, its monochrome CMOS IR camera acquires an image. The

D. Kırçalı (✉) · F.B. Tek
Robotics and Autonomous Vehicles Laboratory, Computer Engineering Department,
Işık University, 34980 Şile, İstanbul, Turkey
e-mail: dogan@isikun.edu.tr

F.B. Tek
e-mail: boray@isikun.edu.tr

disparities between the expected and the observed patterns are used to estimate a depth value for each pixel. Kinect works well indoor. However, the depth reading is not reliable for regions that are far more than 4 m; at the boundaries of the objects because of the shadowing; reflective or IR absorbing surfaces; and at the places that are illuminated directly by sunlight which causes IR interference. Accuracy under different conditions was studied in [1–3].

Regardless of the method or the device that is used to obtain the depth map, several works approach to the ground plane detection problem based on the relationship between a pixel's position and its disparity [4–9].

Li et al. show that the vertical position (y) of a pixel of the ground plane is linearly related to its disparity $D(y)$ such that one can seek a linear equation $D(y) = K_1 + K_2 * y$, where K_1 and K_2 are constants, which are determined by the sensor's intrinsic parameters, height, and tilt angle. However, ground plane can be directly estimated on the image coordinates using the plane equation based on disparity $D(x, y) = ax + by + c$ without determining mentioned parameters. A least squares estimation of the ground plane can be performed offline (i.e., by pre-calibration) if a ground plane only depth image of the scene is available [5]. Another common approach is to use RANSAC algorithm which allows fitting of the ground plane even the image includes other planes [4, 10, 11].

Some other approaches aim to segment the scene into relevant planes [11, 12]. The work of Holz et al. clusters surface normals to segment planes and reported to be accurate in close ranges [11].

In [7], histograms of the disparity image rows were used to model the ground plane. In the image formed of the row histograms (named as *V-disparity*), the ground plane appears as a diagonal line. This line, which is detected by Hough Transform, was used as the ground plane model.

In this paper, we present a novel and simple algorithm to detect the ground plane without the assumption that it is the largest region. Assuming a planar ground plane model which may probably cause problems if the floor has significant inclination or declination [6, 7], we use the fact that if a pixel is from the ground plane, its depth value must be on a rationally increasing curve placed on its vertical position. Although the degree of this curve is not known, it can be estimated by an exponential curve fit to use it as the ground plane model. Later, the pixels which are consistent with the model are detected as ground plane whereas the others are marked as obstacles. While this is our base model which can be used for a fixed viewing angle scenario, we provide an extension of it for dynamic environments where sensor viewing angle changes from frame to frame. Moreover, we note the relation of our approach to V-disparity approach [7], which rely on the linear increase of disparity and fitting of a line to model the ground plane, and compare our method by tests on the same data.

2 Method

2.1 Detection for Fixed Pitch

In a common scenario, the sensor views the ground plane with an angle (i.e., pitch angle), in which we can assume that the sensor is fixed and its roll angle is zero (Fig. 1b). The sensor’s pitch angle (Fig. 1a) causes allocation of more pixels for the closer scene than the farther. So that linear distance from the sensor is projected on the depth map as a rational function. This is demonstrated in (Fig. 1c). Any column of the depth image shows that the depth value increases not linearly but rationally from bottom to top (i.e., right to left in Fig. 1d). Furthermore, a “ground plane only” depth image must have all columns equal to each other, which is estimable by a curve fit of sum of two exponential functions in the following form:

$$f(x) = ae^{bx} + ce^{dx} \tag{1}$$

where $f(x)$ is the pixel’s depth value and x is the its vertical location (i.e., row index) in the image. The coefficients (a, b, c, d) depend on the intrinsic parameters, pitch angle, and the height of the sensor.

A least squares fitting estimation of these coefficients make it possible to reconstruct a curve, which is named as the *reference ground plane curve* (C_R). In order to detect ground plane pixels in a new depth frame, its columns(C_U) are compared to C_R ; any value under C_R represents an object (or a protrusion), whereas values above

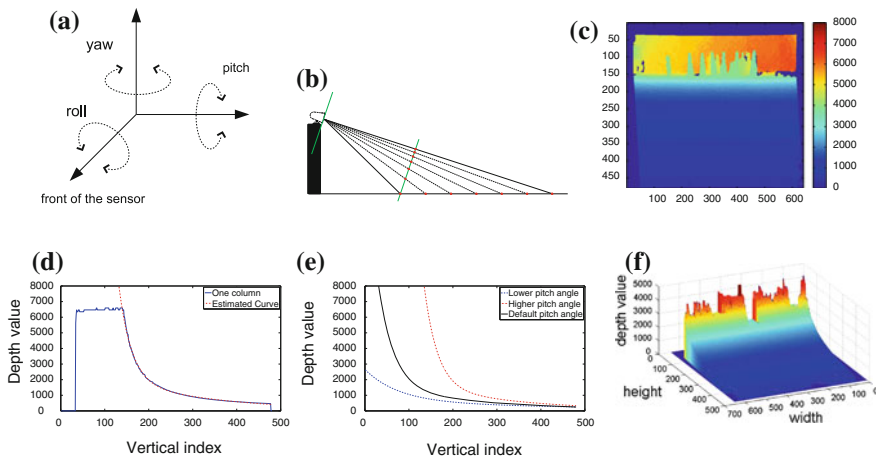


Fig. 1 a Roll & pitch axis, b sensor view pitch causes linearly spaced points to mapped as an exponential increasing function. c An example depth map image, d one column ($y = 517$) of the depth map and its fitted *curve* representing the ground plane, e ground plane curves for different pitch angles, f depth map in three dimensions showing the drop-offs caused by the objects

the reference curve represent drop-offs, holes (e.g., intrusions, downstairs, edge of a table). Hence, we compare the absolute difference against a pre-defined threshold value T ; mark the pixels as ground plane if difference is less than T . Here, the depth values that are equal to zero were ignored as they indicate sensor reading errors. The related experiments are in Sect. 3.

2.2 Detection for Changing Pitch and Roll

The fixed pitch angle scheme explained above is quite robust. However, it is not suitable for the scenarios where the pitch and roll angles of the sensor changes. Such as of the mobile robots that cause movements on the sensors' platform. These can be compensated by using an additional gyroscopic stabilization [13]. However, here we propose a computational solution in which a new reference ground curve is estimated for each new input frame.

A higher pitch angle (sensor almost parallel to the ground) will increase the slope of the ground plane curve. Whereas a non-zero roll angle (horizontal angular change) of the sensor forms different ground plane curves along columns of the depth map (Fig. 1e). Such that at one end the depth map exhibits curves of higher pitch angles, while toward the other end, it has curves of lower pitch angles, which complicate the use of a single reference curve for that frame.

To overcome the roll angle effects our approach aims to rotate the depth map to make it orthogonal to the ground plane. If the sensor is orthogonal to the ground plane it is expected to produce equal or very similar depth values along every horizontal line (i.e., rows), which can be captured by a histogram of the row values such that a higher histogram peak value indicates more similar values along a row. Let h_r shows the histogram of the r th row of a depth image (D) of R rows, and let us denote the rotation of depth image with D_θ .

$$\operatorname{argmax}_\theta \left(\sum_{r=1}^R \operatorname{argmax}_i (h_r(i, D_\theta)) \right) \quad (2)$$

Thus for each angle value θ in a predefined set, the depth map is rotated with an angle θ and the histogram h_r is computed for every row r . Then, the angle θ that maximizes the sum of the histogram peak values is estimated as the best angle to rotate the depth map prior to the ground plane curve estimation. This removes the roll angle effect.

The changes of pitch angle create different projection and different curves along the image columns (Fig. 1e). However, in a scene that consists of both the ground plane and objects, the maximum value along a particular row of the depth map must be due the ground plane, unless an object is covering the whole row (as in Fig. 1f). This is because the objects are closer to the sensor than the ground plane surface that they occlude. Therefore, if the maximum value across each row (r) of the depth map

(D) is taken, which we name as the *depth envelope* (E), it can be used to estimate the *reference ground plane curve* (C_R) for this particular scene and frame.

$$E(r) = \max_i(D(c_i, r)) \quad (3)$$

The estimation is again performed by fitting the aforementioned exponential curve (1). Prior to the curve fitting we perform median filtering to smooth the depth envelope. Moreover, depth values must increase exponentially from bottom of the scene to the top. However, when the scene ends with a wall or group of obstacles this is reflected as a plateau in the depth envelope. Hence, the envelope (E) is scanned from right to left and the values after the highest peak are excluded from fitting as they cannot be a part of the ground plane. After the curve is estimated pixels of the frame are classified, as described in Sect. 2.1.

Two conditions affect the ground plane curve fit adversely. First, when one or more objects cover an entire row, this will produce a plateau in the profile of the depth map. However, if the rows of the “entire row covering object or group” do not form the highest plateau in the image, ground plane curve continues afterwards and the object will not affect the curve estimation. Second, drop-offs of the scene cause sudden increases (hills) on the depth envelope because they exhibit depth values higher than the ground plane’s: If a hill is found on the depth envelope, the estimated curve will be produced by a higher fitting error.

3 Experiments

We tested our algorithm on four different datasets comprised of several 640×480 frames. Dataset-1 and dataset-2 were composed of 300 frames captured on a robot platform which moves on the floor among several obstacles. Dataset-3 was created with the same platform; however, the pitch and roll angles change excessively. Dataset-4 included 12 individual frames acquired from difficult scenes such as narrow corridors, wall only scenes etc. Dataset-1 and dataset-2 were manually labeled to provide the ground truth and were used in plotting ROCs (Receiver Operating Curve), whereas the other two were visually examined.

We compared three different versions for our approach: fixed pitch (A1), pitch compensated (A2), pitch and roll compensated (A3). There is only one free parameter for A1 and A2 that is the threshold T , which is estimated by ROC analysis, whereas the third roll compensation algorithm requires a pre-defined angle set for the search for best rotation angle: $\{-30^\circ, -28^\circ, \dots, +30^\circ\}$. Least squares fit was performed by Matlab curve fitting function with default parameters.

Moreover, we compared the results with V-disp method [7], which is originally developed for stereo depth calculation where the disparity is available before depth. To implement V-disp method, we calculated disparity from the Kinect depth map (i.e., $1/D$); calculated the row histograms to form V-disp image; and then run Hough

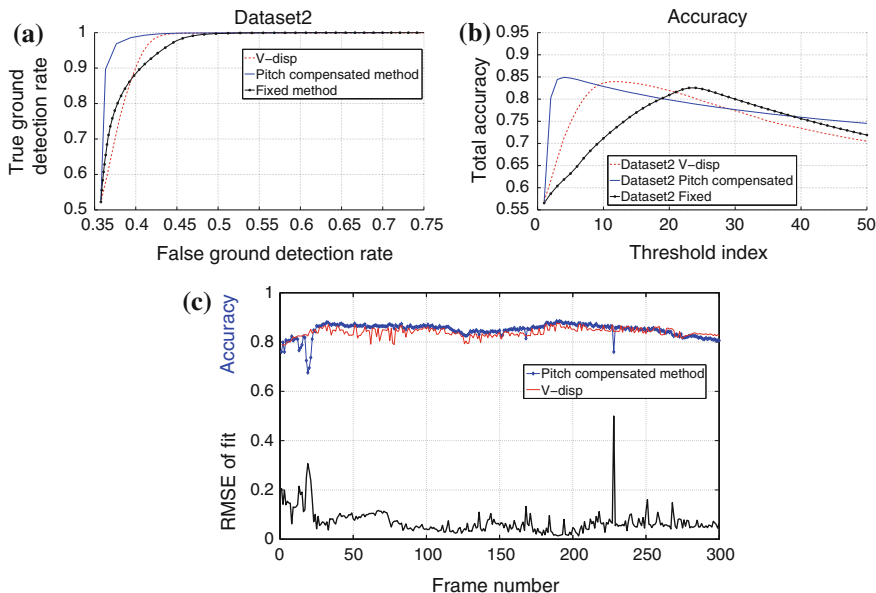


Fig. 2 **a** ROC curves comparing V-disp and our fixed and pitch compensated algorithms (A1-A2), **b** average accuracy over 300 frames versus thresholds, **c** accuracy and *curve* fit error of A2 for individual frames

transform to estimate the ground plane line. We put a constraint on the Hough line search in $[-60^\circ, -30^\circ]$ range.

Since A3 and A2 algorithms are same, except for the roll compensation, we examine and compare results of A2 to A1 and V-disp; however, we compare A3 results only against A2 to show the effect of the roll compensation.

Figure 2a, b show ROC curves and overall accuracies plotted for our fixed and pitch compensated algorithms (A1 and A2) and V-disp method on dataset-2. It can be seen that our pitch compensated algorithm is superior to V-disp which is better than our fixed algorithm.

When we select the best accuracy point thresholds and run our algorithms on dataset-2, we obtained accuracy versus frames (Fig. 2c). In addition, we recorded the curve fitting error for the pitch compensated algorithm (A2). Both methods were quite stable with the exception of some high curve fitting error frames for A2. Those frames can be automatically rejected to improve accuracy.

Some example inputs and outputs of our algorithm A2 is shown in Fig. 3. The examples include a cluttered scene (Fig. 3a-c), stairs (Fig. 3d-f), one of the frames from dataset-3, where the sensor is rolled almost 20° degrees (Fig. 3g). Figure 3h, i shows the respective outputs of A2 and A3. It can be seen that the roll compensation provides a significant advantage.

Finally, Fig. 3j, k shows output pairs (overlaid on RGB) for A2 and V-disp. Both methods detect the ground planes in the scenes where ground plane is not the largest

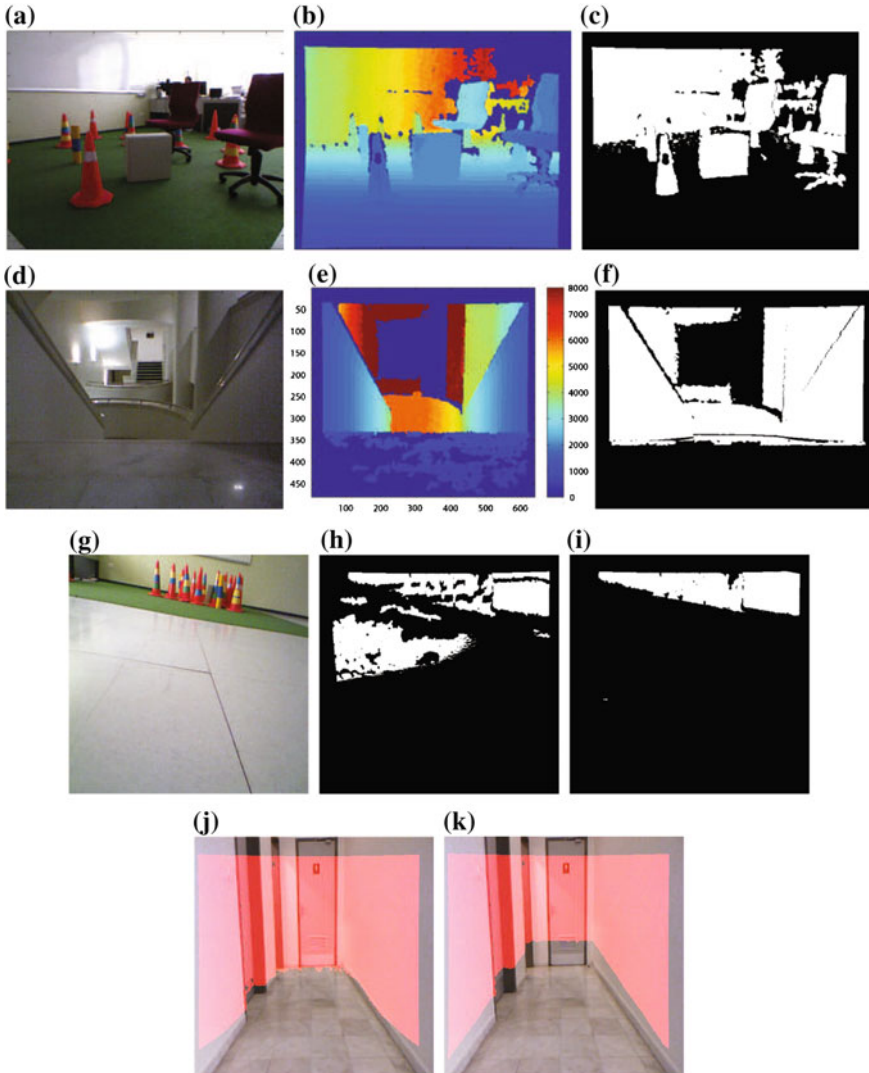


Fig. 3 Experimental results from different scenes. RGB, depth-map and pitch compensated method output (*white* pixels represent objects whereas black pixels represent ground plane): **a, b, c** lab environment with many objects and reflections; **d, e, f** stairs **g, h, i** respective outputs of pitch compensated (A-2) and pitch&roll compensated method on an image where sensor was positioned with a roll angle (A-3). **j, k** Comparison of pitch compensated (*left*) and V-disp method (*right*) in narrow corridor

nor the dominant plane. Note that A2 is better than V-disp, though the thresholds used by both methods were determined for the highest respective overall accuracy for dataset-1, -2.

If the frames are buffered beforehand, our algorithm A2 processed 83 fps on a Pentium i5 processor using Matlab 2011a. Datasets and more results can be found in our web site.

4 Conclusion

We have presented a novel, and robust ground plane detection algorithm which uses depth information obtained from an RGB-D sensor. Our approach includes two different methods, where the first one is simple but quite robust for fixed pitch and no-roll angle scenarios, whereas the second one is more suitable for dynamic environments. Both algorithms are based on an exponential curve fit to model the ground plane which exhibits rationally increasing depth values. We compared our method to the popular V-disp [7] method which is based on detection of a ground plane model line by Hough transform which relied on linear increasing disparity values. We have shown that the proposed method is better than V-disp and produces acceptable and useful ground plane-obstacle segmentation for many difficult scenes, which included many obstacles, different surfaces, stairs, and narrow corridors.

Our method produce errors especially when the curve fitting is not successful. Our future work will focus on these situations that are easy to detect by checking the RMS error of the fit, which has been shown to be highly correlated with the accuracy of segmentation.

Acknowledgments This study was supported by FMV Işık University Internal Research Funding Grants project BAP-10B302.

References

1. J. Stowers, M. Hayes, A. Bainbridge-Smith, Altitude control of a quadrotor helicopter using depth map from microsoft kinect sensor. in *IEEE International Conference on Mechatronics (ICM)*, pp. 358–362 (2011)
2. C. Rougier, E. Auvinet, J. Rousseau, M. Mignotte, J. Meunier, Fall detection from depth map video sequences. in *International Conference on Smart Homes and Health Telematics*, pp. 121–128 (2011)
3. K. Khoshelham, S.O. Elberink, Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors* **12**(2), 1437–1454 (2012)
4. F. Li, J.M. Brady, I. Reid, H. Hu, Parallel image processing for object tracking using disparity information. in *2nd Asian Conf. on Computer Vision ACCV*, pp. 762–766 (1995)
5. S. Se, M. Brady, Ground plane estimation, error analysis and applications. *Robot. Auton. Syst.* **39**(2), 59–71 (2002)
6. Q. Yu, H. Araújo, H. Wang, A stereovision method for obstacle detection and tracking in non-flat urban environments. *Auton. Robots* **19**(2), 141–157 (2005)
7. R. Labayrade, D. Aubert, J.P Tarel, Real time obstacle detection in stereovision on non flat road geometry through “v-disparity” representation. in *IEEE Intelligent Vehicle Symposium*, vol. 2, pp. 646–651 (2002)

8. C.J. Taylor, A. Cowley, Parsing indoor scenes using rgb-d imagery. in *Robotics: Science and Systems* July 2012
9. K. Gong, R. Green, Ground-plane detection using stereo depth values for wheelchair guidance. in *Proceedings 7th Image and Vision Computing New Zealand*, pp. 97–101 (2009)
10. C. Zheng, R. Green, Feature recognition and obstacle detection for drive assistance in indoor environments. in *Proceedings 11th Image and Vision Computing New Zealand* (2011)
11. D. Holz, S. Holzer, R. Bogdan Rusu, S. Behnke, Real-Time Plane Segmentation using RGB-D Cameras. in *Proceedings of the 15th RoboCup International Symposium*, vol. 7416, pp. 307–317, Istanbul, Turkey, July 2011
12. C. Erdogan, M. Paluri, F. Dellaert, Planar segmentation of rgb-d images using fast linear fitting and markov chain monte carlo. in *The 12' Conference on Computer and Robot Vision*, pp. 32–39 (2012)
13. L. Wang, R. Vanderhout, T. Shi, Computer vision detection of negative obstacles with the microsoft kinect. *Uni. British Columbia. ENPH 459 Reports* (2012)