An Effective Ground Plane Extraction using Depth Map Estimation from a Kinect Device

Dang Khanh Hoa[*], Pham The Cuong, Nguyen Tien Dzung

Hanoi University of Science and Technology, No. 1, Dai Co Viet, Hai Ba Trung, Hanoi, Viet Nam

Abstract

This paper presents a new approach to extract ground planes from a depth map which is provided by Kinect. The proposed system applies an robust algorithm to calculate the depth gradient maps (GDM) with high accuracy. Then the correct partition provides a set of candidates for the selection of ground. Last, it uses an efficient filter to find out the truth ground planes. The results prove the certainty of the algorithm in both cases consisting of the perfect data and actual scenes. For first case, the percentage of truth ground pixel detection R1 is common over 90%. The percentage of incorrect ground pixels detection R2 is lower than 5%. For the second case, the process of implementing the proposed algorithm on a depth map from Kinect also is compared with RANSAC algorithm and Enhanced V-Disparity algorithm. The result demonstrates that the proposed method's R1 is usually greater than RANSAC method and V-Disparity method 2%, while R2 of the proposed method is less than half of R2 of the compared methods, respectively. The experimental results show the ability to respond in real time when this work is deployed as a stereo vision-based navigation system.

Keywords: Depth map, gradient, ground plane, Kinect, vehicle

1. Introduction

In recent years, visual-base navigation field for robots is more and more interesting. Many researchers propose new approachs to extract infomation from images for controling a mobile robot or a wheel vehicle. The results are usually evaluated by two basic criteria including implementation time and accuracy. The calculation speed are often improved when the system select a compact data such as single image or applies an low complexity algorithms. But if it requires to increase the accuracy, the approach is a lot more complicated. Some recent results are amazing in specific cases. Several papers base on the classic algorithms such as improved RANSAC algorithm [1-4] or Hough transform [5]. Even there is an conjunction of both algorithms to distinguish flat land certainly. It is clear that the high complexity is still a difficult issue which is not be solved. It is hard to implement these works on an embedded system. So a navigation system could only work in real time condition easily if it is equipped with a powerful hardware. High accuracy of results depend on the optimized parameters of RANSAC algorithm. In fact, they are not stable when a vehicle is put into an fast variable environment. Moreover, the method of Probabilistic Hough transformation has a pretty criticism with large volumes of 3D point cloud data [6-7]. An image homograph method has been demonstrated in [8-9] with an simple calculation. But this method' results is only suitable for environments with non-complex ground. The works using a data stream collected from a single camera are quite outstanding [10-12]. They only demand a simple image acquisition system without depth data. The proposed process take a 2D color image sequence as an input which consists of three RGB basic color channels. Therefore, the actual number of operations is three times larger than case of gray image input.

The articles [13-15] focus on exploiting the difference map to reduce the volumes of input data. [13] The proposed method compares the difference of disparity values on each line in a disparity map. So the robustness of the algorithm is not high if the source is probably affected by the context of scene. In a real difference map, it is evident that there is always noise that are appeared by many external and internal factors. Moreover the results are only shown in the limited context with non-obstructions. The improved V-Disparity algorithm [14] lead the rate of detected ground point higher than but it was compensated for by the complexity of the handling process which is integrated one kind of Hough transform or RANSAC algorithm in order to filter the raw results. In addition, the results can only be assessed well in outdoor limited environments with no sidewalks.

[*] Corresponding author: Tel.: 0989123114
Email: hoa.dangkhanh@hust.edu.vn

In our previous work [16] the GDM algorithm has been successfully applied to some input data.. In this paper, the GDM will be modified to achieve the ground extraction from depth maps which are supported by Kinect. The experienced results is compared with other classic and recent approachs, such as RANSAC algorithm and Enhanced V-Disparity.

The paper is structured in five sessions where Section 2 introduces some basic mathematical fundamentals in differential depth problem. Section 3 then illustrates the implementation of proposed method. Section 4 discusses the experimental results and performance evaluation and followed by conclusion and future work in Section 5.

2. The basis mathematics theory of system

Capture devices with focal length f are placed at O with height h from the ground and its direction is parallel to the ground which is considered to be flat absolutely (Fig. 1.) [16]. Let O' be the perpendicular projection of O on the ground plane and O'' be the projection of O on the image . Let M_1 be a considered ground points and p denotes a distance from O'' to the projection of M_1 on the image. Assuming that OO'' is parallel to $O'M_1$, we have the distance z from O to M_1:

$$z = h\sqrt{1 + \frac{f^2}{p^2}} \qquad (1)$$

Taking the differential both sides of formula (1), it leads to (2) as below:

$$dz = -hf \frac{1}{p^2\sqrt{p^2 + f^2}} d\rho \qquad (2)$$

From (2) and the actual figures of the camera we can approximate the differential Δz follow Δp .

Suppose there is one more point M_2 that is also located on the ground plane and M_1M_2 is perpendicular to the direction of the camera's view. Then the images of M_1M_2 is a horizontal segments. The length of that segment:

$$\Delta z = \sqrt{\Delta M^2 + z^2} - z \qquad (3)$$

From (3), the differential of horizontal direction Δz will be smaller than z. Now, assuming the depth images obtained from the camera horizontally x axis (direction from left to right), the vertical axis is y (direction from top to bottom); depth value is quantized and only get a finite value. Considering the obtained image, it is a digital image so from (3) we find that the depth difference of the ground along x axis (called $gradient\_x$) will be 0 and the other only be kept within near zero.
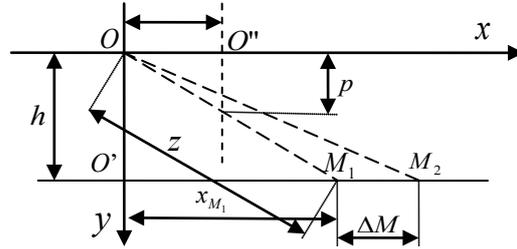

Fig. 1. Principles of calculation the differential depth.

From (2) and going over a few simple changes, we can see the differential depth along y axis (called $gradient\_y$) is different from zero for the pixels which have a depth value greater than a certain value T. And it maybe get zero if the point's depth is less than T .

Thus, we can assume that adjacent pixels belong to a region if both x-axis gradient and y-axis gradient are equal, respectively. The set of regions with $gradient\_x = 0$ and $gradient\_y \neq 0$ (maybe be nix if the depth of the plane is small or scope lies entirely in the bottom of the images. The size of this area compared with quarter-size images) will create shape the ground in the images.

3. The system implementation

The block diagram of the ground extraction system is depicted in Fig. 2, where the GDM plays an important role to detect planes captured by the Kinect's camera sensor from the input block. The candidate planes will be finally detected after refining process to remove unreliable planes.
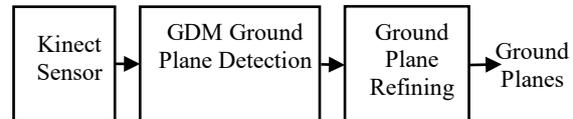

Fig. 2. Block diagram of the implementation system.

3.1 Kinect Sensor

Kinect [17] is a device that provides input data for a system of self-propelled vehicles, where multiple sensors are mounted in the device. One of the input data that the Kinect can support is the VGA monochrome depth video stream with 11 bits to store the depth values. The frame rate of video stream is up to 30 Hz that shows a relative smooth motion. The

angular field of view is 57° horizontal and 43° for vertical angle. As the geometric parameters of the Kinect's camera are shown in Fig. 3, the distance from the camera O to a given object P is 0.8m. The image size obtained by the camera then is approximately 87 cm in horizontal and 63 cm in vertical, respectively. It's equivalent to 1.3 mm per pixel in the resolution point of view.
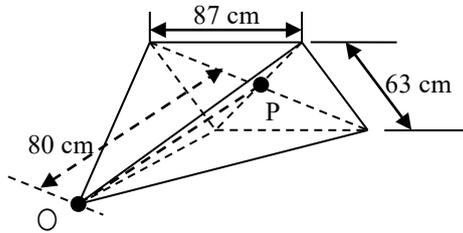

Fig. 3. Geometric parameters of the Kinect's camera

3.2 Ground Plane Calculation based on GDM

The block diagram of the proposed method is presented in Fig. 4. The first stage calculates $gradient\_x$ and $gradient\_y$ for each pixel using the depth map as an input to construct a gradient depth map. Then, the second stage groups the adjacent pixels that have a similar gradient into a range. The candidate ground plane is then formed by the ranges that meet the ground hypotheses. Since the candidate ground plane typically are affected by sporadic and random noise, the final stage will have to refine noise to construct the final ground plane by splitting the input image into blocks of size B.
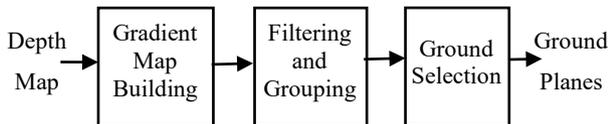

Fig. 4. A block diagram of the GDM Ground Plane Calculation

3.2.1 Construction of Gradient Depth Map

The task of this stage is to create a map of depth difference, also called a gradient map from the depth map input performed by calculation of gradients in y and x directions using equations (2) and (3) between two successive points, respectively. The resulting gradient depth map is further smoothed by consideration of depth difference of a point within a given window of size w, because of possible presence of noise in the input depth map.

3.2.2 Filtering and grouping

The goal of this stage is to group the points having similar gradients in the gradient depth map into a homogeneous region called range, and then eliminate inappropriate regions which do not satisfy the following constraints of the ground plane:

- The number of pixels of the region must be greater than a predetermined threshold;

- $gradient\_x = 0$ and $gradient\_y \neq 0$; or if $gradient\_x = 0$ and $gradient\_y = 0$ then the region must be located completely in the quarter area from the bottom of the input image for higher accuracy in ground plane detection.

The grouping and elimination algorithms are illustrated in the pseudo code in Fig. 5.

```
//Algorithm: Grouping.
//Input: Image.
//Output: Ranges
for each pixel do
    if this pixel is not in other Collection then
       Range add this pixel
       Range add all pixels satisfying Range's
    conditions
    end if
    if number pixel of Range > Range threshold
    and Range satisfies Ground Plane's conditions
    then
       Ranges Add Range
    end if
    Renew Range
end for
```

Fig. 5. The grouping and elimination algorithms

As the result, the ground plane from the acquired image would be roughly determined.

3.2.3 Ground Selection

In order to extract more exact and smooth ground plane, this correction stage starts dividing the initial difference depth map into square blocks of size B and then estimates the ratio R between the ground pixels inside each block and the block size. This is an important parameter used to classify the blocks into ground or non-ground ones and then generate the final map which includes ground and non-ground regions. If R is greater than a given threshold θ, then the block is considered as ground, and vice versa. In order to evaluate the value of θ, the smallest rectangular bounding the detected ground regions is determined, and the ratio between the number of all ground pixels $Pground\_of\_ranges$ over the square of the rectangular $Prec$ as depicted in equation (4).

$$\theta = \frac{\sum P_{ground\_of\_ranges}}{P_{rec}} \qquad (4)$$

Obviously the non-ground areas which belong to obstacles appearing with large enough size would be detected. The pseudo code illustrates the algorithm is shown in Fig. 6.

```
//Algorithm: Dividing Block.
//Input: Ranges.
//Output: Ground Plane
Divide gradient depth map into blocks of size B
Calculate threshold θ
for each block in Image do
    Ratio R of this block = number of ground
    pixel in a block / block size
    if R > θ then
            This block is assigned as Ground
            Plane
    else
            This block is eliminated
    end if
end for
```

Fig. 6.   The ground selection algorithms

3.3 Ground Plane Refining

This is a mandatory work that aims to make the ground plane with higher reliability. In fact, the depth map calculated by Kinect is not perfect because the noise always appears while capturing. This interference has two kind of forms. The first interference is in the form of small pieces where their distribution is scattered on the ground. And the second interference is in form of a large array. Both of them appear at the locations where the reflectance infrared signal is too weak for Kinect's receiver. The proposed solutions is going to fill fully error black holes using smoothing windows B. The program experiments with some appropriate smoothing window size for the purpose of looking out a most suitable window size value. Continuing with the second case of the black clusters error where Kinect cannot be determined in depth, there is no a feasible algorithm to remove them because of the depth map loses a lot of convergent depth information. However, these errors usually occur in remote locations which are quite far from the vehicle mounted a Kinect. Moreover, this phenomenon can be reduced or disappear when vehicles move cause the change of the signal's angle of reflection to the receiver on the Kinect sensor.

4. Experience Results And Discussion

Firstly, the proposed algorithm is tested on the depth maps with high quality from the Middlebury's library [18-19] as depicted in Fig. 7. In the ground plane refining step, the algorithm uses three smoothing windows denoted by B with different sizes

to compare the results with each other as shown in Fig. 8. On a visual assessment, as larger the smoothing window B is, the error detected ground pixels are higher increased in non-ground zones.
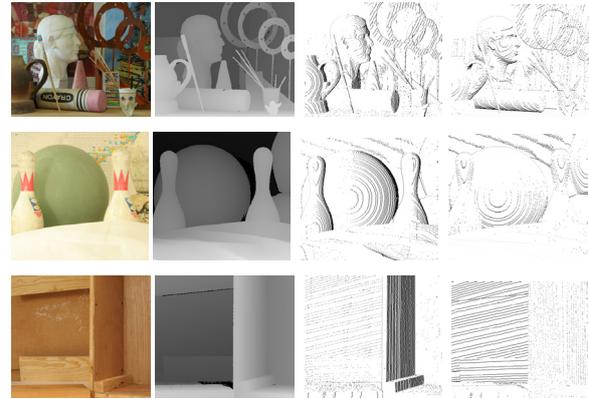

Fig. 7.   The gradient maps of the tested Midlebury images.

From left to right, the first colum is color images, the second colum is depth images, the third colum is x-gradient maps, the fourth colum is y-gradient maps; From top to down, the first row is Art image, the second row is Bowling1 image and the last row is Wood1 image, respectively.
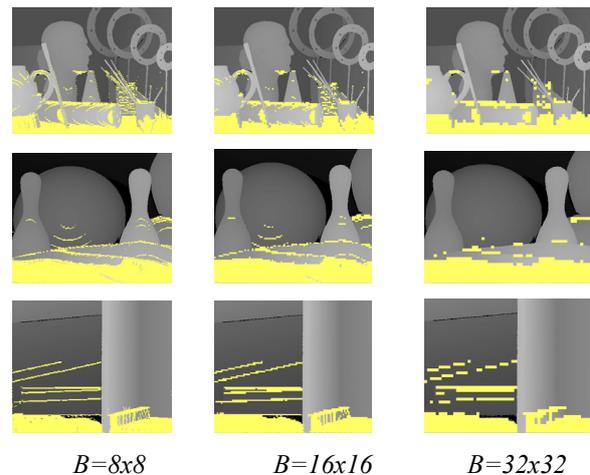

B=8x8	*B=16x16*	*B=32x32*

Fig. 8.   The results of the tested images in the plenty of case study.

From top to down, the first row is Art image, the second row is Bowling1 image and the last row is Wood1 image, respectively.

Secondly, the results are performed on four depth maps acquired by a Kinect in case with or without obstacles, less and more obstacles, simple and complex background, respectively as demonstrated in Fig. 9. The boundaries of the extracted ground are uniform all over the actual ground of the given scene. In the illustrated results, the ground areas are detected without confusion with

the surrounding obstacles at different sizes. As one can see, the detected grounds are completely matched with the actual ground areas. However, a few small holes have appeared where the algorithm has considered as non-ground areas from the depth map due to the Kinect's sensors. In order to evaluate the effectiveness of the proposed method, the percentage rate of detected ground pixels R1 and the percentage rate of incorrect detected ground pixels R2 are commonly determined within a given smoothing window size B. In this framework, three windows size 8 × 8, 16 × 16 and 32×32 respectively as shown in Fig. 10. In case of non-obstacles depth map, the proposed method outperforms with R1 greater than 96% and the value of R2 less than 2% (see Fig.11).
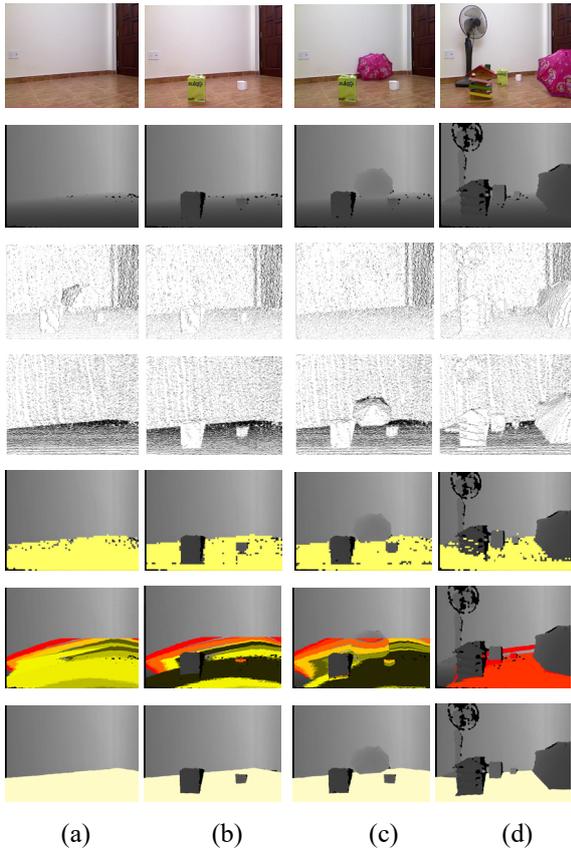

(a) (b) (c) (d)

Fig. 9. The results of the tested images in the plenty of case study.

From top to bottom, the first row is color images, the second row is depth images, the third row is x-gradient maps, the fourth row is y-gradient maps, the fifth row is detected ground planes, the sixth row is 3D RANSAC algorithm's detected ground planes and the last row is the truth ground planes, respectively;

In series of depth map containing the obstacles on the ground, the percentage of correct detected ground pixels R1 is best and stable at window size B = 16 × 16 (see Fig. 10). As complexity of ground

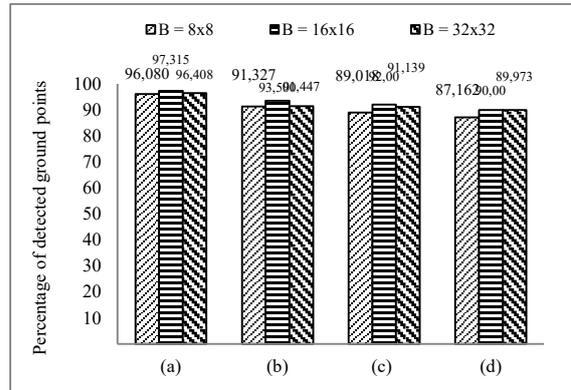detection process is increased, value of R2 also increased to around 5% (see Fig. 11.)

Fig. 10. The rate of detected ground pixels *R1* according to the smoothing window sizes *B*.
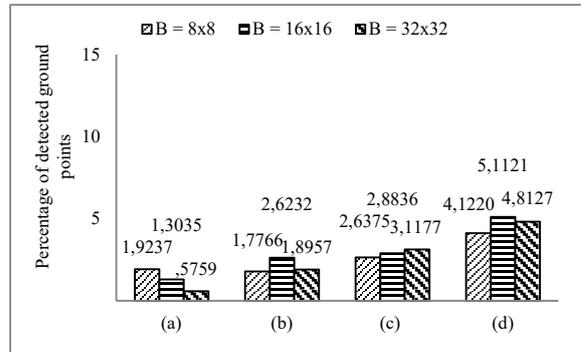

Fig. 11. The rate of error ground pixels according to the smoothing window sizes *B*.
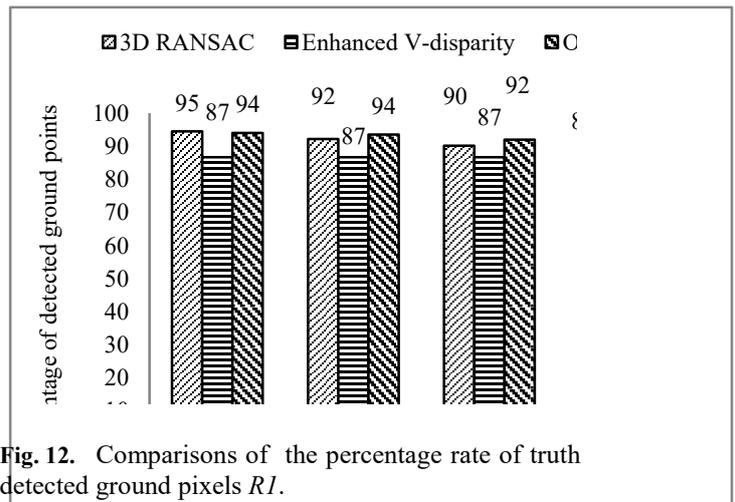

Fig. 12. Comparisons of the percentage rate of truth detected ground pixels *R1*.
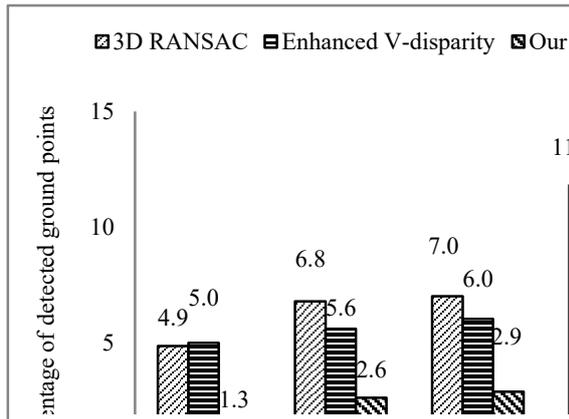
Fig. 13. Comparisons of percentage rate of wrong detected ground pixels *R2*.

Moreover, the *R1* and *R2* of this work are compared with the results of that for 3D RANSAC algorithm used in [1] and V-Disparity method used in [14] as illustrated in Fig. 12 and Fig. 13, respectively. The rate *R1* of the proposed method is greater than *R1* of 3D RANSAC and Enhanced V-Disparity methods. These comparisions are implemented with the optimal size of window *B = 16 × 16*. Meanwhile, the rate *R2* of the proposed method is always lowest among those performed by 3D RANSAC and Enhanced V-Disparity, respectively.

5. Conclusions

In this paper, a robust method of ground plane detection using GDM algorithm is proposed. The results demonstrate the effective depth map-base approach of ground plane detection with lower complexity. By a comparison with RANSAC và Enhanced V-Disparity algorithms, the average of recognition rate for ground plane detection always higher than the compared methods in most cases. The proposed approach's *R1* is greater than the compared methods 2%, while the *R2* of the proposed approach is smaller than half of the compared method's *R2*. The experienced results are also consistent with the actual environment certainly. This work could be used for autonomous vehicle driving in off-road environment in the future.

Next work will focus on removal of non-ground areas which are possibly appeared due to the camera sensors by utilization of combined smoothing windows at different sizes.

Acknowledgment

References

[1] Sunglok Choi et al; "Robust ground plane detection from 3D point clouds"; KINTEX, 14th International Conference on Control, Automation and Systems (ICCAS 2014); Gyeonggi-do, Korea; 2014; 1076-1081.

[2] Anders Hast, Johan Nysjö, Andrea Marchetti; "Optimal RANSAC – towards a repeatable algorithm for Finding the optimal set", Journal of WSCG 21, (2013), 21–30.

[3] Xiao Hu, Rodriguez, Gepperth; "A multi-modal system for road detection and segmentation"; 2014 IEEE Intelligent Vehicles Symposium Proceedings; Michigan, USA; 2014; 1365–1370.

[4] Atsushi Sakai, Yuya Tamura, Yoji Kuroda; "Visual odometry using feature point and ground plane for urban environment"; Robotics (ISR), 41st International Symposium on and 2010 6th German Conference on Robotics (ROBOTIK); Munich, Germany; 2010; 1- 8.

[5] Tarsha-Kurdi, F., Landes, T., & Grussenmeyer, P. ; "Hough-transform and extended ransac algorithms for automatic detection of 3d building roof planes from lidar data"; Proceedings of the ISPRS Workshop on Laser Scanning, Vol. 36; Espoo, Finland; 2007; 407-412.

[6] Borrmann, D., Elseberg, J., Lingemann, K., N¨uchter, A.,"The 3D Hough Transform for plane detection in point clouds: A review and a new accumulator design", Journal 3D Research 2(2), (2011), 1-13.

[7] Ogundana, O. O., Coggrave, C. R., Burguete, R. L., Huntley, J. M.,"Automated detection of planes in 3-D point clouds using fast Hough transforms", Optical Engineering 50(5), (2011), 53609–053609.

[8] Zhongli Wang, Jie Zhao; "Optical flow based plane detection for mobile robot navigation"; In Proceedings of the 8th World Congress on Intelligent Control and Automation; Taiwan; 2011; 1156-1160.

[9] Arshad Jamal et al; "Real-time ground plane segmentation and obstacle detection for mobile robot navigation"; 2010 International Conference on Emerging Trends in Robotics and Communication Technologies (INTERACT 2010); Chennai, India; 2010; 314–317.

[10] J. Arro´spide L. Salgado M. Nieto R. Mohedano, "Homography-based ground plane detection using a single on-board camera", IET Intell. Transp. Syst. 4(2), (2010), 149–160.

[11] N. Mostof, M.Elhabiby, N. El-Sheimy; "Indoor localization and mapping using camera and inertial measurement unit (IMU)"; 2014 IEEE/ION Position, Location and Navigation Symposium - PLANS 2014; Monterey, CA, USA; 2014; 1329 – 1335.

[12] Prabhakar Mishra et al; "Monocular vision based real-time exploration in autonomous rovers"; 2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI); Mysore, India; 2013; 42 – 46.

[13] K. Gong and R. Green; "Ground-plane detection using stereo depth values for wheelchair guidance"; In 24th International Conference Image and Vision Computing New Zealand (IVCNZ); Wellington, New Zealand; 2009; 97-101.

[14] Dai Yiruo, Wang Wenjia, and Kawamata Yukihiro; "Complex ground plane detection based on V-disparity map in off-road environment"; IEEE Intelligent Vehicles Symposium (IV); Gold Coast, Queensland, Australia; 2013; 1137 – 1142.

[15] CheeWay Teoh, ChingSeong Tan, Yong Chai Tan; "Ground plane detection for autonomous vehicle in rainforest terrain"; IEEE Conference on Sustainable Utilization and Development in Engineering and Technology (STUDENT); Wellington, New Zealand; 2010; 7–12.

[16] Nguyen Tien Dzung et al; "Gradient depth map based ground plane detection for mobile robot applications"; 8th Asian Conference on Intelligent Information and Database Systems - ACIIDS 2016, Part I, LNAI 9621; Da Nang, Vietnam; 2016; 721–730.

[17] Wikimedia Foundation, Inc. https://en.wikipedia.org/wiki/Kinect.

[18] Middlebury College, Microsoft Research, and the National Science Foundation, http://vision.middlebury.edu/stereo/data/scenes2005/

[19] Middlebury College, Microsoft Research, and the National Science Foundation, http://vision.middlebury.edu/stereo/data/scenes2006/

25