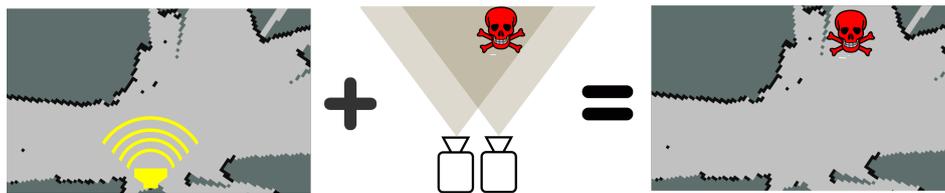


LIDAR and stereo camera data fusion in mobile robot mapping

Jana Vyroubalová*



Abstract

LIDAR (2D) has been widely used for mapping and navigation in mobile robotics. However, its usage is limited to simple environments. This problem can be solved by adding more sensors and processing these data together. This paper explores a method how measurements from a stereo camera and LIDAR are fused to dynamical mapping. An occupancy grid map from LIDAR data is used as prerequisite and extended by a 2D grid map from stereo camera. This approach is based on the ground plane estimation in disparity map acquired from the stereo vision. For the ground plane detection, RANSAC and Least Squares methods are used. After obstacles determination, 2D occupancy map is generated. The output of this method is 2D map as a fusion of complementary maps from LIDAR and camera. Experimental results obtained from Willow Garage Stereo and Hokuyo UTM-30LX Laser are good enough to determine that this method is a benefit, although my implementation is still a prototype. In this paper, we present the applied methods, analyze the results and discuss the modifications and possible extensions to get better results.

Keywords: LIDAR — stereo camera — mapping — least squares plane — sensors fusion — disparity map — ground plane detection

Supplementary Material: N/A

*xvyrou04@stud.fit.vutbr.cz, Faculty of Information Technology, Brno University of Technology

1. Introduction

Environment mapping is one of the most important part of mobile robotics. It is the basis for localization and navigation - especially in real time in unknown environment. There are plenty of research about mapping by 2D LIDAR and stereo camera separately (not only because of their frequency of use). LIDAR provides very accurate information and its frequency is high. However, this sensor provides information only in one plane. For example, a LIDAR is able to see the legs of a table on a right position but not the table surface which presents as obstacle to the robot. On the other hand, a stereo camera can provide 3D structural

information with color and texture of any object. However, the camera is less accurate compared to a laser range detector and the range of view is thus limited.

The problem we address in this paper is primarily which data from LIDAR and stereo camera should be fused and next how to fuse the data to get any new or better information for mapping than just from the data of one sensor only. Furthermore, we explore which methods we should use to achieve the fusion with focus on stereo camera (shortly stereo) data processing. The objective of this work is to find out if our method can be useful. We do not provide the best possible results of this approach. Of course we can achieve better

results with a superior camera and with more suitable data.

There are some presented techniques which could solve some parts of our problem. For example, there is a 3D object detection technique presented in [1]. However, this approach is performed by using three cameras and they do not focus on environment mapping, but only on the objects detection. In these works [2, 3] there are obstacles and road detected. Their method is based on "v-disparity". However, this method will fail in environment where obstacles occupy the most part of the image (mostly indoor) or in outdoor environment where, in the presence of a slope variation, the 2D line representing the ground plane (road) is not a straight line. There is also paper [2], which prefers transformation of disparity map to 3D point cloud and then uses depth information to extract ground plane and obstacles before raw extraction from disparity map [3, 4]. However, working with 3D point cloud is not reliable due to the nonlinear transformation involved in range estimation, where the inaccurate parameters of camera are used.

Our method is based on ground plane detection in disparity map from stereo camera. Then we determine obstacles in environment and project this scene into 3D point cloud, which is used for 2D map construction. Finally, we complement the 2D map from LIDAR by this 2D map from stereo. Basic information and problems about both sensors used in mapping are presented in Section 2. Details about our approach in stereo mapping based on the ground plane detection are described in Section 3. In Section 4, we present the idea of data fusion and finally we show and discuss the results of our approach in Section 5. All the conclusions are summarized in Section 6.

2. Sensors in mapping

A LIDAR and a camera are the most frequently used sensors in mobile robotics, although a laser range finder is relatively expensive. Conversion from data to information is on high level.

2.1 LIDAR

A 2D LIDAR is non-contact optical sensor, which uses laser beams to scan an environment in two-dimensional area. It scans from small angles to full 360 degrees and a scanning frequency can be from 5 Hz to 50 Hz for example of the most used LIDAR sensors.

For many years, the most used approach to get a map for localization and navigation from LIDAR is *Occupancy grid mapping*. This method has been investigated for long time and therefore we use it as a

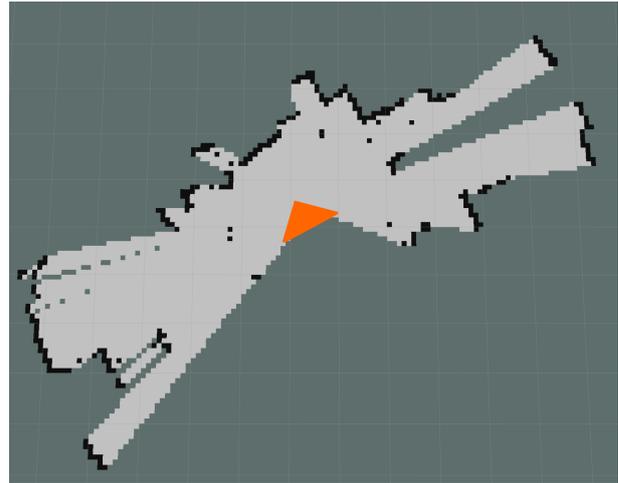


Figure 1. Occupancy grid map with robot pose (the orange triangle).

prerequisite for our method. For better idea, see Figure 1. The map is composition of small cells - you can see the cells as rasterization or pixelization. Value of each cell represents probability of whether the cell is occupied by an obstacle or not (in grayscale - black means 100% occupied, white means free). The green cells represent unknown environment. As we can see, in the area in front of the robot in Figure 1 there are many objects and on the back side, the area is not investigated. For details about implementation of this method, see [5, 6, 7].

The main disadvantage of LIDAR mapping is that the sensor sees only in two dimensions, so some obstacles may not be detected by it. However, this sensor is very accurate, therefore this is the basis for localization and navigation. Due to these advantages, we try to extend this method by information from stereo camera.

2.2 Stereo camera

A stereo camera provides important information about scene: color, texture, intensity and mainly 3D structure. Moreover, the camera is able to see behind the obstacles. However, the narrow field of view and limited range are disadvantages.

If we want to work with stereo camera properly, we need to calibrate the camera to get the camera intrinsic and extrinsic parameters to repair the images (so our images are no longer deformed). Based on the parameters from camera calibration and epipolar geometry, we are able to *rectify* the stereo images and find the pixels correspondences [8].

Finally, we can compute the *binocular disparity* between two points, which we need for information about depth. This principle is based on difference between the indexes of corresponding pixels (see Figure



Figure 2. Pixel correspondences in left and right rectified images.

2):

$$d = (x_l - x_{cl}) - (x_r - x_{cr}), \quad (1)$$

where $[x_l, y_l]$ and $[x_r, y_r]$ are the corresponding points (endpoints of each line in Figure 2) in left and right images. Coordinate y is the same in pair and x_{cl} and x_{cr} are indexes of the center in left and right images.

The 3D scene is based on disparity map, from which the scene is backward reconstructed. For this backward projection into camera frame, we need the parameters from camera calibration:

$$z = \frac{fB}{d}, \quad (2)$$

$$x = \frac{(u - u_c)z}{f}, \quad (3)$$

$$y = \frac{(v - v_c)z}{f}, \quad (4)$$

where f is the focal length, B is the baseline, $[u, v]$ is the computed pixel position (index) in disparity map and $[u_c, v_c]$ is the position of pixel in the center of image. For more details see [8, 9].

We can see that there are many steps, where the stereo camera cumulates software accuracy error:

1. estimation of calibration parameters,
2. images rectification,
3. finding the pixel correspondences for disparity algorithm,
4. backward projection from disparity map to 3D coordinates.

Notice that from the principle of disparity computation (Figure 2 and equation 1), there can be a problem with large homogeneous regions.

3. Stereo mapping

We work with ground robots, so we consider the robot motion in two dimensions. Although the stereo camera provides 3D information, we transform it into 2D grid map similar to the LIDAR map in Section 2.1. This

map complements the LIDAR map (both in 2D), which is suitable approach for localization and navigation for ground robots.

In our approach, we determine obstacles and free regions from the disparity map directly - notice that the backward projection from 2D perspective to 3D orthogonal system is on the end of chain. Scheme of our method is depicted in Figure 3:

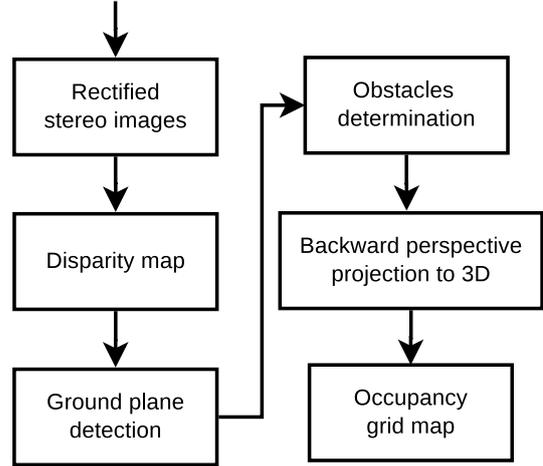


Figure 3. Scheme of creating the map from stereo camera.

3.1 Disparity map

Disparity map allows us to work with raw data from camera sensor, obviating all needless transformations based on some estimated parameters. We recommend some preprocessing such as smoothing images by bilateral filter. Then pure disparity computation is not enough, because we lose information about exact edge and the objects are incomplete. We can solve this problem by post-filtering of raw disparity map on base of *Weighted Least Squares* method [10].



Figure 4. Source image for filtered disparity.

Our result of disparity map is shown in Figure 5 which was computed from source Figure 4. Some information can be better to seen in Figure 6 or Figure 7. Disparity of the ground plane looks relatively fine, although we notice the fact that this ground plane is homogeneous area. There are some papers [11], which



Figure 5. Example of our filtered disparity map.

explore disparity map in homogeneous areas. We can also upgrade ground plane disparity map by image segmentation or hypothesis about ground plane such as flatness and parallelism. However, it is not the aim of this paper.

3.2 Ground plane detection

We use a geometric model of the plane to find the ground plane in front of the robot. We could simply assume that the ground plane is of constant height. It can be useful for indoor applications, where there is no change in position of the ground plane relative to the camera. However, the dynamic approach (which we use) works in more cases. The geometric definition of plane is:

$$ax + by + cz = d, \quad (5)$$

where $n = (a, b, c)$ is normal vector. Distance R_i of point i (coordinates of i are $[x_i, y_i, z_i]$) from the plane (5) is

$$R_i = ax_i + by_i + cz_i - d, \quad (6)$$

assuming that the plane is normalized, so

$$a^2 + b^2 + c^2 = 1. \quad (7)$$

To get concrete definition of our plane, we need for example 3 points, which fit the plane. For this purpose, *Random Sample Consensus* (RANSAC) [12] is an effective technique developed from within the computer vision community. Basically we iteratively and randomly generate 3 points in region of interest, then fit a plane and compute the ratio of inliers and outliers. Ground plane mostly covers bottom half of an image, so we can use only this part as region of interest for better efficiency.

Then we do *Least Squares Plane* (LSQP) fitting to find the best solution. We need to minimize Q :

$$Q = \sum_{n=1}^N R_i^2, \quad (8)$$

where N is the set of inliers from RANSAC. Using (6) we can rewrite (8) as:

$$Q = \sum_{n=1}^N (ax_i + by_i + cz_i - d)^2. \quad (9)$$

Then from (9):

$$\frac{\partial Q}{\partial a} = \sum_{n=1}^N 2x_i(ax_i + by_i + cz_i - d) = 0, \quad (10)$$

and similarly the equations $\frac{\partial Q}{\partial b} = \frac{\partial Q}{\partial c} = \frac{\partial Q}{\partial d} = 0$. To avoid the trivial solution $a = b = c = 0$, we define the condition (7). With this condition the solution of these equations becomes *Eigenvalue problem*. You can find more information about this LSQP approach in [13, 14, 15].

Our result of the ground plane detection is shown in Figure 6 and Figure 7. The green points (the biggest area) are the disparity map in 3D space. The three red points define the plane. The black plane which fits the red points is our ground plane. We can also see the tilt of black plane, which copies the decreasing disparity values to the center of the image.

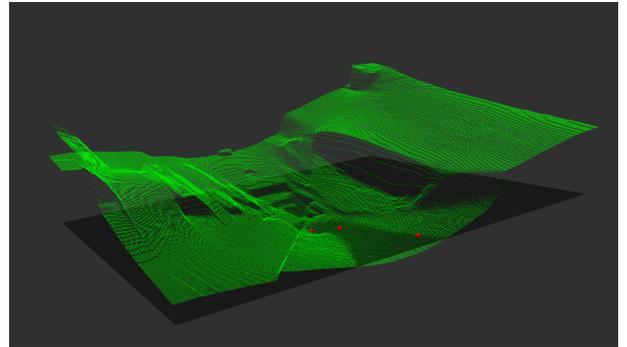


Figure 6. Disparity map visualized in 3D space with detected ground plane.

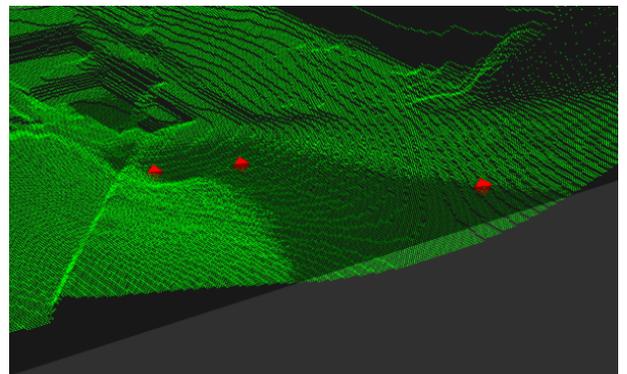


Figure 7. Detail of ground plane from Figure 6.

3.3 Obstacles determination

For determination between obstacles and free areas for robot movement, we use simple decision rule:

$$p_d(x,y) = \begin{cases} free, \Delta d < low \\ occupied, low \leq \Delta d \leq high \\ free, \Delta d > high \end{cases} \quad (11)$$

where $p_d(x,y)$ is pixel in disparity map, Δd is difference between estimated disparity (from ground plane) and measured disparity and low and $high$ are two thresholds which are obtained by our experiments. Although the lower part of an obstacle can be resolved as free, it does not affect our mapping, because the lower part is covered by upper part in grid map (see Section 3.4).

You can see our result in Figure 8. The green area is free and the red is occupied. Notice that the estimator determined the area at the bottom of the walls as free. The reason is that we have to set the $high$ parameter high to detect whole ground plane because of inaccuracy in disparity map.

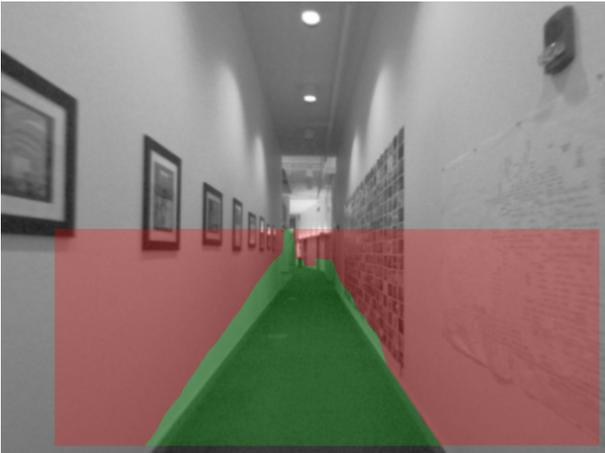


Figure 8. Result of determination between free and occupied area.

3.4 Occupancy grid map

To represent the result in occupancy grid map, we need to transform our information from perspective (disparity map) to orthogonal 3D. For this purpose we use the equations mentioned in Section 2. When we have orthogonal coordinates of points, we can project these into the 2D space. This is the step, where the wrong determined bottom parts of obstacles are covered properly by the top parts.

We do not discuss the occupancy probabilities of cells in map, because it is not the aim of this paper. We simplify our approach only to 2 values in grid - occupied and unknown (see Section 5), because this is enough for our research. Moreover, omission of free cells from stereo camera is justified by its inaccuracy.

For complete implementation of correct disparity map see Section 2.1 or directly [5].

4. Data fusion

The aim of our fusion is to bring new information to mapping, which is for example used in path planning. Because of the great accuracy of LIDAR (see 2.1), the localization should be computed mainly from LIDAR map. To include the camera into localization, the priority of this sensor should be set to right value. On the other hand, the path planner definitely benefits from having information from our approach.

We do not include the information about obstacles from camera to localization (for this purpose, we can use something like camera SLAM [16]). However, we recommend to use this method to upgrade path planner - it should know about obstacles, which are invisible for LIDAR, although the camera is a little bit inaccurate. Advanced implementation of the path planning, such as *Vector Field Histogram* [17] based on a local cost map, can be used.

We implemented only simple fusion of our two maps from LIDAR (Section 2.1) and stereo camera (Section 3), which is suitable for our approach. The occupied cells from camera are mapped into the LIDAR occupancy grid map. We use something like *OR* operation:

$$M_i = \begin{cases} free, (L_i = free) \wedge (C_i = unknown) \\ unknown, (L_i = unknown) \wedge (C_i = unknown) \\ occupied, (L_i = occupied) \vee (C_i = occupied) \end{cases} \quad (12)$$

where M_i is cell in the fused map M on index i , L_i is cell in the LIDAR map on index i and finally C_i is cell in the stereo camera map on index i . This map is then given to path planner.

5. Implementation and experiments

We implemented our method in C++ with *Robotic Operation System* (ROS) framework [18]. Especially for LIDAR mapping, we used GMapping module from ROS. To process some graphical calculations, the *OpenCV* library [19] was used. Complexity of our implementation is

$$O(n^2), \quad (13)$$

where n is input image resolution. The most critical section is computation of disparity map. However, some optimizations for finding disparity exist, but it is not the topic of this work.

Experimental results were obtained from MIT Stata Center Data Set [20], where they have Hokuyo UTM-30LX laser range finder and Willow Garage Stereo

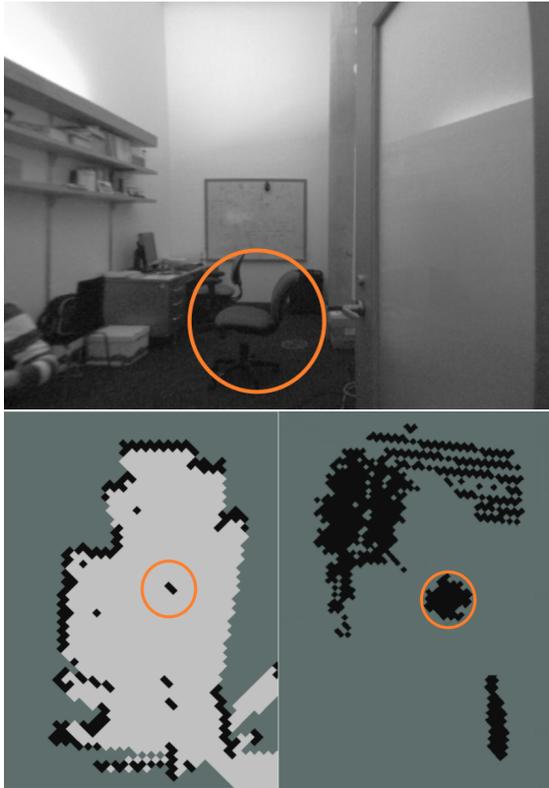


Figure 9. Chair scene - source image on the top, LIDAR map at the bottom left, camera map at the bottom right.

camera. Partial results (of each significant step) are shown during this paper in appropriate section. Our final results are shown in next several figures, where are illustrated different situations. In every example we show a source image and next we show corresponding occupancy maps from LIDAR sensor (on the left side) and stereo camera sensor (on the right).

As it shows in Figure 9, the laser scanner is not able to distinguish the chair (in the orange ellipse in source image), because its height is more than the plane scanned by the LIDAR. LIDAR is able to see only the chair leg, which is much smaller than the real obstacle for mobile robot. The camera detected this chair, although stereo vision data is noisy and inaccurate in the range measurement, especially over large distance.

The appropriate example of the problem with large distance detection by stereo camera is shown in the Figure 10. There we can see narrow and straight passage. Length of this passage is more than 25 meters. The LIDAR is able to see whole range without any problem. However, the stereo camera is lost in this range - it sees very inaccurately (the orange ellipse). The reason is evident from Figure 5, where disparity map is wrong in rear part of the passage.

Figure 11 shows situation, where are many chairs with thin and shiny legs around the table. LIDAR



Figure 10. Passage scene - source image on the top, LIDAR map at the bottom left, camera map at the bottom right.



Figure 11. Table scene - source image on the top, LIDAR map at the bottom left, camera map at the bottom right.

is only able to see table legs. At the opposite side, the camera sees all the chairs with table (the orange ellipse). However, we can see failure of the camera detection in the red ellipse. The reason of this failure is similar to Figure 7. Disparity map of this ground plane is cambered (because it is large homogeneous area), therefore the ground plane detection is a little bit inaccurate (especially at wrong tilt). So there is a possibility to detect part of the ground plane as obstacle.

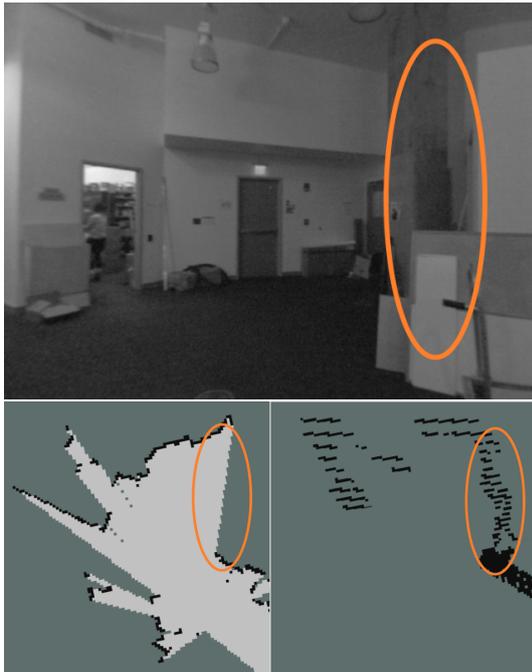


Figure 12. Wall scene - source image on the top, LIDAR map at the bottom left, camera map at the bottom right.

The last shown experiment is in Figure 12, which is great illustration of invisible obstacle for LIDAR, but stereo camera knows about this obstacle. Vision of LIDAR is overshadowed.

Based on all these experiments, we can see many facts about our approach. Camera vision is very noisy. However, with the LIDAR background, it can be helpful in many situations. The problem with inaccurate disparity map occurs very often, which can lead to the situation in Figure 11. We can eliminate it by improving the disparity map extraction, which is not simple topic for homogeneous areas. We have already opened this theme in Section 3. Moreover, data from camera in large distance is so sparse. The reason is that the perspective projection shrinks real information into small part of camera image - so there are only few pixels with information in large distances, which are reprojected back into 3D orthogonal system.

6. Conclusions

In this paper we have presented the method for LIDAR and stereo camera data fusion to reach better mapping and navigating. Particularly, we focused on new information extraction from stereo camera compared to LIDAR mapping. This approach includes techniques such as Least Square (Plane) method, Random Sample Consensus and extraction of disparity map from raw stereo images.

We evaluated every step of this method and we find out, that there are big reserves in disparity map detection. However as our tests shows, we reached what we wanted to - we shew that this approach is useful and it brings new information to mapping of mobile robots. Also we discussed the modifications of this method leading to better results.

One possible way to further improve the accuracy of the presented method is to investigate the side of disparity map extraction, because we did not focus on it so much. We only used some implemented method. Next there are some topics to continue this method. For example, how to assign weights of detected object from stereo camera to navigation or localization compared to the LIDAR. And next one is how to adapt path planner to fused map [17, 21].

Acknowledgements

I would like to thank my colleague Radim Luža for many discussions and suggestions.

References

- [1] L. Romero, A. Núñez, S. Bravo, and L.E. Gamboa. *Fusing a Laser Range Finder and a Stereo Vision System to Detect Obstacles in 3D*. Advances in Artificial Intelligence – IBERAMIA 2004, 2004. ISBN: 978-3-540-30498-2, pages 555-561.
- [2] L. Matthies and P. Grandjean. *Stochastic Performance Modeling and Evaluation of Obstacle Detectability with Imaging Range Sensors*. IEEE Transactions on Robotics and Automation, 1994. ISSN: 1042-296X, pages 783-792.
- [3] R. Labayrade, D. Aubert, and J.-P. Tarel. *Real Time Obstacle Detection in Stereovision on Non Flat Road Geometry Through "V-disparity" Representation*. IEEE Intelligent Vehicle Symposium, 2002. Pages 646-651.
- [4] F. Ferrari, E. Grosso, G. Sandini, and M. Mastrassi. *A Stereo Vision System for Real Time Obstacle Avoidance in Unknown Environment*.

- IEEE International Workshop on Intelligent Robots and Systems, Towards a New Frontier of Applications, 1990. Pages 703-708.
- [5] T.-D. Vu, O. Aycard, and N. Appenrodt. *Online Localization and Mapping with Moving Object Tracking in Dynamic Outdoor Environments*. IEEE Intelligent Vehicles Symposium, 2007. ISBN: 1-4244-1067-3, pages 190-195.
- [6] G. Grisetti, C. Stachniss, and W. Burgard. *Improved Techniques for Grid Mapping with Rao-Blackwellized Particle Filters*. IEEE Transactions on Robotics, 2007. Volume 23, pages 34-46.
- [7] G. Grisetti, C. Stachniss, and W. Burgard. *Improving Grid-based SLAM with Rao-Blackwellized Particle Filters by Adaptive Proposals and Selective Resampling*. In Proc. of the IEEE International Conference on Robotics and Automation (ICRA), 2005. ISBN: 0-7803-8914-X, pages 2432-2437.
- [8] M. S. H. Achmad, W. S. Findari, N. Q. Ann, D. Pebrianti, and M. R. Daud. *Stereo camera — Based 3D object reconstruction utilizing Semi-Global Matching Algorithm*. 2016 2nd International Conference on Science and Technology-Computer (ICST), 2016. ISBN: 978-1-5090-4357-6, pages 194-199.
- [9] L. Yan, X. Zhao, and H. Du. *Research on 3D measuring based binocular vision*. 2014 IEEE International Conference on Control Science and Systems, 2014. ISBN: 978-1-4799-6397-3, pages 18-22.
- [10] D. Min, S. Choi, J. Lu, B. Ham, K. Sohn, and M. N. Do. *Fast Global Image Smoothing Based on Weighted Least Squares*. IEEE Transactions on Image Processing, 2014. ISSN: 1941-0042, pages 5638-5653.
- [11] J.-E. Kim, K.-D. Kim, and K.-H. Jung. *Reliable estimation of disparity map in textureless region of roadway*. 2017 19th International Conference on Advanced Communication Technology (ICACT), 2017. ISBN: 978-89-968650-9-4, pages 399-402.
- [12] M.A. Fischler and R.C. Bolles. *Random consensus: A paradigm for model fitting with applications to image analysis and automated cartography*. Communications of the ACM, 1981. Volume 24, pages 381-395.
- [13] K. Pearson. *On lines and planes of closest fit to systems of points in space*. Philosophical Magazine, 1901. Volume 2, pages 559-572.
- [14] V. Schomaker, J. Waser, R.E. Marsh, and G. Bergman. *To Fit a Plane or a Line to a Set of Points by Least Squares*. Acta Cryst, 1959. Volume 12, pages 600-604.
- [15] D.M. Blow. *To fit a plane to a set of points by least squares*. Acta Cryst, 1960. Volume 13, page 168.
- [16] A.J. Davison, I.D. Reid, N.D. Molton, and O. Stasse. *MonoSLAM: Real-Time Single Camera SLAM*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007. ISSN: 0162-8828, pages 1052-1067.
- [17] J. Borenstein and Y. Koren. *The vector field histogram-fast obstacle avoidance for mobile robots*. IEEE Transactions on Robotics and Automation, 2007. ISSN: 1042-296X, pages 278-288.
- [18] Group of authors. *ROS: Robot Operating System*. Cit. [2016-12-23] URL <http://www.ros.org/>.
- [19] G. Bradski and A. Kaehler. *Learning OpenCV: Computer Vision in C++ with the OpenCV Library*. O'Reilly Media, 2013. 2nd edition.
- [20] M. Fallon, H. Johannsson, M. Kaess, and J. Leonard. *The MIT Stata Center Dataset*. IJRR Dataset Paper. Cit. [2017-04-06], under review, URL: <http://projects.csail.mit.edu/stata/index.php>.
- [21] I. Ulrich and J. Borenstein. *VFH*: local obstacle avoidance with look-ahead verification*. IEEE International Conference on Robotics and Automation, 2000. Volume 3, pages 2505-2511.