Reliable Multi-baseline Stereovision filter for Navigation in Unknown Indoor Environments

Khalid Al-Mutib, Muhammad Emaduddin, Mansour Alsulaiman, Hedjar Ramdane
Dept. of Computer Engineering
College of Computer Science and Information
King Saud University, Riyadh, Saudi Arabia
{muteb, memaduddin, msuliman, hedjar}@ksu.edu.sa

Ebrahim Mattar
Electrical & Electronics Engineering Dept.
College of Engineering, University of Bahrain
Kingdom of Bahrain
ebrgallaf@eng.uob.bh

Abstract

A multi-baseline stereo-vision based approach is proposed to produce reliable obstacle information for real-time mapping and path-planning within an unknown indoor environment. The proposed method uses Sum of Absolute Differences (SAD) based stereo-matching to generate multiple point-clouds using a multi-baseline stereovision camera. The presented approach focuses on generating low-noise, high-confidence 3D point clouds that serve as a reliable input to ICP-SLAM, mapping and path-planning modules of a full-scale navigation system. The resultant mobile robot navigation system performs superior when compared to a system fed with typical stereovision based 3D point clouds utilizing state-of-the-art noise-removal filters. In our work, we feed low-noise, high confidence point clouds to a pre-implemented ICP-SLAM module and the resultant 3D point cloud map is projected onto a 2D stochastic occupancy grid-map. The proposed mapping process is unique and is optimized for minimizing shortcomings and false positives (e.g. featureless surfaces, specular surfaces) involved in a typical stereo-matching process. Authors have shown through a statistically significant number of experiments that the resultant stochastic occupancy grid-map can be reliably used for path-planning within an unknown, dynamic indoor environment.

Keywords- robotics; navigation; Stereo-vision; SLAM; point cloud filtering; multi baseline

1. INTRODUCTION

Through past two decades, special interest is paid to unmanned guided vehicle navigation using laser and vision based sensors. Recently, a renewed focus has emerged towards vision based navigation of mobile robots within indoor environments in order to perform complex tasks such as goods handing at factory floors, transporting office and healthcare materials within offices and hospitals, cleaning/wiping floors for general household etc. Almost any kind of indoor scenario usually exhibits a high amount of unstructured clutter leading to challenges in the perception of obstacles and navigable space. Our work in this article focuses on solving problems that arise from shortcomings in local stereo matching algorithms such as handling of *specular reflections*, *lack of discriminative image features* and *repetitive patterns* [1] within indoor environment. Most local stereo matching methods strive to remain within the realm of real-time algorithms while remaining accurate e.g. AdaptGCP [2], ADCensus [3] and SAD-IGMCT [4], to name a few. Instead of presenting completely new local stereo matching technique, we propose a novel higher-level filtering method that aims at minimizing the noise and patchy 3D information generated by local stereo matching techniques such as SAD. This higher-level filtering approach makes the proposed method suitable for filtering point clouds generated by any local stereo matching technique. In order to prove the viability and robustness of our proposed filter, we have integrated our filter into a fully functional mobile robot navigation system. A multi-baseline stereovision camera by Point Grey is used as the primary input sensor for our proposed method. We chose this camera due to its relatively high accuracy and performance profile [5]. The stereo camera gathers point cloud using 640x480 resolution images @7.5FPS (frames per second) in both wide (24cm) and narrow (12cm) baseline configuration. The camera initialization in dual baseline mode and limited support for stereo-camera in 64bit Windows 7, are two implementation related factors that restrict the camera fps to down to 7.5. For the same reason we perform obstacle avoidance for obstacles at very short distances using a custom designed Laser scanner based Fuzzy-logic Motion Controller [17]. It must be mentioned here that even state-of-the-art filters such as surface validation filter [6] and back-forth validation filter [7] are unable to remove false positives due to specular surface reflections. Also these filters are unable to preserve 3D points belonging to featureless patches of stereo images while trying to minimize false positives. In our proposed work, three distinct 3D point clouds are generated using the two combinations (center-left, center-right) of narrow baseline and one combination (right-left) of wide baseline lenses. The multi-baseline camera is shown in Figure 2 for reference. The observations for these point clouds are taken near simultaneously. The method then chooses ROI (Region of Interest) within each of the point cloud in order to choose the most accurate region within the point clouds. Each of the ROI is then projected onto a separate stochastic occupancy grid-map using a floor and obstacle detection method by Emaduddin et.al [8]. The updated cells lying within sensor observation cone, within each stochastic occupancy grid-map are then consolidated and mapped onto a single occupancy grid-map based on an

intelligent criterion. The consolidated map is then passed through a custom designed filter that removes false positives from the occupancy grid-map using the occlusion and sensor cone visibility information from multiple observations that are separated temporally. The resultant grid-map is free, to a very high degree, from false positives generated by *specular reflections* and *repetitive patterns*. The final grid-map also stores, for each cell, the 3D location of a centroid calculated from all the points that were projected onto that particular

Figure 1 PowerBot by Adept Mobilerobots capturing stereo observations during a navigation task

cell. This extra information is used to reconstruct a down-sampled, low-noise, highly reliable 3D point cloud, ideal for use in a landmark based SLAM algorithm. We use this reliable point cloud in a pre-implemented generic version of ICP-SLAM (Iterative Closest Point – Simultaneous Localization and Mapping) algorithm. The resultant localization proved to be far more accurate in the conducted experiments rendering accurate maps for navigation. Stereovision camera is mounted via a Pan-Tilt Unit by FLIR Systems, Inc(see Figure 3), on-board PowerBot – a mobile robotics development platform by Adept Mobilerobots Inc (see Figure 1)

After outlining the previous work conducted in the domain of multi-baseline stereovision based navigation in section 2, section 3 illustrates the inner details of our proposed method by describing (i) Stereo Capture & ROI extraction (ii) Multi grid-map projection and consolidation (iii) Filtering via visibility checks (iv) 3D point cloud reconstruction from consolidated map (v) Navigation using the consolidated map. Section 4 details the results of conducted experiments.

2. PREVIOUS WORK

The first practical use of occupancy grids can be attributed to Elfes [9]. Among the earliest uses of trinocular camera along with area tessellated into grid-cells, the work of Murray et al. [10] can be considered as reference. The proposed system in [10] has many short-comings including sensitivity to false positives when environment observation time is less (due to slower convergence rate while updating occupancy probability). In [1], an interesting occupancy grid-mapping technique is presented that uses both dense and sparse point clouds from Stereovision and SAM respectively (structure and Motion). In this case, although the algorithm appears to handle false positives relatively well but there exists a great margin for encountering outliers as no multi-view multi-baseline stereo matching technique is utilized. Several other techniques were found in literature that attempt to utilize multi-baseline stereo at the level of local stereo matching or multi-view stereo in order to filter noise e.g. [11-14] but no technique was found to be using a higher-level multi-baseline filtering or using it in combination with multi-view stereo filtering.

3. PROPOSED METHOD

The method is implemented as a system having client-server architecture. A multi-threaded thin client application lives on-board the mobile robot while another multi-threaded much heavier server-side application resides on the server. Client-side consists of sensor data capture, motion control, ROI extraction, point cloud compression and network communications modules. Server-side includes Network communications, point cloud decompression, Multi grid-map projection & consolidation, 3D point cloud reconstruction and Navigation modules. Client-Server architecture is deployed in order to distribute the computation load of point cloud and grid-map manipulation to high-end server side processors and GPUs. While optimum navigation speed, real-time obstacle avoidance is not the focus of our work in this article, above architecture is put into place for the same reasons.

Figure 2 Bumblebee XB3 multi-baseline stereovision camera by Point Grey Research Inc.

Figure 3 Pan-Tilt Unit D-46 by FLIR Systems, Inc.

3.1 Point-Cloud Capture & ROI Extraction:

The client application that executes onboard the robot, constantly captures the 3D point clouds with three distinct baselines, mentioned in section 1, employing the stereo camera. None of the state-of-art filters listed in section 1 are applied to the stereo rectification process. After point cloud

generation, ROIs are extracted from these point cloud following the presented limits.

(i) P': ROI for narrow baseline (left & center camera)
z-axis: 0.1 m till 5.0 m

(ii) P'': ROI for narrow baseline (right & center camera)
z-axis: 0.1 m till 5.0 m

(iii) P''': ROI for wide baseline (left & right camera)
z-axis: 2.5 m till 5.0 m

The system then compresses the data and transmits the three ROI to the server. Point-Clouds are compressed using Octree-based point-cloud compression technique [15], in order to minimize the observation transmission time and keep the method viable for real-time processing. It takes on an average, over 150ms for the data to be captured and transmitted over the network. The Bumblebee camera-pose $Pose_{BB}$ and the robot pose $Pose_{Rob}$ are associated with each point cloud observation. This enables the method to reliably deduce localization information from ICP-SLAM module.

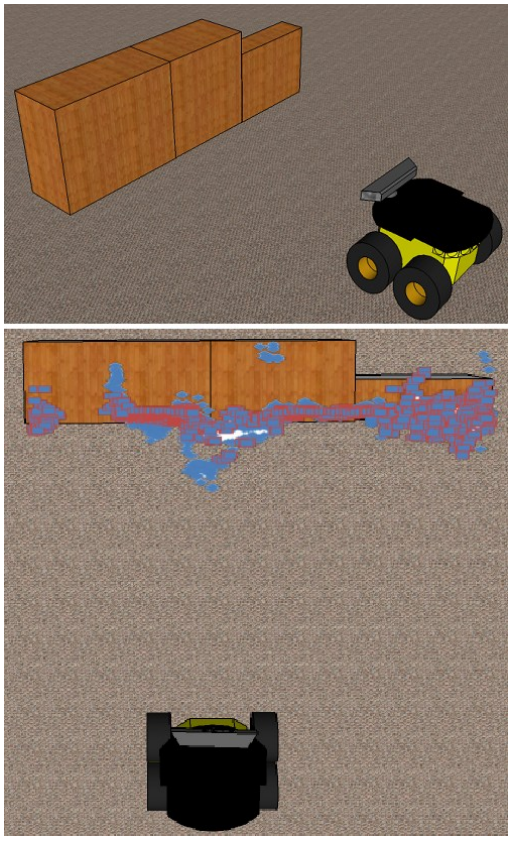

Figure 4 Top: A view of an obstacle configuration. Bottom: Top view of actual 3D stereo clouds from both baselines (Blue shows narrow, Red shows wide and the over-lapping area has high confidence.

3.2 Multi-grid Map Projection and Consolidation

The point clouds received from the client application are fed into a well-tested Hough-transform based plane fitting technique detailed in [8]. This technique is employed to distinguish floor from the obstacles. Thus, for three point clouds (P', P'', P'''), after the execution of plane fitting technique, three stochastic 2D occupancy grid-maps (G', G'', G'''), will be populated respectively. In order to perform multi-baseline consolidation at this each vertex $U_{x,y}$ belonging to (G', G'', G'''), is loaded with further information apart from a usual probability value i.e.

$$U_{x,y} = \begin{cases} prob_{u_{xy}} \\ centroid\_xyz_{u_{xy}} \\ update\ status_{u_{xy}} \\ confidence_{u_{xy}} \end{cases} \dots \text{(I)}$$

where

$$prob_{u_{xy}\ at\ time\ t} = \log\left(\frac{p(U_{x,y}=occupied|p_1,\dots,p_t)}{1-p(U_{x,y}=occupied|p_1,\dots,p_t)}\right) \dots \text{(II)}$$

Equation II implements a binary Bayes Filter. Here $prob_{u_{xy}\ at\ time\ t}$ is represented in log odds form [16] for time t. $p(U_{x,y} = occupied|p_1, \dots, p_t)$ represents the occupancy probability given the measurement p (p qualifies as a point that belongs to an obstacle). Details regarding the points that qualify as an obstacle are given in [8]. In Equation I, $update\ status$ indicates that the particular vertex of grid-map received an update from the current observation (currently received point cloud from the sensor). $centroid\_xyz_{u_{xy}}$ contains the centroid location of all 3D points contributing to the cell's occupancy probability. The $confidence$ value can have three possible states i.e. *low*, *medium* and *high*.

The consolidation process goes as follows:

1. At the beginning of consolidation process set $confidence_{u_{xy}}$ for all vertices to *low*
2. For all the vertices $U_{x,y}$ in (G', G'', G'''), which receive update in form of a single or multiple points belonging to an obstacle, the value of $confidence_{u_{xy}}$ is set to *medium*.
3. For all the vertices $U_{x,y}$ in $G' or\ G''$, whose distance from robot center is less than 1.5 meter and which also receive update in form of a single or multiple points belonging to an obstacle, the value of $confidence_{xy}$ is set to *high*. This is done as wide baseline based stereo output produces very inaccurate 3D points for initial 1.5 meters ahead of camera for the current configuration of stereo parameters that have be initialized for the our system.
4. Now for three vertices *a, b* and *c* where
$a \in G', b \in G'', c \in G'''$
Do the following

$if\ (prob_a = medium)\ AND\ (prob_b = medium) AND\ (prob_c = medium)$
$\{\ prob_a = high;$
$prob_b = high;$
$prob_c = high;\ \}$

5. Add vertices $U_{x,y}$ to the final consolidated grid-map G for which $prob_{u_{xy}}$ is *high*

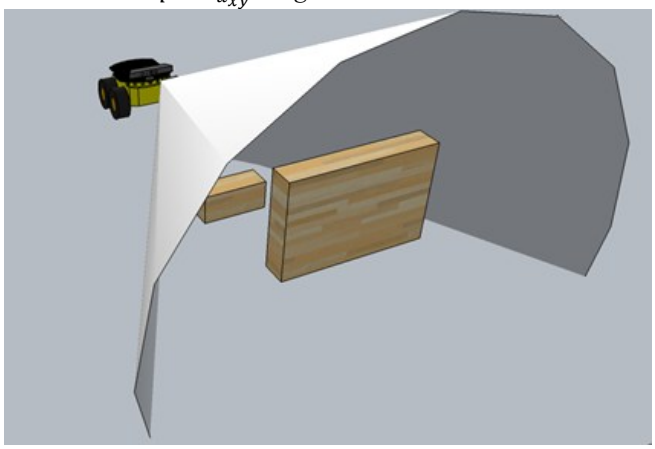

Figure 5 The 66 degree 3D visibility cone for the Bumblebee stereo vision camera

3.3 Filtering via visibility checks

In this step, a set of view-rays is computed in 2D for the field-of-view (FOV) of the camera up to a predefined distance i.e. 5 m (see Figure 5). Each one of the view-ray is traversed while beginning at the current camera position. Each traversed cell probability $prob_{u_{xy}}$ in grid-map G is set to a value of 1.0 until the first obstacle cell is reached. The remainder of the view-ray is skipped and the remaining cells lying on the view ray remain unchanged. False positives in this case mostly occur due to specular reflections or surfaces that do not follow Lambertian reflectance model. The distance of the first obstacle cell along with height of the $centroid\_xyz_{u_{xy}}$ is stored in an indexed array $closest\_obstacle[359][2]$. The index runs from 0 to 359 degree, each index representing the view ray after every 1 degree. After this, all cells lying on the view ray beyond the first obstacle cell are assessed for height, in case any cell is found to be with the less than the height of the closest obstacle, the cell's probability value is increased (cleared) via a Bayesian update. The false positives in this case occur mostly due to reflections of surroundings on translucent or ultra-smooth surfaces. In case of a sensor having a 2D cone, the height check will be unnecessary and the cell will be cleared instantly.

3.4 3D Point Cloud Reconstruction from Consolidated Map

After the consolidated grid-map G is filtered and contains minimal false positives, now the process of reconstruction of down-sampled 3D point cloud can begin. The process is fairly simple as it consists of three short steps.

1. Initialize a new point cloud P'.
2. Start iterating the filtered consolidated point cloud G. Whenever an occupied cell is encountered, a new point is pushed into P' with $centroid\_xyz_{u_{xy}}$ values of associated with the cell.
3. When all occupied cells in G are exhausted, stop.

3.5 Navigation using the Consolidated Map

The focus of work in this article is the enhancement in the reliability of the stereovision input. For the purpose of evaluating the improvement in the localization, mapping and path-planning, reconstructed point cloud P' is submitted to pre-implemented ICP-SLAM algorithm, robot pose along with camera pose is also submitted to the ICP-SLAM for each observation. The actual accumulation of 3D point clouds in a 3D global map is done by ICP-SLAM. The resultant 3D global map constantly updates a 2D global stochastic grid-map. This global grid-map is utilized by path-planning module for the purpose of path generation. A motion-planning module executing on-board the mobile robot awaits new path-points whenever path planning and re-planning is performed.

4. EXPERIMENTATION AND RESULTS

A brief comparison of 2D global grid-maps filtered by Triclops SDK (Point Grey Research Inc.) with the global grid-maps filtered by the proposed method, is presented in Figure 7 for reference. As already mentioned, we gather stereo point cloud sequences through the stereo-vision camera mounted on top of a mobile research platform. These sequences are gathered while the robot is in motion and robot/camera pose is bundled with each observation. The observations are processed online by the proposed method and in real-time. The maximum speed the robot achieves during navigation is 0.47 m/s. For comparison, the ground truth for the arena is also shown in Figure 6.

Figure 6 Ground Truth for the occupancy grid-map

5. DISCUSSION

Figures 8 & 9 detail the path-planning and path-execution performance for the generated map via our proposed filter. It should be reminded here that whenever

filters are relaxed within low-feature environments to allow maximum number of feature points belonging to low-featured obstacles, noise creeps in. In Figure 9, specifically, plain walls with vertical dark-colored stripes were part of the experiment environment. Vertical stripes were just 0.05m wide and repeated after every 2 meters on the wall. With such low number of features, extracting nearly 70% of wall points was considered to be a success for our method. The robot path is evidently optimized and can be considered optimal for unknown low-featured environments. Robots travelled with an average speed of 0.45 m/s during the experiments. Intelligent camera gazing was also part fot eh experiment implementation. This module helped focus the "gaze" of camera via a DPPTU towards the low-featured obstacles such as plain walls.

6. CONCLUSION

In this paper, a reliable multi-baseline stereovision filter for enhanced navigation in unknown indoor environments is proposed. Although the filter acts as small unit within a large navigation system, nevertheless its role is primary and crucial for successful and error-free navigation. Resultant grid-maps generated with the help of the filtered reconstructed point clouds by ICP-SLAM module show marked reduction. Noise present in stereo 3D point clouds causes a great deal of inaccuracy in localization and mapping output. Usually feature points based key points are used to improve localization output but this approach, in no way improves the inaccuracies and false positives in the mapping output. Thus the presented filter successfully reduces the noise within stereo point clouds rendering their effective usage in SLAM algorithms.

6. ACKNOWLEDGEMENT

This work is supported by NPST program by King Saud University (Project No. : 08-ELE-300-02).

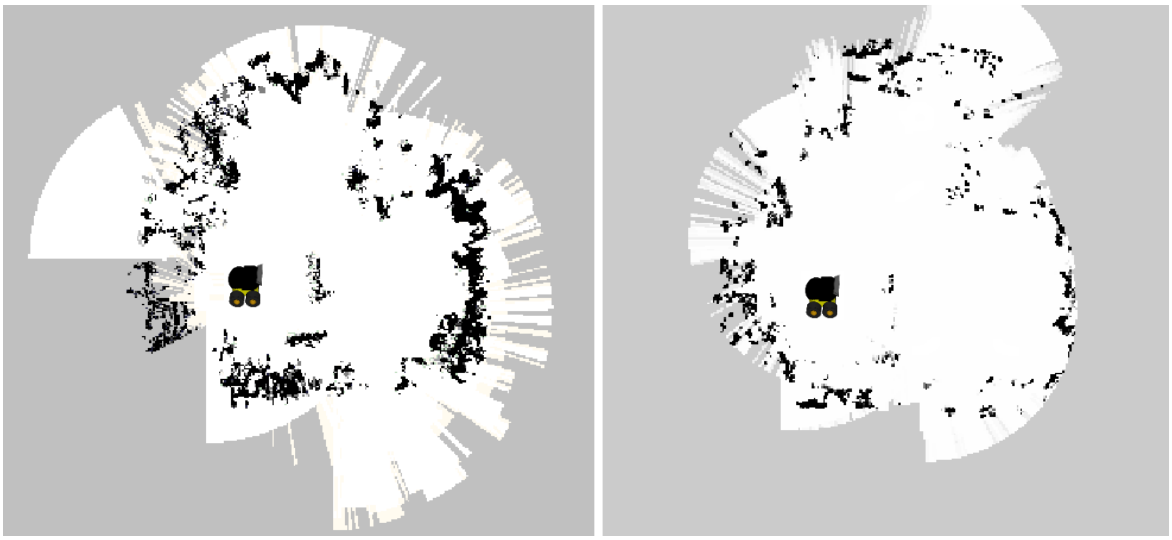

Figure 7 Stochastic Occupancy Grid-maps. LEFT: Mapping output from ICP-SLAM after application of proposed filter. The map contains minimal noise without the loss of fidelity. RIGHT: Mapping out from ICP-SLAM after application of Surface-size, Texture validation and Back-forth filter by Triclops SDK, Point Grey Research. The map looses fidelity at the cost of noise removal.

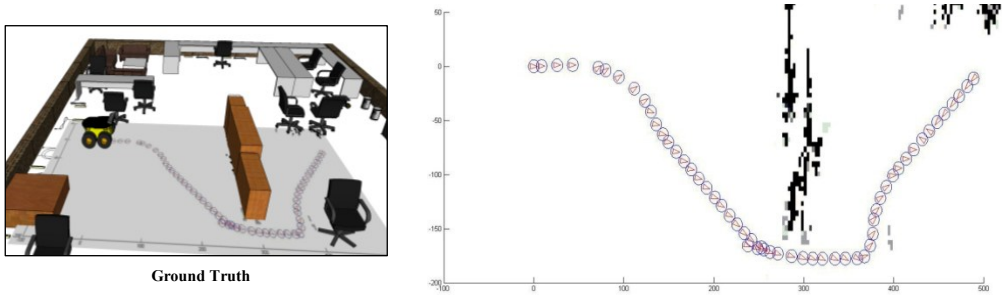

Figure 8 Mapping output from ICP-SLAM used for path-planning and traversal within a room. Red arrows indicate camera gaze angle. Blue circle represents actual robot path. X-Y axis coordinates are shown in centimeters.

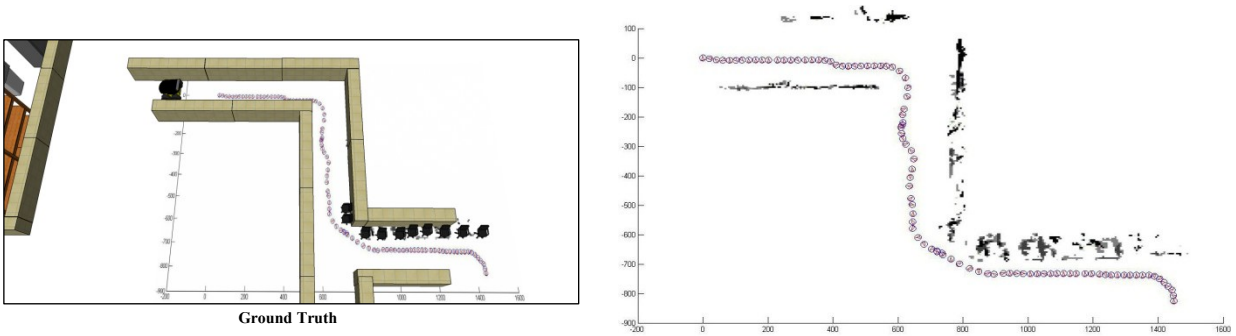

Figure 9 Mapping output from ICP-SLAM used for path-planning and traversal in a corridor. Red arrows indicate camera gaze angle. Blue circle represents actual robot path. X-Y axis coordinates are shown in centimeters.

7. REFERENCES

[1] H. Lategahn, W. Derendarz, T. Graf, B. Kitt, and J. Effertz, "Occupancy grid computation from dense stereo and sparse structure and motion points for automotive applications," in Intelligent Vehicles Symposium (IV), 2010 IEEE, 2010, pp. 819–824.

[2] C. Shi, G. Wang, X. Pei, H. Bei, and X. Lin., "High-accuracy stereo matching based on adaptive ground control points," Submitted to IEEE TIP 2012.

[3] Xing Mei; Xun Sun; Mingcai Zhou; shaohui Jiao; Haitao Wang; Xiaopeng Zhang, "On building an accurate stereo matching system on graphics hardware," *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on* , vol., no., pp.467,474, 6-13 Nov. 2011 doi: 10.1109/ICCVW.2011.6130280

[4] K. Ambrosch and W. Kubinger., "Accurate hardware-based stereo vision." in Computer Vision Image Understanding, vol. 114, no. 11, pp. 1303-1316. November 2010.

[5] Point Grey Research, Inc., "Stereo accuracy and error modeling," Product Support: Knowledge Base. 30-Aug-2012.

[6] D. Murray and J. Little, "Using Real-Time Stereo Vision for Mobile Robot Navigation," Autonomous Robots, vol. 8, no. 2, pp. 161–171, Apr. 2000.

[7] P. Fua., "Combining stereo and monocular information to compute dense depth maps that preserve depth discontinuities." in Proceedings of the 12th International Joint Conference on Artificial intelligence, Vol. 2. Morgan Kaufmann Publishers Inc., CA, USA, 1292-1298. 1991.

[8] M. Emaduddin, K. Al-Mutib, M. AlSulaiman, H. Ramdane, and E. Mattar, "Accurate floor detection and segmentation for indoor navigation using RGB+D and stereo cameras," in Conference on Image Processing, Computer Vision, and Pattern Recognition, Las Vegas, Nevada, USA, 2012.

[9] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," Computer, vol. 22, no. 6, pp. 46–57, 1989.

[10] D. Murray and C. Jennings, "Stereo vision based mapping and navigation for mobile robots," in IEEE International Conference on Robotics and Automation, vol. 2, pp. 1694–1699. 1997.

[11] H. Wang, J. Xu, J. Guzman, R. Jarvis, T. Goh, and C. Chan, "Real Time Obstacle Detection for AGV Navigation Using Multi-baseline Stereo," in Experimental Robotics VII, vol. 271, D. Rus and S. Singh, Eds. Springer Berlin Heidelberg, 2001, pp. 561–568.

[12] A. Kuhn, H. Hirschmüller, and H. Mayer, "Multi-Resolution Range Data Fusion for Multi-View Stereo Reconstruction," in Pattern Recognition, vol. 8142, pp. 41–50. 2013.

[13] A. Milella, G. Reina, and M. M. Foglia, "A multi-baseline stereo system for scene segmentation in natural environments," in Technologies for Practical Robot Applications (TePRA), 2013 IEEE International Conference on, 2013, pp. 1–6.

[14] D. Gallup, J.-M. Frahm, P. Mordohai, and M. Pollefeys, "Variable baseline/resolution stereo," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1–8.

[15] R. Schnabel and R. Klein, "Octree-based point-cloud compression," in *Eurographics Symposium on Point-Based Graphics*, 2006, pp. 111–120.

[16] S. Thrun, W. Burgard, and D. Fox, "Probabilistic Robotics," 3rd Ed., Intelligent robotics and Autonomous Agents. Cambridge, Mass, MIT Press, 2006.

[17] M. Faisal, R. Hedjar, M. AlSulaiman and K. Al-Mutib , "Fuzzy Logic Navigation and Obstacle Avoidance of Mobile Robot in Unknown Dynamic Environment," International Journal of Advanced Robotic Systems, vol. 10, 2013.