Improving Path Planning and Mapping Based on Stereo Vision and Lidar

Peyman Moghadam, Wijerupage Sardha Wijesoma, Dong Jun Feng
School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore, 639798

Abstract—2D laser range finders have been widely used in mobile robot navigation .However, their use is limited to simple environments containing objects of regular geometry and shapes. Stereo vision, instead, provides 3D structural data of complex objects. In this paper, measurements from a stereo vision camera system and a 2D laser range finder are fused to dynamically plan and navigate a mobile robot in cluttered and complex environments. A robust estimator is used to detect obstacles and ground plane in 3D world model in front of the robot based on disparity information from stereo vision system. Based on this 3D world model, 2D cost map is generated .A separate 2D cost map is also generated by 2D laser range finder. Then we use a grid-based occupancy map approach to fuse the complementary information provided by the 2D laser range finder and stereo vision system. Since the two sensors may detect different parts of an object, two different fusion strategies are addressed here. The final occupancy grid map is simultaneously used for obstacle avoidance and path planning .Experimental results obtained form a Point Grey's Bumblebee stereo camera and a SICK LD-OEM laser range finder mounted on a Packbot robot are provided to demonstrate the effectiveness of the proposed lidar and stereo vision fusion strategy for mobile robot navigation.

Keywords—mobile robot,sensor fusion,stereo vision,path planning

I. INTRODUCTION

There are a plethora of effective techniques reported [1, 2, 3] for path planning and mapping using a single planar laser range finder. While the laser range finder is generally accurate and fast in the range measurements over small to long ranges, it only provides information about objects in the plane of the scanning laser beam. Obstacles above or below the scanning plane are invisible to the 2D laser range finder. For example, the range finder may only be able to detect the presence of the legs of a table but not the table surface that presents as obstacle to the robot. It is thus more suitable for sensing simple environments consisting of simple and regularly shaped objects. On the other hand, color stereo vision can provide not just 3D structural data but also color and texture information. However, as compared to a scanning laser range finder the field of view and the range are very limited, albeit the relatively slow speed. Further, the fidelity of the range measurements provided by stereo vision is susceptible to changes in lightning and lack of texture in the environment.

The fusion of multiple sensor data is an obvious attempt to overcome deficiencies of the different sensing modalities in

building accurate and reliable maps [4].Sensor fusion has been studied in many robotics research topics such as, mapping, localization, path-planning ,exploration and SLAM. In this paper, the main goals of the sensor fusion are map building and path-planning. Much work has been reported in the area of fusing lidar and stereo vision for reliable and robust detection of obstacles [6, 7, 8].In [5] sonar and infrared data are used to complement stereo occupancy grid mapping using a one-dimensional linear Kalman filter. In [9] omnidirectional stereo camera data and laser range finder data are used to generate a free space of the environment. A separate probabilistic grid map is maintained for each sensor and each of these grid maps are subsequently combined using logical rules to generate a final free space map. The map data is integrated over time based on the robot's ego-motion estimates.

In this paper, a robust algorithm is used to construct the 3D world model from disparity information in front of the robot. Based on this 3D world model, 2D cost map is generated .A separate 2D cost map is also generated by 2D laser range finder. An effective strategy is addressed by fusing this narrow field of view, short range stereoscopic vision cost map (5-8m) with a 360 degree long range 2D range scanning lidar. The proposed technique is fast and robust for real-time implementation in complex and cluttered indoor and outdoor environments.

The rest of the paper is organized as follows: Part II describes the major characteristics of the sensors and their probabilistic models. Part III details the probabilistic occupancy grid mapping based on ground plane obstacle detection using stereo vision. Part IV describes the grid based probabilistic sensor fusion algorithm that combines 2D lidar cost map with stereo vision cost map for dynamic path-planning. Part V provides experimental results obtained in an indoor environment using a Bumblebee stereo camera and SICK LD-OEM lidar sensor mounted on an iRobot's Packbot. Part VI concludes the paper with a discussion on our future plans for improving the methodology.

II. RANGE SENSORS

A. Stereo Vision System

The stereo vision system provides not just 3D structural information about the environment but also intensity, color and texture. However, it may suffer from narrow field of view and limited range. There are some situations where stereo vision

cannot provide promising range information such as in large homogenous regions like plain walls or in poor lightning conditions. The vision sensor used in our experiments is a short baseline Point Grey Bumblebee color stereo vision camera that computes disparity using Sum of Absolute Differences correlation method (SAD).In some cases, such as occlusion and lack of texture disparity estimated values may contain many mismatches. In order to avoid such inaccuracies and errors, several validation steps such as Texture Validation and Surface Validation are applied [10].

The 3D position (in the camera coordinate frame) of a feature with disparity d is given by:

$$z = \frac{fB}{d} \qquad (1)$$

$$x = \frac{u - u_c}{f} z \qquad (2)$$

$$y = \frac{v - v_c}{f} z \qquad (3)$$

Where, f (pixel) is the focal length, B (m) is the baseline of the stereo camera, and (u, v) is the pixel position of the feature point from the center of the image (u_c, v_c) on the 2D disparity image.

B.   Probabilistic Stereo Model

Mathies and Grandjean [11] have shown that the disparity errors of a stereo camera are Gaussian distributed with a constant standard deviation. However, as would be expected it is better to assume a varying disparity standard deviation correlated with the estimated disparity rather than a constant standard deviation. For faraway features the standard deviation is smaller since the expected disparity is small and it increases when features come closer. Thus, based on the manufacture's specifications of the camera the distribution of disparity errors is modeled as Gaussian with varying standard deviation started for close range by 0.4 .The 3D range measurement computed from disparity has a more complicated distribution as it is shown in the nonlinear triangulation equation (1) .However, it is quite common to use a Gaussian approximation to probabilistically model the range estimates obtained by stereoscopic vision [11]. Therefore, in our implementation, a Gaussian sensor noise model is used for range estimates obtained form stereo vision, with the range standard deviation (σ_z) proportional to the disparity standard deviation (σ_d) and the range (z^2), i.e.

$$\sigma_z \propto \sigma_d z^2 \qquad (4)$$

The uncertainties for x, y, z are:

$$\Delta z = \frac{fB}{d - 0.2} - \frac{fB}{d} \qquad (5)$$

$$\Delta x = \frac{p\Delta z}{f} \qquad (6)$$

$$\Delta y = \frac{p\Delta z}{f} \qquad (7)$$

Where p, known as the pointing accuracy and is limited by the accuracy of in-house calibration. This value is indicative the accuracy of the 3D ray associated with each pixel in the image. At the native CCD resolution (1024*768) the RMS value of p was measured to be 0.4 pixels. This value should be scaled by the stereo resolution/native resolution (320/1024).This decrease in accuracy of range estimates derived from stereo vision with increasing range (z) for the Point Grey Bumblebee camera is shown in Fig 1.A sick laser was used to measure the nominal distance between the robot and the object .
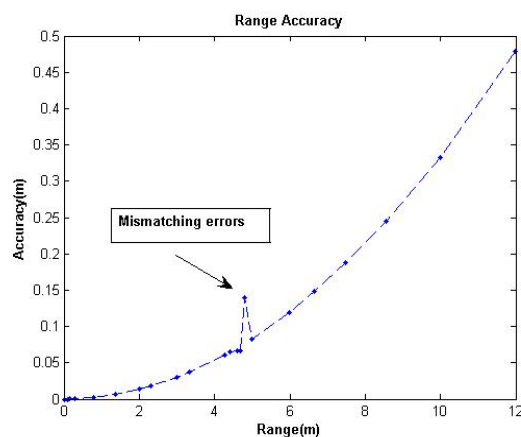

Figure 1.   Accuracy in range vs. nominal distance

Another important attribute in stereoscopic sensor modeling is the range resolution. That is the smallest change in range Δz ,for given a change in disparity Δd ,and is determined by:

$$\Delta z = \frac{z^2}{Bf} \Delta d \qquad (8)$$

It is easy to see that the range resolution is a function of the square of the range itself. At near distances to the camera, the resolution is much better than distances farther away.

C.   Laser Range finder

The laser range finder used is a SICK LD-OEM 1000 and is an electronic, non-contact optical sensor that uses laser beams to scan the contours of its environment on a single level. It measures the environment in two-dimensional polar coordinates. When a measuring beam strikes an object, the position is determined with regard to the distance and direction. It scans the full 360 degrees and has a scanning frequency that can be varied from 5 Hz to 15 Hz. We have found that the LD-OEM 1000 is very efficient and accurate in range measurements and can be used effectively in implementing algorithms for Simultaneous Localization and Mapping (SLAM), navigation and obstacle avoidance. For ease of

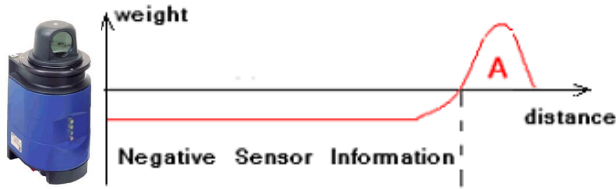implementation and simplicity, we model the range measurements obtained by the LD-OEM as Gaussian.

Figure 2.   Sensor modeling of LD-OEM laser range finder

III.   OCCUPANCY GRID MAP

Occupancy grid mapping is a classical approach that addresses the problem of integrating consistent maps of sensor data with uncertainly over time [12]. Occupancy grid map has been used in the past for laser range finder [13, 14] and stereo vision [15]. Each grid cell in an occupancy map can have one of three possible values: Occupied, free, and unknown as shown in Fig 3. If d is the observed distance to the nearest object, the region with a standard deviation measured in (4) is marked as occupied ,the region between sensor and the object is labeled as free and the cells beyond are noted as unknown .

To observe the probability of three possible values for a grid cell, an obstacle detection algorithm is applied. Obstacle detection on laser range finder point clouds is straightforward. When the laser beams intersect with an object it returns a set of points corresponding to the distance and direction of the object. Now using the Gaussian distributed measurement model is applied as described above to label the grid cells in the range (up to 20m) and field of view (360 degree) of the laser. The dual-lens camera provides complementary 3D information in a narrow field of view (66 degree) over a relatively short range (5 – 8m) in the environment. These observations provide measurements of objects of various heights and positions that permit the classification of the terrain into 'traversable' and 'nontraversable' regions taking into consideration the dimensions of the robot. We collapse the 3D structural and obstacle information provided by the stereoscopic vision into a 2D cost map for fusion with the laser generated 2D cost map.
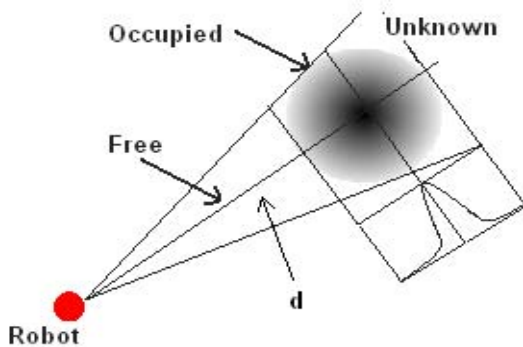

Figure 3.   Grid attributes

A.   Mapping based on Stereo Vision

We apply obstacle and ground plane detection on stereo data to find the probability of occupancy of a grid cell. There are two approaches to ground plane and obstacle detection using stereo vision. Those that find the objects and ground plane directly from the disparity map [17, 18] and those that first transform the disparity to 3D point clouds and then use depth information to extract ground plane and obstacles [11]. In [11] they assume an obstacle as a near-vertical step displacement, so their algorithm checks for such obstacle in the range image by using a pair of pixels in the same column. It measures a displacement in height and range between two pixels in the same column number in the range image. If the measured change exceeds a given pre-defined threshold, there is an obstacle present; however, working on 3D point clouds is not reliable due to the non-linear transformations involved in range estimation as described in section II.

Disparity map allows us to work with data produced directly from the imaging sensor, obviating the need for intrinsic camera parameters whilst improving robustness. A simple and real time method known as "v-disparity" is due to Labayrade et al. [11].The v-disparity image is computed by first generating a histogram of disparity values on every row of the disparity image. Since it is reasonable to assume that a majority of the points result from the ground plane, the histogram bin with maximum number of pixels is assumed to correspond to the ground plane. The road surface is estimated as a 2D straight line in v-disparity image using the Hough transform instead of 3D plane estimation in the disparity image. In some cases such as indoor environments where obstacles occupy the most part of the image or outdoor environment where in the presence of a slope variation the 2D line representing the ground plane is not a straight line, the v-disparity fails to extract a ground plane. In contrast ,the strategy ,which is used in this paper ,does not need dominant plane assumption in front of the robot to detect the ground plane and it works in environments 50% occupied by obstacles.

Model based ground plane detection algorithms use a geometric model of the plane to find the ground plane in front of the robot and to detect obstacles. To model a ground plane there are two different approaches. One is to simply assume that the plane in front of the robot is of constant height. Any entities whose height is above this ground plane are marked obstacles. Similarly, the dynamic approach models the ground plane based on the disparity map in image coordinates (u, v, d).In this case, ground plane equation takes the linear form of

$$au + bv + c - d = 0 , \qquad (9)$$

where (a, b, c) are the ground plane parameters. The dynamic model is implemented in this paper.

In [18], they used this linear approach by predefining ground floor disparity and fast comparison technique between current disparity value and a reference disparity map to detect obstacles. One drawback of this method is it doesn't adapt with changing environment in such case which robot moves from flat surface to the slop area. Another problem is it needs a free space in front of the robot each time for initialization and finding the predefined ground floor disparity values.

For indoor applications moving over flat ground, there is no change in position of the ground plane relative to the camera, and the cyclopean ground-plane disparity function is therefore fixed. However, in unstructured environments where there is no dominant flat ground, a robust estimator is required that not only tolerates a large outlier percentage but also tolerates several discontinuities. RANSAC (Random Sample Consensus) [19] is an effective technique developed from within the computer vision community to estimate the ground plane robustly, that take all the image features in the region of interest, fits the ground plane features and discards the obstacle features as outliers. Points that are above the ground plane but lower than the robot's height are labeled as obstacles. By interpreting this 3D map, the system estimates the local ground plane, as well as the location of slopes, ditches, and above ground objects. This yields computing free space near robot. The 3D map, with identified obstacles is then projected down to a 2D map with traversal costs assigned to the identified obstacles. This method does not require any information about velocity and motion of the vehicle. It is expected that once the obstacle distribution near the vehicle is known, a path-planning algorithm will generate a suitable path to the goal for the vehicle to follow. The robust estimator method comprises of two stages:

1. Randomly select three feature points (subset) from the region of interest (ROI) of disparity image to fit a plane. Then compute its residuals with respect to all data points in the ROI and count the points with distances less than some threshold (inliers).The distance of a point P_a from ROI of the disparity image to the plane is

$$\text{Minimum distance} = (P_a - P_b)dotn / \|n\|, \quad (10)$$

n is the normal of the plane and $P_b = (u_b, v_b, d_b)$ is a point on the plane where

$$au_b + bv_b - 1d_b + c = 0, \quad (11)$$

2. Until N iterations repeat step 1.The triple points with the maximal number of inliers are chosen. Then do total least-squares fitting for the plane parameters using its identified inliers by orthogonal regression.

Ground plane fitting by RANSAC uses the supporting points instead of the whole area of interest prevents from failures by selecting obstacles instead of true ground as model parameters [20]. It automatically discards the outliers from data points that finally used for model fitting. In RANSAC, adaptively choosing the number of good samples (N) with probability p can be obtained by:

$$N = \frac{\log(1-p)}{\log(1-(1-e)^s)}, \quad (12)$$

where e is the contamination fraction and is often unknown a priori ,so we pick the worst case where the 50% of the ROI is occupied with obstacles , s is the size of the sample . The size

of the sample in this case is 3, as we need 3 points to define a plane, and the probability p is set to 0.99.The pseudo-code for finding the number of sample is described in Algorithm I.

ALGORITHM I. THE PSEUDO-CODE FOR FINDING THE NUMBER OF SAMPLE

- N=∞, sample_count =0
- While N >sample_count repeat
  1. Choose a sample and count the number of inliers
  2. Set e=1-(number of inliers)/(total number of points)
  3. Recompute N from e
  4. Increment the sample_count by 1
- Terminate

Fig. 4 and Fig. 6 show the result for an outdoor and indoor environment with objects at different distances. The dynamic estimator finds the ground plane (shown in green) and discards outliers (shown in red). Fig. 5(a) shows the height map based on the static disparity assumption. The height map is very sensitive to the robot pitch and yaw, and hence even small vibration of the robot results in marking free space detected as obstacles. The two ellipses in Fig. 5(a) shows the free space area detected wrongly as obstacles. The result of the 2D map of the dynamic ground-fitting algorithm is shown in Fig. 5(b) and Fig.6 (b). The result is less prone to vibrations and different terrain conditions, especially in outdoor environments where there are no planar surface in many instances.
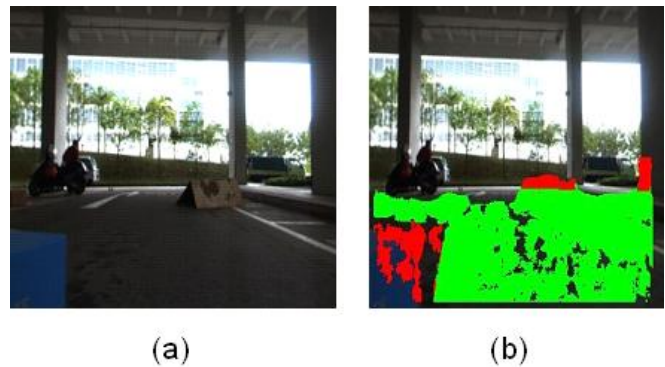

(a)   (b)

Figure 4.   Dynamic ground plane finding in urban environment

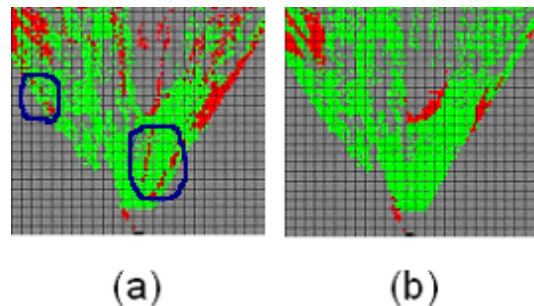

(a)   (b)

Figure 5.   (a) Height map (Z=Const.) (b) Dynamic Cost map

The benefits of using stereo vision to complement laser data are clear from Fig. 4. The blue box seen on the left of Fig. 4(a),
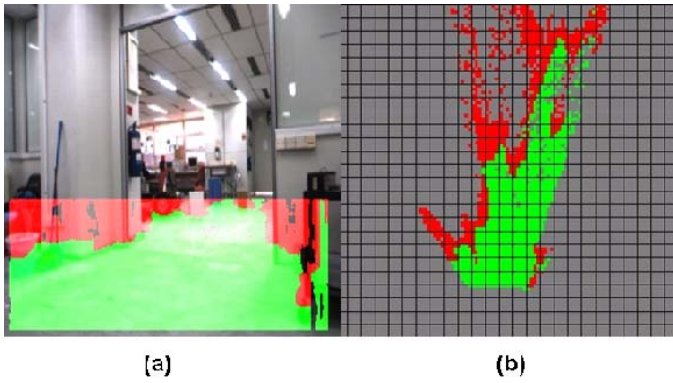
Figure 6.   (a) Ground plane in indoor (b) Dynamic cost map

blocks the view of the laser beyond the blue box. Thus, it is not possible to know the terrain condition beyond the box purely based on laser data. However, as shown in Fig. 4(b), stereo vision is capable of seeing beyond the box thereby complementing the laser. The fusion of laser and vision for cost mapping helps effective dynamic path panning.

The information provided by the ground-plane and obstacle detection algorithm is used to generate the vision cost grid map as follows. The weight of the ith grid cell (cost value) from stereo vision measurement at (x, y) is computed using

$$w_i(x, y) = \begin{cases} 0, \Delta d < low \\ a, low \le \Delta d < high \\ 0, \Delta d > high \end{cases} \quad (13)$$

Δd is the difference between estimated disparity and the real value ,and low and high are two thresholds which are obtained by our experiments. Only points whose measured disparities are outside of the expected ground plane will be considered as obstacles with a traversability cost of *'a'*. Therefore, lower part of obstacles may not be detected as its disparity differences may be not significant enough. However, it may not affect our mapping since the lower part of an object will be covered by its upper part in the grid map. For a series of N measurements the final weights of the grid cell is obtained by summing over all measurements, i.e.

$$w(x, y) = \sum_{i=1}^{N} w_i(x, y) \quad (14)$$

IV.   SENSOR FUSION FOR INDOOR PATH PLANNING

At the beginning, all cells are typically unknown and are initialized to a nominal value. When the robot moves an accurate cost map of the environment can be constructed by fusing the laser and vision data by integrating over time. This fused map dynamically can be used to plan paths.

Local path planning algorithms such as the vector field histogram (VFH) [21] can be implemented based on a local cost map. We construct the fused cost map by taking the weights of the individual cost maps. If the traversability weight at grid cell position (x, y) is p for the laser and q for vision then

we use the maximum of the two to obtain the fused weight W, at grid cell position (x, y), i.e.

$$W(x, y) = \max[p(x, y), q(x, y)] \quad (15)$$

This ensure that the traversable environment space is indeed traversable and minimizing the probability of any failure to the crashing on the obstacles. Safety factor is the first priority in our concern.

Now this fused cost map is used for local path planning and obstacles avoidance using a VFH algorithm. The advantages of grid map based sensor fusion with VFH algorithm are three-fold. First, the grid map can be used to handle the uncertainty of each sensor modeling. Secondly, the VFH algorithm itself is implementing with an internal occupancy grid map. In such a case, the mapped environment can be directly transformed into VFH, minimizing the unnecessary computational cost. Thirdly, the occupancy grid map obtained from the sensor fusion can be implemented with higher-level path planning algorithms.

In our experiment, the A* path planning algorithm is fused with VFH algorithm similar to the VFH*[22], implementing with the same environmental mapping. The approach makes use of the occupancy grid map obtained into both global and local path planning algorithms.

V.   EXPREMENTAL RESULTS

Experimental results were obtained from a Bumblebee camera and LD-OEM laser range finder mounted on a Packbot robot. Fig. 7 shows images depicting the test cases. Fig. 6(a) shows the real images obtained from the camera in an indoor environment. Different objects are placed in front of the robot in the overlapping region of the fields of view of laser and the stereo camera. As it shown in Fig. 7(c) laser scanner is not able to distinguish the obstacles because their heights are less than the plane scanned by the laser.

In Fig. 7(b), the occupancy grid map provided by stereo camera shows that 3D information derived from camera can detect all obstacles that were not detected by the laser. However, as viewed in Fig. 7(b), stereo vision data is often noisy and inaccurate in the range measurement, especially over large distances. This drawback restricts the usage of the stereo vision camera for mobile robot navigation. The laser range finder, on the other hand, mitigates this weakness.

Fig. 8 shows the planned path. White dot represents the position of the robot and the red dot is the goal. Black space denotes the obstacles detected. As shown in Fig. 8(b), the table is not detected by laser and the path is planned through the table. The A*star algorithm is used in conjunction with the VFH on the fused occupancy grid map, providing a smooth and reliable path as shown in the Fig 8(c).

VI.   CONCLUSION AND FUTHER WORKS

In this paper, we have presented a path planning and mapping approach based on fusing laser range finder and stereo vision into an occupancy grid map. We investigated the need of sensor fusion for mapping unknown space.

Through our experimental results presented here, we have clearly shown the limitation of using either laser or vision
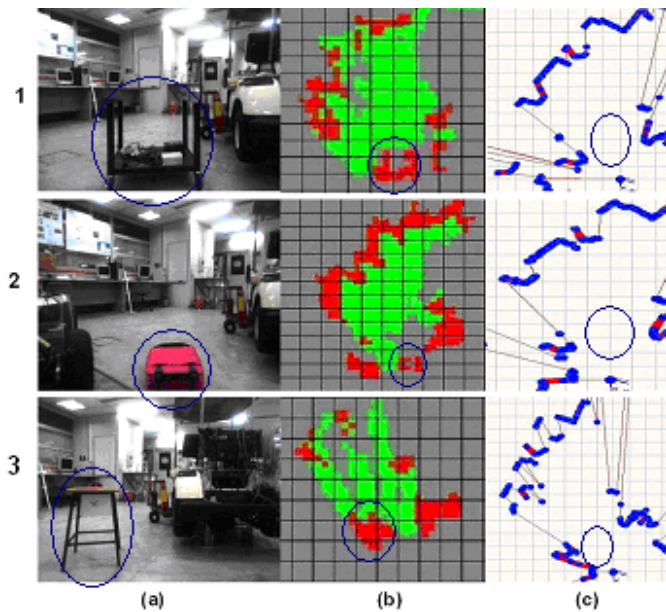
Figure 7. (a) Real images with different obstacles (b) Stereo 2D map (c) Laser beam scan

in isolation for cost map generation and path planning. Both laser and vision data are necessary for the indoor path planning. We plan to extend our work by using color information to improve the ground floor classification by using adaptive learning algorithms. We also plan to investigate how to plan a path directly from the disparity image instead of projecting it to 2D grid map first. Although it may be useful only for local path planning, it is more efficient and gives more information than the standard 2D cost map.
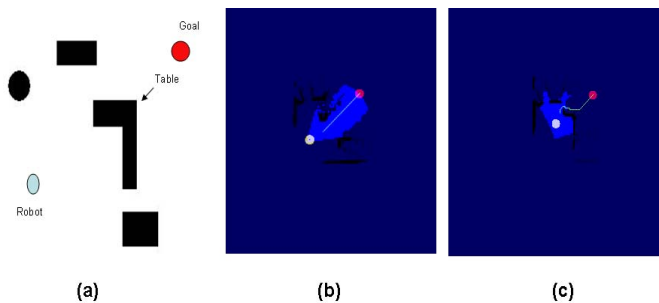

Figure 8. (a)An example scene (b)Laser cost map(c)Fused cost map

REFERENCES

[1] S. Thrun,W. Burgard, and D. Fox. "A real-time algorithm for mobile robot mapping with applications to multi-robot and 3D mapping". In IEEE Intl. Conf. on Robotics and Automation, 2000

[2] S. Thrun , A. Buecken ,W. Burgard, D. Fox , T. Froehlinghaus, D. Hennig,T. Hofmann,M. Krell , T. Schmidt "Map learning and high-speed navigation in RHINO". Technical report IAI-TR-96-3, 1996.

[3] J.-S. Gutmann and K. Konolige. "Incremental mapping of large cyclic environments". In Int. Symp. on Computational Intelligence in Robotics and Automation (CIRA'99), Monterey, November 1999.

[4] R.R. Brooks and S.S. Iyengar, "Multi-sensor fusion "Prentice Hall, New Jersey, USA, 1998.

[5] M. Lanthier, D. Nussbaum, A. Sheng, "Improving vision-based maps by using sonar and infrared data", in Proc. 10th Int. Conf. on Robotics and Applications, IASTED 2004, Honolulu, Hawaii, USA, August 2004, pp. 118-123

[6] S. Kagami, K. Okada, M. Inaba, and H. Inoue. "Design and implementation of onbody real-time depthmap generation system". In Proc. of 2000 IEEE Int. Conf. on Robotics and Automation, pp. 1441–1446, 2000.

[7] D. Murray and J. Little. "Using real-time stereo vision for mobile robot navigation". autonomous robots, 2000 ,Vol. 8, No. 2, pp. 161–171.

[8] L. Romero, A. Núñez, S. Bravo, Luis E. Gamboa. "Fusing a Laser Range Finder and a Stereo

[9] J. Miura , Y. Negishi, and Y. Shirai, "Mobile robot map generation by integrating omnidirectional stereo and laser range finder", Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 250-255, Lausanne, Switzerland,Oct 2002

[10] PTGrey, 2004. Point grey research inc. http://www.ptgrey.com/

[11] L. Matthies, P. Grandjean "Stochastic performance modeling and evaluation of obstacle detectability with imaging range sensors". IEEE Trans Robot Autom ,1994 , 10:783–792

[12] A. Elfes, "Using occupancy grids for mobile robot perception and navigation", IEEE Computer Magazine, special issue on Autonomous Intelligent Machines, June 1989, pp. 46-58.

[13] S. Thrun, "Robotic mapping: a survey", In G. Lakemeyer and B. Nebel, editors, Exploring Artificial Intelligence in the New Millenium, 2002.

[14] G. Dissanayake, P. Newman, S. Clark, H.F. Durrant-Whyte, and M. Csorba. "An experimental and theoretical investigation nto simultaneous localisation and map building (SLAM)". In P. Corke and J. Trevelyan, editors, Lecture Notes in Control nd Information Sciences: Experimental Robotics VI, page 265–274, London, 2000. Springer Verlag.

[15] F. Dellaert, S. M. Seitz, C. E. Thorpe, and S. Thrun. "EM, MCMC, and chain flipping for structure from motion with unknown correspondence". *Machine Learning*, 50:45–71, 2003.

[16] A. Talukder, R. Manduchi, L. Matthies, and A. Rankin. "Fast and reliable obstacle detection and segmentation for cross country navigation". IEEE Intelligent Vehicles , 2002 .

[17] R. Labayrade, D. Aubert, and J.-P. Tarel. "Real time obstacle detection on non flat road geometry through V-disparity representation." In IEEE Intelligent Vehicles Symposium, Versailles, pages 646–651 June 2002.

[18] F. Ferrari, E. Grosso, G. Sandini, M. Magrassi, "A stereo vision system for real time obstacle avoidance in unknown environment", in Proceedings of IEEE International Workshop on Intelligent Robots and Systems 1990, pp. 703–708.

[19] 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, 24(6), 381-395.

[20] S. Se, M. Brady, "Ground plane estimation, error analysis and applications". Robotics and Autonomous Systems ,2002,39(2): 59-71

[21] J. Borenstein and Y. Koren, "The Vector Field Histogram: Fast Obstacle Avoidance for Mobile Robots," *IEEE Trans. Robotics and Automation,* Vol. 7, No. 3, June 1991, pp. 278-288.

[22] Ulrish and J. Borenstein, "VFH*: local obstacle avoidance with look-ahead verification", ICRA, 2000.