Mapping and Localization in 3D Environments Using a 2D Laser Scanner and a Stereo Camera

KUEN-HAN LIN, CHUN-HUA CHANG, ANDREAS DOPFER AND CHIEH-CHIH WANG
Department of Computer Science and Information Engineering
National Taiwan University
Taipei, 106 Taiwan

2D laser scanners have been widely used for accomplishing a number of challenging AI and robotics tasks such as mapping of large environments and localization in highly dynamic environments. However, using only one 2D laser scanner could be insufficient and less reliable for accomplishing tasks in 3D environments. The problem could be solved using multiple 2D laser scanners or a 3D laser scanner for performing 3D perception. Unfortunately, the cost of such 3D sensing systems is still too high for enabling AI and robotics applications. In this paper, we propose to use a 2D laser scanner and a stereo camera for accomplishing simultaneous localization and mapping (SLAM) in 3D indoor environments in which the 2D laser scanner is used for SLAM and the stereo camera is used for 3D mapping. The experimental results demonstrate that the proposed system is lower cost yet effective, and the obstacle detection rate is significant improved compares to using one 2D laser scanner for mapping.

Keywords: localization, mapping, navigation, stereo vision, range sensing

1. INTRODUCTION

Mapping and localization are two of the most important functions in a number of robotics applications. Performing mapping often needs accurate localization, and accurate maps are critical for localization. When a robot does not have sufficient information of its surroundings, it is difficult to use the robot’s sensors to perform localization; when the robot has low confidence about its location, it is difficult to build a map. The Simultaneous Localization and Mapping (SLAM) approach solves the localization and mapping problems concurrently [1] which has become one of the most important functions in robotics. Due to the high accuracy of 2D laser scanners in terms of both range and bearing measurements, 2D laser scanners have been widely used to solve the SLAM problem. Our previous work [2] has demonstrated the feasibility of SLAM with generalized objects using a 2D laser scanner from a moving car in a crowded urban area. However, a 2D laser scanner only provides measurements from the plane where it is placed. For obstacles and objects at different heights, a 2D laser scanner could provide no information. For instance, tables and chairs are very common in indoor environments, but it is very likely that only legs of tables and chairs are detectable using 2D laser scanners. The surfaces of tables or chairs might lie on different planes, which is not detectable by a 2D laser scanner and can be dangerous to robots.

To solve this issue, it could be straightforward to use a 3D laser scanner to perform...
obstacle avoidance in a 3D environment [3]. Unfortunately, the cost of 3D laser scanners is still too high to enable applications. It could be also feasible to place several 2D laser scanners or to rotate a 2D laser scanner to collect the 3D information of obstacles at different height [4-6]. The cost of multiple 2D laser scanners could be still high. Using motors to rotate a 2D laser scanner could limit the real-time performance of 3D sensing. Instead of using laser scanners, stereo cameras are often used to perform 3D sensing and obstacle avoidance which can provide near distance and partial 3D information of obstacles. SLAM using a stereo camera has been demonstrated [7, 8]. However, stereo-based SLAM often represents the map by sparse features in which the description of the environment could be insufficient for robot navigation. In addition, the accuracy of localization from a stereo camera is generally inferior to that using a 2D laser scanner.

There are a few works in which the 2D laser scanner and the stereo camera are used together for accomplishing some specific tasks. In [9], the stereo camera was used to estimate the ground which is then used to filter out laser scan points generated by hitting the ground. The two sensors were also used together to eliminate false positives in object detection. However, their work focused on providing frame by frame obstacle information while ours aims at providing a 2D map which is embedded with 3D obstacle information and can be further used in applications such as navigation. In [10], the 2D laser scanner and the stereo camera were used for building the 3D environment map. However, their approach required the prior knowledge of the environment for defining semantic elements and needs classifier training in advance.

The objective of this work is to effectively describe obstacles at different heights into the 2D map. The proposed approach combines SLAM using a 2D laser scanner and dense 3D depth information from a stereo camera. With using a stereo camera, it can provide obstacles information at the plane different from the 2D laser scanner. The proposed system is more affordable and needs no extra motors. Robots with the proposed sensing system can have rich scene information for better path planning and obstacle avoidance. Fig. 1 shows an indoor environment mapping result built using the proposed system.

The rest of the paper is organized as follows. In section 2, the background knowledge of laser scanner based SLAM and the extrinsic parameters estimation of the laser scanner – camera system is introduced. The fusion approach of the laser scanner and stereo camera measurements is described in section 3. The experiment results are shown in sections 4 and 5 give the conclusion.
2. FUNDAMENTAL KNOWLEDGE

In this section, the fundamental knowledge of robotic mapping in an unknown environment is presented. Laser scanner based SLAM and a probabilistic map representation which effectively fuses temporal measurements into a 2D map are described. A calibration method of the laser scanner-camera system is introduced.

2.1 Laser Scanner-Based Simultaneous Localization and Mapping

In SLAM, robot localization and scene mapping are jointly estimated. This SLAM problem can be theoretically described with the robot motion prediction stage and the measurement update stage through the Bayes filter.

\[
Bel(x_t, M) = p(x_t, M | u_{t-1}, z_{t-1}) = \eta P(z_t | x_t, M) \int P(x_t | u_t, x_{t-1}) Bel(x_{t-1}, M)
\]

where \(x_t\) is the robot position at time \(t\). \(M\) is the map to be estimated, \(u_t\) and \(z_t\) are robot’s motion command and measurements at time \(t\). Through proper modeling of the sensor probability model \(P(z_t | x_t, M)\) and the motion model \(P(x_t | u_t, x_{t-1})\), measurements from motion command and measurements can be incrementally fused in a recursive fashion and more robust state estimates of the system state can be achieved.

While a 2D laser scanner is used to accomplish SLAM, the scan matching techniques [11-13] in the computer vision literature are often applied. If it is assumed that the best match between a scan collected at one specific robot pose and the map has the highest sensing probability and the rest locations has zero sensing probability, then the SLAM problem can be simplified as a map problem by incrementally matching measurements from the laser scanner. Fig. 2 shows a mapping and localization result using the scan matching approach.

Fig. 2. A mapping and localization result. Circles indicate the robot’s positions, red points are current laser scan points, white are the reconstructed occupancy grid map.

2.2 Occupancy Grid Map

During the process of laser scanner-based SLAM, the sensor noise and the uncertainty of robot pose might cause inconsistency between laser measurements and the map. If every measurement is forced to plot on the map, the belief of the map could be over-
confident. In addition, incorrect mapping results could deteriorate localization performance. Therefore, an occupancy grid mapping approach [14] was proposed to describe the uncertain map. In the occupancy grid map, the map is divided as grid cells. Each cell’s probability of being occupied is calculated by a binary Bayes filter [14]:

$$l_i^t = l_i^{t-1} + \log \left( \frac{P(m_i | z_i, x_i)}{1 - P(m_i | z_i, x_i)} \right)$$

(2)

where $l_i^t$ is the log odds ratio with the initial value $l_i^0 = 0$. $P(m_i | z_i, x_i)$ is the probability of the $i$th grid cell being occupied and observed by the robot at time $t$. $\log(\cdot)$ is the logarithm operator. Then its probability can be calculated as:

$$P(m_i) = \frac{\exp(l_i^t)}{1 + \exp(l_i^t)}$$

(3)

where $\exp(\cdot)$ is the exponential operator. This representation has the benefit of fusing new measurements through easy addition to update grid cell’s occupancy probability.

2.3 Laser-Camera Extrinsic Parameters Calibration

In order to fuse measurements from the laser scanner and the stereo camera, the extrinsic parameters between these two sensors have to be known in advance. As a laser scanner’s laser point cannot be detected by a regular camera, a calibration object is needed to provide relationship between the laser points and the visual image. Li and Liu [15] proposed a laser scanner-camera calibration method based on providing the boundary of the calibration object and laser point projections on the image. The same calibration approach is used in this work which is illustrated in Fig. 3.

![Fig. 3. Laser points hit on the two sides of the triangle plate. Their projections on the image are points $E, F$. If the extrinsic parameters are correct, these points $E$ and $F$ should lie on the edges $AB$ and $AC$, respectively.](image)

The extrinsic parameters between the laser scanner and the camera can be represented by a rotation matrix $\Phi$ and a translation vector $T$ as:

$$X = \Phi X_L + T$$

(4)
where $X_L$ and $X$ are points in the laser scanner and the camera coordinates respectively. Using Eq. (4) and camera’s intrinsic parameters $K$, laser points can be projected back into images as:

$$x \sim K(\Phi X_L + T)$$

(5)

where $x$ is the projected point on the image.

In order to calibrate, laser points which hit the triangle board has to be extracted first, and then use the initial guess of $(\Phi_0, t_0)$ to project laser points into the image as points $E, F$. Let the distances between these two points and two edges are:

$$\text{dist}(E, AB) = \frac{\|\overline{EB} \times \overline{AB}\|}{\|\overline{AB}\|}$$

(6)

When extrinsic parameters are correct, the distances should be zero. Therefore, the error function can be defined as:

$$\min \{ \sum_{i=1}^{N} \text{dist}(E_i, A_i B_i) + \text{dist}(F_i, A_i C_i) \}$$

(7)

where $N$ represent the laser data and image numbers. Using the non-linear optimization method such as the Levenberg-Marquardt algorithm [16], the extrinsic parameters $(\Phi, t)$ between the laser scanner and the camera can be determined. The final calibration result is shown in Fig. 4.

Fig. 4. A laser scanner-camera calibration result. Note that the laser points (red) successfully hit on the edges (blue and green) of the triangle board.

3. ASSISTANCE OF A STEREO CAMERA IN LASER SCANNER-BASED SLAM

This section describes the proposed approach of fusing stereo measurements into an occupancy grid map. First, depth noises of stereo camera measurements are filtered by a
downsampling technique. Then the error model of the stereo camera is considered during mapping. Last, the fusion method of laser scanner and stereo camera measurements are described.

3.1 Depth Noise Filtering

For a pair of stereo images, the Block Matching (BM) method [17] can be used to calculate the depth of images effectively. Since the BM method is based on local image matching results, there might be some outliers. In order to filter out these incorrect depth measurements, a common downsampling method is used. For a depth image with width $w$ and height $h$, it is divided as several squares with width $a$. The median depth within each square is extracted to represent the depth of each square, thus a depth image with noise filtering can be acquired. Using the previous calculated extrinsic parameters, depth points that either are above the robot (e.g. ceiling), or too low (e.g. floor) are removed. Fig. 5 shows a noise filtering result; only depth points that could possibly affect robot’s motion were reserved.

![Fig. 5](image)

(a) The original depth image calculated using the BM method; (b) Down sampling the image by dividing the depth image as grid cells; (c) Filter the depth points that are too high or too low.

3.2 Error Model of Stereo Camera’s Depth Measurements

As stereo camera depth measurements are more uncertain than laser scanner’s measurements, the depth uncertainty of the stereo camera has to be considered in the following mapping procedure:

$$
\Delta D = - \frac{bf}{d^2} \Delta d
$$

where $\Delta D$ is the depth uncertainty, $b$ and $f$ is the stereo camera’s baseline and focal length. $d$ is the disparity calculated by the BM method, and $\Delta d$ is the disparity error. It can be seen that depth error became larger when the disparity became smaller. Thus the depth error will become larger when the depth is farther. If $\Delta d$ is assumed to follow a normal distribution with zero mean and sigma of $\sigma$, then $\Delta D$ will follow a normal distribution with zero mean and sigma of $\sigma' = \sigma \cdot \frac{bf}{d^2}$. 
For each depth point projected on the $xy$-plane of the laser scanner coordinate, only grid cells within $2\sigma'$ bound of that depth point will be updated by Eq. (2). Grid cells outside the bound are ignored.

### 3.3 Fusion of Laser and Camera’s Measurements

After acquiring a filtered depth image, each depth point can be projected back to the $xy$-plane of the laser scanner coordinate and the corresponding points between the laser scanner and the cameras can be found. If the depth of each depth point from the stereo camera is farther than the laser scanner depth measurement, the point will be ignored and is not considered in mapping. On the contrast, if the depth point is nearer than the laser measurement, the point will be added to the occupancy grid map. In addition, depth points with the same laser beam direction will be organized as the same group and reordered according to their depths in an ascending order. Only the depth points whose depths are between 25% and 75% of the group are considered in mapping. This procedure can further remove several sparse error depth points. Fig. 6 shows the depth errors of upper lamps in the image where these depth estimates are too near. Within the same direction of depth point’s group, these sparse error depth points will locate on the two sides of depth distribution. The proposed filtering procedure removes depth points with extreme values.

![Fig. 6. (a) Original image; (b) Depth image calculated by the BM method; (c) Depth points after filtering. Note that there are still some sparse error depth measurements on the top; (d) A depth points distribution in one laser beam direction shown in (c). The extreme near depth point is filtered out in which only depth points between 25% to 75% (red dashed lines) are maintained.](image)

After acquiring occupancy grid maps build by the laser scanner and the stereo cam-
era, the maps can be fused by choosing higher occupancy grid cells from two maps as the final fused map. An example is shown in Fig. 7. The process diagram of the propose approach is shown in Fig. 8.

![Fig. 7](image1)

Fig. 7. (a) The laser scanner occupancy grid map; (b) The stereo camera occupancy grid map; (c) The fused map. Green points are the map build from the stereo camera.

![Fig. 8](image2)

Fig. 8. The process diagram.

### 4. EXPERIMENTS

This section gives the evaluation of the proposed system in indoor environments. The experimental platform that used in this work is first described, and the system was tested in an indoor environment and compared with the ground truth to demonstrate the obstacle detection performance. Finally, the feasibility of the proposed system is shown in a large indoor environment.

#### 4.1 Experimental Platform

In this work, a mobile-robot, NTU-PAL 7, is used as the experimental platform. The robot is equipped with a 2D laser scanner, SICK-S200, with the field of view of 270 degrees and the maximum range of 50 meters with the resolution of 0.1 mini-meter. For the stereo camera, a Point Grey’s Bumblebee X3 camera with baseline of 12 cm is used for the experiment. The field of view of the camera is 60 degrees, and the resolution of the image is $640 \times 480$. The depth resolution of the stereo camera can be calculated using the Eq. (8); the resolution is 0.4 m when the depth is 5 m with 1 pixel image error. The image is already undistorted and rectified. Extrinsic parameters of the laser scanner and the camera have been pre-calibrated. The camera is on top of the laser scanner by 23 cm, and the laser scanner is above the ground by 20 cm. In this work, the square size of $a = 8$ is used to filter depth image noise. The overall processing time per each frame without display is 120 ms on a desktop PC with a quad-core 2.8Ghz processor.
4.2 Experiment Result with Ground Truth

Fig. 10 shows the experiment scenario of a room of 8.4 m by 8.4 m. Four tables are within the room. The height of the laser scanner can only detect the legs of the tables. The edges of the tables are pasted with wall paper for the stereo camera to calculate its depth.

The ground truth was acquired from a laser scanner based SLAM result [2] of the experiment scenario. By adjusting laser’s height up 30 cm, the height of the laser scanner could detect the tables. It is noted that we manually set the obstacles detectable on this height, thus the ground truth map could be generated by rising up the laser scanner.

Three experiment data and one ground truth data are collected for comparison.

Fig. 11 shows mapping results using the 2D laser scanner only, using a stereo camera only, and the fused map. Compare to the ground truth, the tables can be detected and added into the map after adding stereo camera’s information. The confidence of each grid cell being an obstacle depends on its occupied probability \( P(m_i) \). In order to evaluate obstacle detection performance, we defined that if \( P(m_i) > 0.5 \), then this grid cell is classified as an obstacle.

\[
\text{obstacle} = \begin{cases} 
\text{True} & \text{if } P(m_i) > 0.5 \\
\text{False} & \text{if } P(m_i) < 0.5 
\end{cases}
\]
Table 1. Obstacle detection rate.

<table>
<thead>
<tr>
<th></th>
<th>Laser map</th>
<th>Stereo map</th>
<th>Fused map</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obstacle detection</td>
<td>48.45%</td>
<td>78.71%</td>
<td>80.12%</td>
</tr>
<tr>
<td>rate</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1 shows the obstacle detection rate of three different maps. As different sensors have different fields of view and depth ranges, only discovered grid cells (non-grey area) in each map are evaluated. The obstacle detection rate of each map is then calculated by comparing correct obstacle grids versus true obstacle grids in the ground truth map. The result showed that the use of the stereo camera is able to detect unseen obstacles from the laser scanner. Table 1 shows that as the tables cannot be detected using a 2D laser scanner, its obstacle detection rate is obviously lower than the other two maps. On the other hand, the fused map has higher obstacle detection rate than using stereo camera only.

The experiment also showed that in some texture less areas (e.g. walls with uniform color), there were some incorrect obstacle detection results near walls as the stereo camera could not provide reliable depth measurements.

4.3 Large Environment Real Data Result

The second experiment scenario is at the basement of our department. The scenario is full of undetectable chairs and tables due to the height of the laser scanner, which is good to demonstrate the performance of the proposed system. Fig. 12 shows the part of the experiment scenario. Fig. 13 shows that laser scanner-based SLAM could not effectively detect obstacles within the environment, while the proposed method successfully
Fig. 12. Basement scenario.

Fig. 13. (a) Laser only map. Most obstacles are walls or the legs of tables; (b) The fused map. Green points are the map build from the stereo camera.

plotted parts of chairs and tables in the map. This information is critical in obstacle avoidance and navigation. The experimental video is available online at: http://robotics.csie.ntu.edu.tw/~linsm/videos/JISE/stereoLaserSlam.mpg.

5. CONCLUSION

This work proposed a method to use a stereo camera to detect unseen obstacles by using a 2D laser scanner only. With the depth image calculated from the stereo camera and the pre-computed laser scanner-camera system’s extrinsic parameters, we can project the depth points into the 2D map built using the laser scanner. After comparing with laser depth measurement’s and stereo depth measurement’s uncertainty, the measurements can be fused into a single occupancy grid map for describing obstacles in 3D environments.

The experiment results have demonstrated that the fused map had a better obstacle detection rate comparing to mapping using only a 2D laser scanner or using only a stereo camera. The large-scale environment experiment demonstrated the feasibility of the proposed system to detect unseen obstacles in general indoor environments.

In this work, the stereo camera is mainly used for mapping, but it can also be used to solve the localization problem. Fusing this capability into the current system could have a better localization and mapping performance. On the other hand, the proposed system can also extend the map from 2.5D to 3D, which will be helpful in obstacle avoidance as well as in object recognition.
REFERENCES

Kuen-Han Lin (林昆翰) received the M.S. degree in Mechanical Engineering from National Taiwan University in 2009. He then worked as a research assistant of the Robot Perception and Learning Laboratory at National Taiwan University from 2009 to 2011. He is a Ph.D. student in Electrical Computer Engineering at University of California, San Diego since 2011. His research interests include robotics, computer vision and machine learning.

Chun-Hua Chang (張峻華) received his B.S. degree in Computer Science and Information Engineering from National Taiwan University, Taipei, Taiwan, in 2009. He is currently a Ph.D. student at the Robot Perception and Learning laboratory at National Taiwan University. His research interests include robot localization, mapping, target tracking, multi-robot cooperation, and dynamic scene understanding.

Andreas Dopfer (薛菲) received his M.S. degree in Computer Science from National Taiwan University, Taipei, Taiwan in 2009. He is currently a Ph.D. student at the Robot Perception and Learning Laboratory at National Taiwan University. His research interests include mapping, multi-sensor perception and dynamic scene understanding.
Chieh-Chih (Bob) Wang (王傑智) earned his Ph.D. in Robotics from the School of Computer Science at Carnegie Mellon University in 2004. He received his B.S. and M.S. from National Taiwan University in 1994 and 1996, respectively. During his graduate study, he worked with the Bayesian vision group at the NASA Ames research center and at Z+F Inc. in Pittsburgh. From 2004 to 2005, he was an Australian Research Council (ARC) Research Fellow of the ARC Centre of Excellence for Autonomous Systems and the Australian Centre for Field Robotics at the University of Sydney. In 2005, he joined the Department of Computer Science and Information Engineering at National Taiwan University, where he is an Associate Professor and is pursuing his academic interests in robotics, machine perception and machine learning. Dr. Wang received the best conference paper award at the 2003 IEEE International Conference on Robotics and Automation (ICRA).