A Probabilistic Approach to Build 2D Line Based Maps from Laser Scans in Indoor Environments

Leonardo Romero and Carlos Lara

Michoacana University
Morelia, Mich., Mexico
lromero@umich.mx, calara@lsc.fie.umich.mx

Abstract. In this work we consider a mobile robot with a laser range finder. Our goal is to find the best set of lines from the sequence of points given by a laser scan. We propose a probabilistic method to deal with noisy laser scans, in which the noise is not properly modeled using a Gaussian Distribution. An experimental comparison with a very well known method (SMSM), using a mobile robot simulator and a real mobile robot, shows the robustness of the new method. The new method is also fast enough to be used in real time.

1 Introduction

This paper focuses on methods to resolve the problem of fitting a set of lines to a sequence of points. Specifically, the sequence of points are acquired by using a laser range finder mounted on a mobile robot. The set of lines can be used in complex activities such as Localization and Mapping in indoor environments.

Localization is a key activity for mobile robots which consists of determining the robot's position in its environment. Localization can be done by matching the newest sensed data against information in a priori map. There are two common matching techniques that have been used in mobile robotics: point–based matching and feature–based matching.

In the feature–based matching approach it is necessary to learn features from the environment. Features are recognizable structures of elements in the environment [12]. The simplest feature is a line or a line segment. A Line Based Map (LM) is a set of lines which represent the robot's environment. Owing to indoor environments are usually rich in planar surfaces, LMs are the natural way to represent them. In addition, LMs are more compact and more accurate than point–based maps.

For extracting a LM from a laser scan, many algorithms have been proposed, but all of them have their shortcomings. Forsyth [6] identifies three principal difficulties in this process: 1) find the best number of lines, 2) determine which points belong to which line, and 3) estimate the line parameters given the points that belong to a line. Furthermore, if the robot navigates within a high cluttered

environment, a lot of points do not belong to any line, and the process becomes more difficult.

This paper introduces a new approach called WSAC–GE, Window SAmpling Consensus with Global Evaluation. The algorithm begins with an empty map, and successively proposes a new map by adding or removing lines from the previous map. The current map is selected from two maps: the proposed new map or the previous map, according to which map is best evaluated by a global function. Problems associated with outliers (atypical data) are handled by a probabilistic search and different segments of the same line are identified easily. Other methods have difficulties with outliers or report many disconnected segments when in fact they belong to the same line.

This paper is organized as follows: Section 2 describes how a laser acquires information from the environment. Section 3 describes related works, and it focuses on the most used algorithms to extract lines. Section 4 describes the proposed method. Section 5 compares the performance of WSAC–GE against a good and well known method called SMSM (Split–Merge–Split–Merge). Finally, section 6 explains the advantages of the proposed method.

2   Data Acquisition

In a single measurement, a laser range finder calculates the distance to the object in a given orientation. Usually, each measurement is expressed in the form (r_i, α_i); where r_i is the distance from the sensor to the detected object at direction α_i. A laser scan performs n consecutive measurements from the environment changing the orientation in $\Delta\alpha$. Also, the laser scan can be expressed in its equivalent cartesian form $\mathcal{P} = \{(x_i, y_i) \,|\, i = 1 \dots n\}$.

3   Related Works

To obtain a LM from a 2D laser scan many algorithms have been proposed. Two approaches, *Line Tracking* (LT) [4] and *Iterative End-Point Fit* (IEPF) [4] represent the classical way to solve the problem. These two algorithms take advantage of the sequentiality of the laser scan. The principal drawback of LT and IEPF is their poor performance when the laser scan is noisy. On other hand they are extremely fast. The Hough Transform Algorithm (HT) [8] is another common alternative used to generate line Maps . This technique has a good performance even when there are outliers. However, the principal drawbacks are associated with the discretization, its slow speed and that it ignores the uncertainty for estimating the line parameters. Finally, the well known Expectation-Maximization algorithm (EM) [3] is a technique used in several kind of problems with missed data. But it requires to know in advance how many line models exist in the environment. In [7] this difficult is solved by incrementally adding and removing models. Due to the SMSM algorithm has been widely used with good results [11], we compare the proposed method against SMSM. The next section describes the SMSM algorithm and section 3.2 focuses in methods commonly used to manage atypical data.
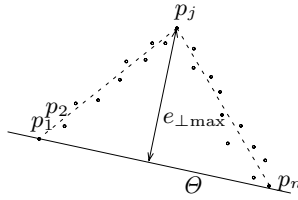
Fig. 1. Iterative End Point Fit

Algorithm 1. IEPF Algorithm

Input: The laser scan $\mathcal{P} = (p_1, p_2, \ldots, p_n)$ and a threshold t

Output: The map of lines $\mathcal{M}$

1. Initialize the set of sequences $\mathcal{L} \leftarrow \{\mathcal{P}\}$ and $\mathcal{M} \leftarrow \{\}$.
2. While $\mathcal{L} \neq \{\}$ do
    (a) Move an element of $\mathcal{L}$ to S
    (b) Calculate the line Θ joining the first and last point of S.
    (c) Detect the point p_j with maximum distance $e_{\perp max}$ to the line Θ
    (d) *If* $e_{\perp max} > t$ *then* Put $S_0 = (p_k | k = 1 \ldots j - 1)$ and $S_1 = (p_k | k = j \ldots n_j)$ into $\mathcal{L}$
    (e) *If* $e_{\perp max} \leq t$ *then* Fit a line Θ^* to all the points in S and put Θ^* into $\mathcal{M}$.
3. Merge collinear segments in $\mathcal{M}$.

3.1 The SMSM Algorithm

The *Split–Merge Split–Merge* algorithm (SMSM)[17] is based in the IEPF algorithm. The IEPF algorithm [4] is described by the algorithm 1. and it is illustrated graphically in the figure 1.The SMSM algorithm is an extended and more robust version of the IEPF algorithm. At the beginning it finds clusters applying a breakpoint detector [1]. The idea behind this step is to detect and eliminate outliers, because they are not going to be included into any cluster. Then it merges two consecutive clusters if their distance (the distance between the final point of the first cluster and the first point of the second cluster) is less than a predefined threshold. In the second phase, SMSM applies the IEPF algorithm to all clusters. Finally it combines collinear segments.

3.2 Robust Regression

The goal of regression is to describe data using a model. Unfortunately, in the data set often exists a kind of data called outliers that makes the regression a hard problem. An outlier is a single observation far away from the rest of the data. Two techniques are frequently used to obtain a model from data contaminated with outliers: The Random Sample Consensus Algorithm (RANSAC) [5] and the M–Estimators [9].

Algorithm 2. RANSAC Algorithm

Input: The laser scan $\mathcal{P} = (p_1, p_2, \ldots, p_n)$, a threshold t and a maximum number of iterations m

Output: Θ^* (the best line given $\mathcal{P}$)

1. $j \leftarrow 0$, $n^* \leftarrow 0$.
2. While $j < m$ do
   (a) $j \leftarrow j + 1$
   (b) Select randomly two points from $\mathcal{P}$ and compute the line Θ, that joins them
   (c) Count the number of inliers, n, from $\mathcal{P}$ given Θ and t
   (d) if $n > n^*$ then $n^* \leftarrow n$, $\Theta^* \leftarrow \Theta$
3. Reestimate Θ^* using all the inliers.

The Random Sample Consensus algorithm consists in to iteratively propose new random models, and to evaluate each model proposed. After a number of tries, RANSAC selects the best evaluated model to represent the data. The version of RANSAC applied to the problem of fitting a line is described in the Algorithm 2.. The RANSAC algorithm has the advantage of estimating parameters of a model with accuracy even when outliers are present in the data set. Frequently RANSAC is a better choice than other algorithms due it can be better adapted to complex data analysis situations [10]. One of the problems of RANSAC is deciding the threshold t for considering a point as an inlier. If t is set too high then the model estimation can be very poor.

Algorithm 3. A generalized RANSAC Algorithm which uses a M–Estimator

---

Input: Tha laser scan $\mathcal{P} = (p_1, p_2, \ldots, p_n)$ a weight function g and a maximum number of iterations m

Output: Θ^* (the best fit line given $\mathcal{P}$)

1. $j \leftarrow 0$, $n^* \leftarrow 0$.
2. While $j < m$ do
   (a) $j \leftarrow j + 1$
   (b) Select randomly two points from $\mathcal{P}$ and compute the line Θ, that joins them
   (c) Calculate $n = \sum_{i=1}^{n} g(e_i)$. Where e_i is the distance from the point p_i to the line Θ
   (d) if $n > n^*$ then $n^* \leftarrow n_j$, $\Theta^* \leftarrow \Theta$
3. Reestimate Θ^*.

In the other hand, M–Estimators reject outliers by weighting each point using a so called weight function. Under certain circumstances, M-estimators can be vulnerable to high-leverage observations [2]. Many authors [13,16,15] proposed the combination of the RANSAC and M–Estimator methods to improve the performance. As an example, the algorithm 3. shows a RANSAC algorithm where the count of inliers in the algorithm 2. is replaced using a weighting function g.

4   Proposed Method: WSAC-GE

Suppose that there is an evaluation function $\mathcal{H}(\mathcal{M}, \mathcal{P})$ which returns a real value related with the correspondence of the map $\mathcal{M}$ (a set of lines) given the laser scan $\mathcal{P}$ (a sequence of points). Using $\mathcal{H}$ it is possible to decide if the map $\mathcal{M}'$ is better than $\mathcal{M}$, for a given laser scan $\mathcal{P}$, if $\mathcal{H}(\mathcal{M}', \mathcal{P}) > \mathcal{H}(\mathcal{M}, \mathcal{P})$. It is the basic idea under the WSAC-GE algorithm. The algorithm 4. proposes a new map $\mathcal{M}'$ from the previous one $\mathcal{M}$ and then it decides which one is better. The line addition and line deletion mechanisms provide ways to obtain $\mathcal{M}'$ from $\mathcal{M}$.

Algorithm 4. WSAC–GE

Input: The laser scan $\mathcal{P} = (p_1, p_2, \ldots, p_n)$ and a maximum number of iterations m
Output: A line based map $\mathcal{M}$

1. Let $\mathcal{M} \leftarrow \{\}$ and $i \leftarrow 0$
2. While $i < m$ do
    (a) $i \leftarrow i + 1$
    (b) Line addition
          i. Select randomly a window w of consecutive points in $\mathcal{P}$
          ii. From the point in w, fit a line Θ
          iii. Propose a new map $\mathcal{M}' \leftarrow \mathcal{M} \cup \{\Theta\}$,
          iv. If $\mathcal{H}(\mathcal{M}', P) > \mathcal{H}(\mathcal{M}, P)$ then $\mathcal{M} \leftarrow \mathcal{M}'$ and $i \leftarrow 0$.
    (c) Line Deletion
          i. Select randomly a line $\Theta_j \in \mathcal{M}$,
          ii. Propose a new map $\mathcal{M}' \leftarrow \mathcal{M} - \{\Theta_j\}$,
          iii. If $\mathcal{H}(\mathcal{M}', P) > \mathcal{H}(\mathcal{M}, P)$ then $\mathcal{M} \leftarrow \mathcal{M}'$ and $i \leftarrow 0$.
3. Associate points with lines and refine the final lines.

4.1   Line Addition Mechanism

The first mechanism, line addition, generates a new map $\mathcal{M}'$ by adding a new line Θ_j to $\mathcal{M}$, where $j = |\mathcal{M}| + 1$. The line's parameters Θ are obtained by fitting a line into a small quantity of points w called a window.

Given the line Θ_j the algorithm calculates the weighted–error c_{ij}, from each point p_i to line Θ_j, using:

$$c_{ij} = g(e_{\perp ij})$$

Where e_{ij} is the orthogonal distance of the point p_i to the line Θ_j and the function $g(x)$ is the Beaton-Tukey function [14] (an M-Estimator):

$$g(e) = \begin{cases} [1 - (e/k)^2]^2 & \text{if } |e| < k \quad \text{inliers} \\ 0 & \text{elsewhere outliers} \end{cases}$$

To improve the results we propose to penalize points which belong to small segments. Recalling that the weighted function $g(\cdot)$ is evaluated with 0 for $|e_{\perp ij}| > k$ and then it is possible to get a set of inliers S. With the ordered set S and using a breakpoint detector [1] it is possible to find the segments of Θ. A Breakpoint

finds the segments of a line by finding the discontinuities of a ordered set of points. Finally, the algorithm penalizes every point p_i which belong to small–length segments by doing: $c_{ij} \leftarrow k\,(c_{ij})$,.where $0 \leq k < 1$ is a penalization constant.

4.2   Line Remotion Mechanism

The objective of the line–remotion mechanism is to remove small-length segments supported by points which belongs to larger segments. This mechanism randomly selects a line Θ from the map $\mathcal{M}$ and generates a new maps by removing Θ, as showed in the algorithm 4..

4.3   The Global–Evaluation

Let $\mathcal{M} = \{\Theta_1, \ldots, \Theta_l\}$ be a map with l lines. Table 1 shows a matrix representation of $\mathcal{M}$, each cell has the weighted error c_{ij}. The i–th column represents the point $p_i \in \mathcal{P}$ and the j–th row represents the line $\Theta_j \in \mathcal{M}$.

Table 1. Weighted–Errors c_{ij}

	p_1	p_2	$\cdots$	p_n
Θ_1	c_{11}	c_{21}	$\cdots$	c_{n1}
	$\vdots$	$\vdots$	$\ddots$	$\vdots$
Θ_l	c_{1l}	c_{2l}	$\cdots$	c_{nl}
	s_1	s_2	$\cdots$	s_n

Where s_i in Table 1 is computed by

$$s_i = \begin{cases} s_i' & \text{if } s_i' > u \\ 1 & elsewhere \end{cases}$$

and s_i' is the maximum value of the i-th column $(c_{ij}, j = 1, \cdots, l)$ and u is a threshold. Using s_i, a normalized weight h_j of the j–th line is given by

$$h_j = \sum_{i=1}^{n} \frac{c_{ij}}{s_i} \tag{1}$$

A high value for h_j denotes a line with a high support given the set of points. Finally the function $\mathcal{H}$ is given by

$$\mathcal{H}(\mathcal{M}, \mathcal{P}) = \sum_{j=1}^{l} h_j^2 \tag{2}$$

4.4   Associating Points with Lines

If After m tries the algorithm can not find a better map, then it finishes. The result is the best map $\dot{\mathcal{M}}$. WSAC–GE associates the point p_i to its the closest line by selecting the maximum value of the i–th column in the table 1.

5   Experimental Tests

We perform two tests: the first one is based on simulated data of a laser range finder mounted on a mobile robot in a structured environment and in the second one we use our mobile robot equipped with a LMS209-S02 laser.
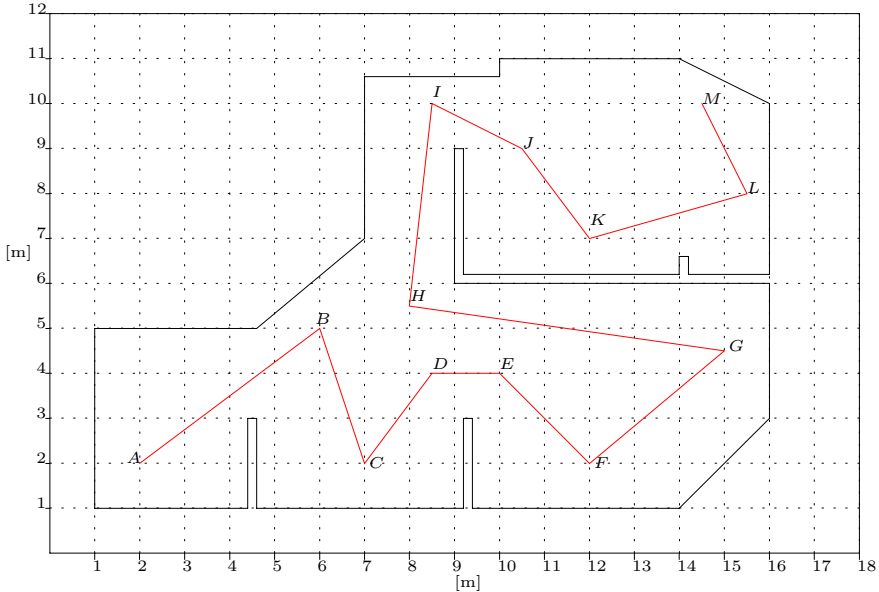

**Fig. 2.** Synthetic Environment

5.1   Test Using Synthetic Data

The computer–simulated environment has 12 walls, as shown in figure 2. The aim of this experiment is to evaluate the robustness of the proposed method against the popular SMSM method. In the experiment the robot follows the path ABCDEFGHIJKLM getting a total of 300 synthetic laser scans.

The parameters of the laser are $\Delta\alpha = 0.5°$, maximum distance $32m$ and the number of lectures $n = 361$. Also, 20 percent of the measures distances were contaminated by spurious noise simulated by adding a uniform random value between 0 and the maximum distance. The remaining measures were only contaminated with a Gaussian random noise with $\sigma_r = 0.03m$.

For this test we use a laptop HP Pavilion with Celeron processor, 1.1 GHz with 256 Mb and we use the C language to program the algorithms. The table 2 summarizes the results in the simulated environment. As it is shown, WSAC–GE is getting favorable results both in the parameters of the line ($|\Delta r|$ and $|\Delta\alpha|$) and in the assertivity. The assertivity a is given by a $= (100) \cdot \frac{n_s}{n_n}$ where n_s is the number of lines that match with lines in the simulated environment, and n_n is the total of lines extracted.
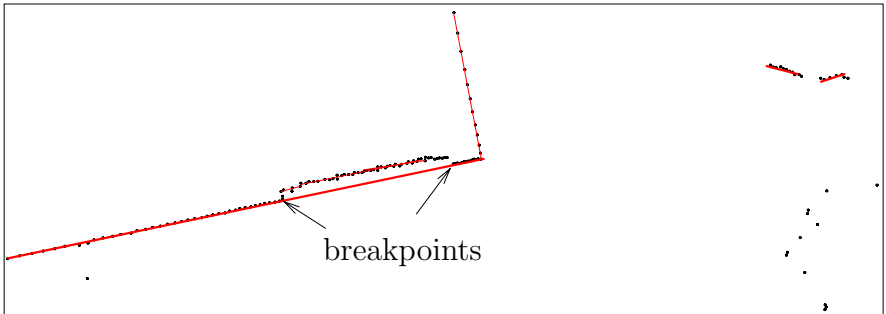
Table 2. Test Results for 300 simulated cases

Algorithm	Maximum time [ms]	$|\Delta\sigma_r|$ [mm]	$|\Delta\alpha|$ [°]	lines n_n	Assertivity %
SMSM	29.5	6.37	0.2670	1557	87.4
WSAC–GE	94.5	4.12	0.1323	1620	94.8
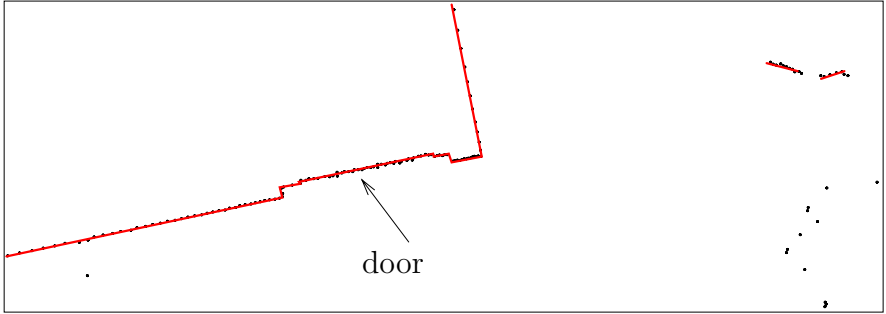
(a) Ideal LM
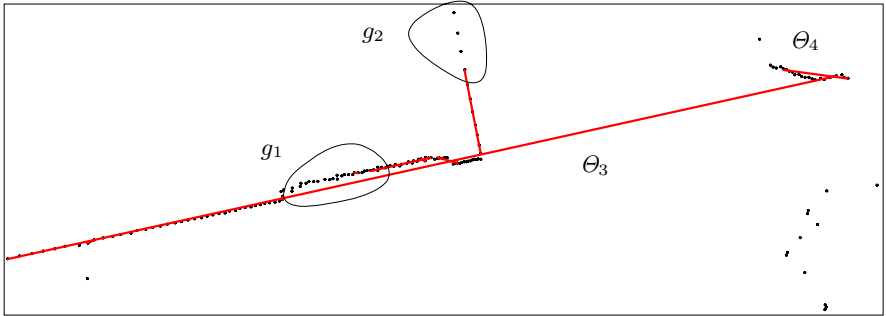
(b) SMSM

(c) WSAC–GE

Fig. 3. Results on real environment

5.2   Test Using Real Data

Figure 3 shows the results of SMSM and WSAC-GE methods in a real environment. The figure 3(a) shows the ideal map for this sample. As shown in the figure 3(b) the SMSM algorithm does not associate some points with lines (groups g_1 and g_2). In other cases, SMSM merges segments which do not represent the same object, lines Θ_3 and Θ_4 obtained does not represent precisely the environment. As shown in the figure 3(c) WSAC–GE finds lines more precise than SMSM and the asociation between points and lines is better.

6   Conclusions

We propose a robust method to find multiple lines in a laser scan, avoiding problems due to noisy data (outliers). Outliers typically does not follow a Gaussian Probability Distribution and in most cases do not have known probability distributions. While simple and fast algorithms work fine with Gaussian noise, the presence of outliers increase the complexity of the problem. The new method merges local and global strategies to solve the problem of outliers. We use a M-Estimator within a RANSAC method to find a line from a short sequence of points of the laser scan (the local strategy), then the line is evaluated and refined using the whole set of points (the global strategy), discarding those points belonging to very small segments. Once a line segment is found, the process continues looking for more lines. The key idea of this approach is to use a global evaluation function to add or remove lines from the map. The global evaluation function have higher values when more points are assigned to lines, when points belongs to single lines and when lines have more points. The proposed method is fast enough to be used in real time and it is able to find better results than the SMSM method, a very well known method reported in the literature as a very good and fast algorithm. Given its probabilistic nature WSAC–GE is able to deal with noise (outliers) with unknown probability distributions.

References

1. Geovany Araujo Borges and Marie-José Aldon. Line extraction in 2d range images for mobile robotics. *J. Intell. Robotics Syst.*, 40(3):267–297, 2004.
2. H. Chen and P. Meer. Robust regression with projection based m–estimators. In *In Proceedings of the 1991 IEEE International Conference on Computer Vision and Patern Recognition*, 2003.
3. A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm (with discussion). *Journal of the Royal Statistical Society*, 1977.
4. Richard O. Duda and Peter E. Hart. *Pattern Classification and Scene Analysis.* Wiley-Interscience, 1976.
5. Martin A. Fischler and Robert C. Bolles. Random Sample Consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.

6. David A. Forsyth and Jean Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, 2003.
7. D. Hähnel, W. Burgard, and S. Thrun. Learning compact 3D models of indoor and outdoor environments with a mobile robot. In *Robotics and Autonomous Systems 44 (2003)*. ELSEVIER, 2003.
8. P.V.C. Hough. Machine analysis of bubble chamber pictures. In *International Conference on High Energy Accelerators and Instrumentation*, 1959.
9. P. Huber. Robust location of a location parameter. *Annals of Mathematical Statistical*, 35, 1964.
10. P. Meer, C. V. Stewart, and D. E. Tyler. Introduction, robust computer vision: An interdisciplinary challenge. *Computer Vision and Image Understanding*, 2000.
11. Viet Nguyen, Agostino Martinelli, Nicola Tomatis, and Roland Siegwart. A comparison of line extraction algorithms using 2d. laser rangefinder for indoor mobile robotics. In *International Conference on Intelligent Robots and Systems (IROS 2005)*, 2005.
12. Roland Siegwart. *Introduction to Autonomus Mobile Robots*. The MIT Press, 2004.
13. Charles V. Stewart. Minpran: A new robust operator for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1995.
14. Charles V. Stewart. Robust parameter estimation in computer vision. *SIAM Review*, 41(3):513–537, 1999.
15. P. Torr and A. Zisserman. MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78:138–156, 2000.
16. P H S Torr and A Zisserman. Robust computation and parameterization of multiple view relations. In *ICCV6*, pages 727 – 732, 1998.
17. Xu Zezhong, Liu Jilin, and Xiang Zhiyu. Map building and localization using 2d range scanner. In *Proceedings 2003 IEEE International Symposium on Computational Intelligence in Mobile Robotics and Automation*, pages 848–853, 2003.