

# Simultaneous Localization and Mapping with Detection and Tracking of Moving Objects

Chieh-Chih Wang and Chuck Thorpe

Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 15213

Email: {bobwang, cet}@ri.cmu.edu

## Abstract

Both simultaneous localization and mapping (SLAM) and detection and tracking of moving objects (DTMO) play key roles in robotics and automation. For certain constrained environments, SLAM and DTMO are becoming solved problems. But for robots working outdoors, and at high speeds, SLAM and DTMO are still incomplete. In earlier works, SLAM and DTMO are treated as two separate problems. In fact, they can be complementary to one another. In this paper, we present a new method to integrate SLAM and DTMO to solve both problems simultaneously for both indoor and outdoor applications. The results of experiments carried out with CMU Navlab8 and Navlab11 vehicles with the maximum speed of 45 mph in crowded urban and suburban areas verify the described work.

## 1 Introduction

Simultaneous localization and mapping (SLAM) as first proposed by Leonard and Durrant-Whyte [16] is to simultaneously estimate positions of newly perceived landmarks and the position of the mobile robot itself while mapping. Throughout the last decade many researchers successfully extended SLAM from indoors [9] to outdoors [8], to underground [26][19] and underwater [29] environments, to airspace [27], and from 2D to 3D [14][31][25].

Until now most of the researchers on SLAM assume that the unknown environment is static, containing only rigid, non-moving objects. Moving objects are taken as noise sources. This assumption is reasonable in some applications, but moving objects and dynamic environments are not avoidable; they are even the main concerns in many applications.

The Navlab group of CMU builds robot cars, trucks, and buses, capable of autonomous driving or driver assistance. Both for automated highways or driver assistance for maneuvering in crowded city environments [21], detection and being able to handle the changes of environments are essential for the successful completion of these tasks.

For most researchers on detection and tracking of moving objects (DTMO), one of the most difficult issues is to separate moving objects and stationary objects. In

surveillance applications, even though the sensors are mounted on stationary platforms, the changes of the environment still make it difficult. Only a small amount of DTMO research mounts sensors on moving platforms [32][5][7]. The successful accomplishment of tasks relies on an accurate pose estimation system.

The results of SLAM will be more accurate if moving objects can be filtered out. Although moving objects are the troublesome parts of SLAM, they are the main concerns of DTMO. Therefore if we can integrate SLAM with DTMO, SLAM won't be affected by moving objects. Also by the more accurate pose estimation of SLAM, DTMO can detect and track moving objects more reliably. SLAM and DTMO are mutually beneficial.

In this paper, we propose a novel approach to tackle these two problems at once. We compare different SLAM approaches and get a suitable approach for general crowded city environments. A matching-based tracking method is also proposed.

The rest of this paper is arranged as follows. In Section 2 SLAM with DTMO is introduced. Section 3 shows the results of the suitable SLAM approach for urban and suburban environments. Section 4 introduces the algorithms of DTMO. Section 5 introduces the methods to accumulate and retrieve the results of SALM with DTMO. The experimental results are in Section 6, and the conclusion and future work are discussed in Section 7.

## 2 SLAM with DTMO

The problem we want to solve is SLAM with DTMO of a mobile robot using range sensing from scanning laser rangefinders. Usually people assume that a horizontal range scan is a collection of range measurements taken from a single robot position. When the robot is moving at high speeds, this assumption is invalid. We use the rotating rate of the scanning device and the velocity of the robot to correct the errors of this assumption.

Fig. 1 shows the flow diagram of SLAM with DTMO algorithm upon which our approach is based. It's not surprising that a lot of work on SLAM has been done since 1991. Here we don't specify any SLAM method. This idea of SLAM with DTMO should be adapted to varied SLAM approaches. More discussion about different kinds of SLAM methods is in Section 3.

For each frame of range data, we begin by segmenting the range data into connected objects. The tracking of

moving object is achieved by registering the objects in the ongoing “moving objects” list with the nearby objects in the current scan. Once the good matching is found, we filter out the object from the current scan and update the moving object list and the local moving object map (MO-Map). In the second step, a registration technique of range scans is used to register filtered data with the local stationary object map (SO-Map) in order to compute relative robot position. After the registration between the current scan and SO-Map is found, the moving object detection algorithm uses the precise pose to separate any new moving objects from stationary objects. Finally SO-Map, MO-Map and the moving objects list are updated, then the whole process iterates.

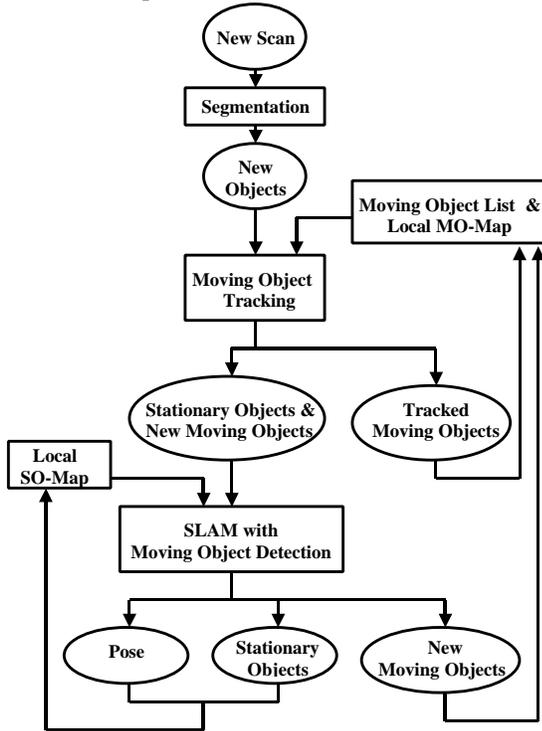


Fig. 1: The algorithm of SLAM with DTMO. The ovals are data and the rectangles are processes.

Pose estimation from Odometry, Inertial Measurement Unit (IMU) or Global Positioning System (GPS) is necessary in those cases where the most parts of the range scan belongs to moving objects or the most parts of the environment are out of the maximum range of laser range finders. Without these sensors, a model of the vehicle is needed. Global localization and mapping is accomplished with limited floor plan or digital map information.

### 3 Simultaneous Localization and Mapping for both indoor and outdoor applications

Basically, in order to build a map of un-constructed environments, we need range sensors. Range information can be from active range sensors or passive range sensors. Hebert did an excellent survey in [20]. Except for Deans and Hebert who used omni-directional camera [6], most published works in the field of SLAM use precise, active ranging sensors. Using passive range sensors in SLAM is beyond the scope of this paper.

There are two dominant approaches to solve SLAM problems. One is probabilistic approach. Here SLAM is treated as a maximum likelihood estimation problem. Thrun *et al.* [28] have demonstrated their probabilistic approach in museum environments. Another one is scan matching approach. Scan matching means finding a maximum overlap of scans through a process of translation and rotation. Establishing correspondences among scans is the first step of this approach. Based on correspondence establishment, we categorized various scan matching methods into four approaches showed in Table 1.

	Approach	Abbr.	Inventors
1	Feature to Feature	F2F	Shaffer [26], Gonzalez <i>et al.</i> [12]
2	Point to Feature	P2F	Cox [3]
3	Point to Point	P2P	Lu and Milios [17,18]
4	Combination	Com	Gutmann & Schlegel [9]

Table 1: Scan Matching Algorithms

Scott *et al.* [24] tried to compare probabilistic approach with scan matching approach quantitatively and qualitatively. Gutmann *et al.* [9,10] compared different methods of scan matching approach for indoor applications. In this paper, we discuss different methods of scan matching approach for both indoor and outdoor environments. Probabilistic approach will be implemented and compared to scan matching approach in future work.

F2F-based methods should have the shortest run-time, since by these methods hundreds of range points are reduced to dozens of features. For most indoor applications, line, circles, corners and other simple geometrical features are rich and easy to detect. Zhao and Shibasaki [31] picked a site that is similar to indoor environments and employed F2F-based method to construct an urban map successfully. Guivant *et al.* [8] used intensity (reflectance) of laser signal and geometrical primitives to define and detect features. Their approaches are still limited to some specific environments or conditions.

Fig. 2 and 3 are samples of range scans from our experiments. The blue box presents a 5 m long and 2 m

wide test vehicle. The SICK laser scanner mounted on the right side of the vehicle was doing horizontal profiling. It is okay to employ F2F approaches in the environments like Fig. 2, because lines or corners are easily detected.



Fig. 2: A laser scan sample of urban areas.



Fig. 3: A laser scan sample of suburban areas.

It is similar to indoor environments. But Fig. 3 shows that it is not reasonable to extract geometric primitives in this kind of environment. The reason is that geometric primitives can't represent all environments well. Outdoors environments contain many different kinds of objects such as bushes, trees, and curved objects whose shapes are hard to be defined. Same problem also occurs in indoor situations. It is very hard to measure the pose of these kinds of objects reliably and accurately because only parts of them can be seen. Fig. 4 illustrates that a results of circle extraction of the same object are different from different measure position. Fig. 5 illustrates the ambiguity of line extraction of curved object.

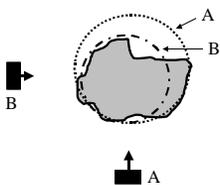


Fig. 4: Circle Extraction

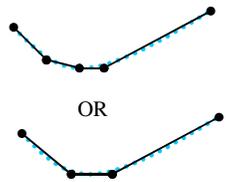


Fig. 5: Line Extraction

Basically, F2F-based approaches try to use less information to represent the raw data in order to speed up algorithms. If features cannot be detected robustly and contain some uncertainties, the whole performance of the approaches will decrease. On the contrary, P2P-based approaches don't have these disadvantages; instead, they use all the raw data.

Based on these considerations, we used P2P-based algorithms. ICP (the abbreviation of Iterative Closest Point or Iterative Corresponding Point) algorithm is one of the most successful and popular P2P algorithms. The basic idea of ICP is that using a closest-point rule to

establish correspondences between points in the current scan and local SO-Map (or the previous scan) with a good

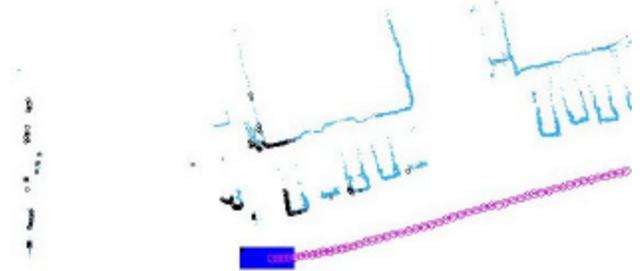


Fig. 6: The SLAM result of Fig. 2.

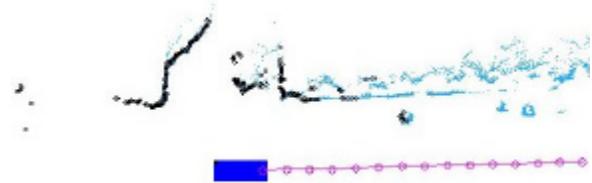


Fig. 7: The SLAM result of Fig. 3.

initial guess of their relative pose, and then solving the point-to-point least-squares problem to compute their relative pose. Finally the relative pose is updated and the whole process iterates until the result is satisfying.

Since ICP introduced by Chen and Medioni [1], many variants have been proposed on the basic ICP concept. Rusinkiewicz and Levoy [23] enumerated and classified many of these variants. Because the current collected data are two-dimensional and are not as complicated as three-dimensional data, we simply use Lu and Milios' iterative dual correspondence algorithm [16]. More efficient method such as Nene and Nayer's projection search technique [22] may be applied for 3D data in future work.

In our applications, the iteration algorithm terminates when the difference of the current and previous estimated translations is less than 5 cm. In practice, the number of iterations is usually less than 20. The results are integrated and saved in SO-Map, which is a grid-map composed by 5 cm x 5 cm cells. The value of the cell is the same as the times that the cell is occupied by stationary objects. The higher the value of the cell is, the higher possibility the cell belongs to a stationary object.

The results of the algorithm are shown in Fig. 6 and 7. The small black circles are the current scan data. The light blue dots belonging to the previous scans are stored in SO-Map. The magenta line is the path of the test vehicle and the small magenta circles are the estimated positions of the vehicle in previous scans. Several parked cars are in front of buildings in Fig. 6. One reason that causes the fuzzy area in the right side of Fig. 7 is a slope, which is not a vertical object. Other reason is that the test vehicle had some pitch motion. This approach successfully registered all kinds of range images in our experiment data.

The main disadvantage of P2P methods is that a good initial guess is necessary. This disadvantage can be overcome easily by integrating with a simple vehicle model or other pose sensors. Although Feature-based approaches like F2F and P2F can't be adapted to all environments, the faster performance without initial pose guess is still attractive. It should be a good direction to combine advantages of both approaches.

## 4 Detection and Tracking of Moving Objects

Moving objects do affect the results of SLAM unless they are properly detected. Fig. 8 and 9 show the results of SLAM without DTMO and with DTMO. In this section, our approaches of DTMO are presented.



Fig. 8: SLAM without DTMO



Fig. 9: SLAM with DTMO

### 4.1 Detection of Moving Objects

Intuitively, any inconsistent part from SLAM should belong to moving objects. But the idea isn't totally correct. Fig. 10 shows two rules to detect moving objects.

#### Rule 1: Approaching Object Detection

From previous scans, we know some space is not occupied. If we find any object in this space, this object must be moving. In Fig. 10, object A must be a moving object.

#### Rule 2: Leaving Object Detection

In Fig. 10, we can't say that object B is a moving object. Object B may be a new stationary object since object C blocked the view of the sensor in previous scans. What we are sure is that object C is a moving object. Although we can't tell if object B is moving or not by registering only 2 scans, the previous information does help us to guess the characteristics of object B.

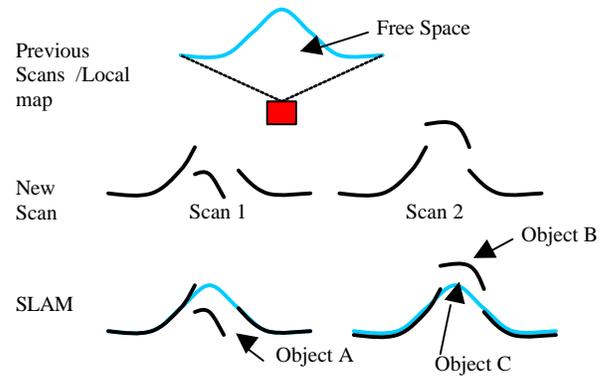


Fig. 10: Rules for Detection of Moving Objects

As mentioned in Section 2, SO-Map and MO-Map are used to store the previous results of SLAM with DTMO. Similar to SO-Map, MO-Map is also a grid map whose cell's size is 5 cm x 5 cm. The value of the cell in MO-Map is the same as the times that the cell is occupied by different moving objects.

The results of the moving object detection are shown in Fig. 11 and 12. The light blue dots and light orange dots respectively belong to SO-Map and MO-Map. The current scan contains black, red and green circles. The black circles mean stationary features. The green circles are new features since we don't have enough information to tell if they are moving or stationary. The red circles are moving objects. Our algorithm found both moving pedestrians and cars successfully.

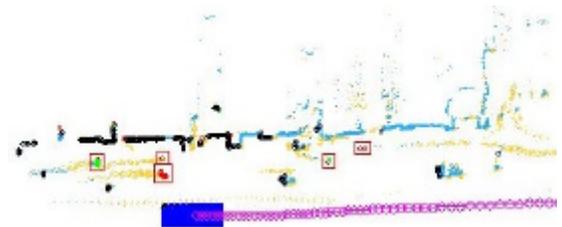


Fig. 11: Pedestrian Detection



Fig. 12: Vehicle Detection

### 4.2 Tracking of Moving Objects

In crowded urban environments, there are many kinds of moving objects, such as pedestrians, wheelchairs, bicycles, motorcycles, cars, buses, trucks, trailers, etc.

The main difficulties in tracking can be summarized as the following.

First, the moving objects have a wide range of sizes and velocities. Second, behaviors of moving objects are not always predictable. Third, moving objects' appearances can change significantly from scan to scan. Fourth, motion of the test vehicle changes suddenly. Also single object's observation may be grouped into several objects (See Fig. 13). And objects may disappear and reappear.

Because feature extraction of moving object is not robust and stable for tracking, we track moving objects by using the same algorithm in our SLAM to find the best match between frames. The idea is the same as map building of SLAM. Once the best match is found, information about this moving object is integrated. We call this method matching-based tracking. By integrating information among frames, we are able to overcome these difficulties. Since we don't presume any model to moving objects, we are able to solve the unpredictable problem. Fig. 9 shows the result of the matching-based tracking. We found that there are still some problems in this method. A modified approach is ongoing.

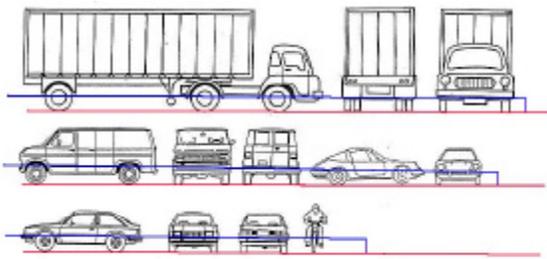


Figure 13: Dimensions of different moving objects. The red line indicates the ground; the blue line indicates the height of the scanner installed on the test vehicle. For trucks and trailers, only wheels are detected in this configuration. This means that trucks and trailers will be segmented into several groups.

The basic idea of the modified approach is that when a moving object is detected, we don't assign any model to it in the beginning. We assign an area for searching the best match. Once the information about this object is enough, we generate a model for this object. If the model generation is difficult or not stable, we only predict the future motion of this object. By motion prediction or model generation, we reduce the computational time for searching and matching and also improve the performance of matching-based tracking.

## 5 Properties of SO-Map and MO-Map

### 5.1 Accumulation and Retrieval of Information

Information can be easily accumulated and retrieved by maintaining SO-Map and MO-Map. If there are many moving objects passing through an area, any object that appears in this area should be recognized as a moving object without any rule. By integrating information from moving cars and pedestrians, we can even detect lanes and sidewalks. Fig. 14 shows that three lanes are detected.

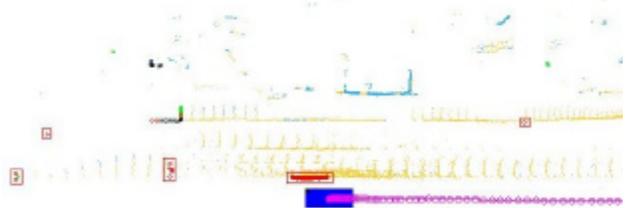


Fig. 14: Lane Detection (Fifth Ave. in Pittsburgh)

### 5.2 Consistency of SO-Map and MO-Map

SO-Map only contains stationary object information; therefore SO-Map is clean without any fuzzy area. In contrast, MO-map only contains moving object information. MO-Map is fuzzy. Any inconsistency of SO-Map and MO-Map provides important information.

Detection of very slow moving objects is difficult. The solution is to use any information we got as much as possible. If any fuzzy area is found in SO-Map, this area may contain slow moving objects.

Compared to vehicles, pedestrians are small objects represented by only several points. If some other points are misclassified as moving objects because of sensor noise and matching errors, these points will be misclassified as pedestrians if we only use the number of points in this moving group as the clue. Using the information of previous moving objects can make detection more reliable, reduce the effects of sensor noise and matching errors, and provide better classification of objects.

If an object is misclassified as a moving object in the beginning but the cells in MO-Map that belong to this object have high values later on, we can say that it is a stationary object.

## 6 Experimental Results

A SICK PLS 100 scanning laser rangefinder was mounted on the right side of our Navlab8 and Navlab11 vehicles, doing horizontal profiling. Navlab11 is equipped with wheel encoders to perform pose estimation. The vehicles were driven through the CMU campus and around nearby streets. The maximum speed of the vehicles was 45 mph and the range data were collected at 6Hz.

### 6.1 SLAM vs. SLAM with DTMO

As it can be seen from Fig. 8 and Fig. 9, SLAM with DTMO can get rid of moving objects and get a more consist mapping result.



Fig. 15: Navlab8, SICK PLS100, and an omni-directional camera.

### 6.2 Global Localization and Mapping

The accuracy of localization in our algorithms is measured by using the digital map shown in Fig. 16 as the ground truth. The result is shown in Fig. 17 with circles indicating intersections. No other pose information was used to get this result.

Since localization errors are cumulative, we need to develop algorithms that can register the local localization with the global. We did this by recognizing landmarks in the environment and correcting the pose of the test vehicle to the known global positions of the landmarks. Since intersections are a large part of the urban road infrastructure, an intersection detector would provide many opportunities for eliminating the cumulative error. The circles in Fig. 17 are intersections found by the current intersection detector. Only one intersection was missed. The reason for this missed intersection is that there are cars or pedestrian, which block the view of the intersection. Once the algorithm determines the intersection is missed, the next intersection will be set as the target for the intersection detector to look for. This is the major step in the construction of reliable global maps.



Fig. 16: Digital Map

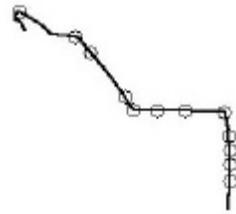


Fig.17: The path of the vehicle

### 6.3 Odometry vs. SLAM with DTMO

Fig. 19 shows the comparison of pose estimation from odometry and SLAM with DTMO. Line 1 is from odometry and Line 2 is from SLAM with DTMO. The line in Fig. 18 shows the route of this experiment. Compared to the ground truth in Fig. 18, the result of SLAM with DTMO is not perfect yet. But it is still much better than odometry. Also by intersection detection, we

can use some global consistent range scan alignment methods to improve SLAM with DTMO.



Fig. 18: Aerial Map



Fig. 19: The results

## 7 Conclusion and Future Work

To authors' knowledge, it's the first time to integrate SLAM and DTMO and to successfully demonstrate SLAM with DTMO at high speeds in large crowded city environments. Also a new matching-based tracking method was presented. Our algorithm found both moving pedestrians and cars successfully.

In future work, we are going to try several approaches to increase the efficiency of the algorithms. Global consistent range scan matching methods will be used to accomplish global localization and mapping. We hope to apply this technique to applications such as driving assistant system, 3D city-sized mapping, and research of dynamic social activities.

## Acknowledgements

The authors wish to thank John Kozar, Liang Zhao, Dave Duggins, Jay Gowdy, Teruko Yata, Hsiao-Ping Chang and Robert MacLachlan for their help with the experiments. This work is partially supported by Robert Bosch Corporation, by funding from PennDOT and by a subcontract from the DARPA MARS program through SAIC.

## References

- [1] Y. Chen and G. Medioni. Object Modeling by Registration of Multiple Range Images. *Proc. IEEE Int. Conf. On Robotics and Automation*, 1991.
- [2] H. Choset, K. Nagatani. Topological Simultaneous Localization and Mapping (SLAM): Toward Exact Localization Without Explicit Localization. *IEEE Transactions on Robotics and Automation*, Vol.17, No. 2, pp. 125-36, April, 2001.
- [3] I. J. Cox. Blanche--An experiment in guidance and navigation of an autonomous robot vehicle. *IEEE Transactions on Robotics and Automation*, Vol. 7, No. 2, pp. 193-204, April 1991.
- [4] J. L. Crowley, World modeling and position estimation for a mobile robot using ultrasonic ranging. *IEEE Int. Conf. on Robotics and Automation*, pp. 674-80, 1989.
- [5] L. Davis, V. Philomin and R. Duraiswami. Tracking humans from a moving platform. *15th Int. Conf. on Pattern Recognition*. Vol. 4, pp. 171-78, 2000.
- [6] M.C. Deans and M. Hebert. Invariant filtering for simultaneous localization and mapping. *IEEE Int. Conf. on Robotics and Automation*, Vol. 2, pp. 1042-47, 2000.
- [7] F. Dellaert, D. Pomerleau and C. Thorpe. Model-based car tracking integrated with a road-follower. *IEEE Int. Conf. on Robotics and Automation*, pp 1889-94, 1998.
- [8] J. Guivant, E. Nebot and S. Baiker. High accuracy navigation using laser range sensors in outdoor applications, *IEEE Int. Conf. on Robotics and Automation*, vol. 4, pp. 3817-22, 2000.

- [9] J.-S. Gutmann and C. Schlegel. AMOS: Comparison of scan matching approaches for self-localization in indoor environments, In *Proc. of the 1st Euro micro Workshop on Advanced Mobile Robots*, IEEE Computer Society Press, 1996.
- [10] J.-S. Gutmann, W. Burgard, D. Fox, and K. Konolige. An experimental comparison of localization methods. In *Proc. IROS'98. IEEE/RSJ*, 1998.
- [11] J.-S. Gutmann, T. Weigel and B. Nebel. Fast, accurate, and robust self-localization in polygonal environments. In *Robocup workshop (ABS1) at IJCAI-99*, 1999.
- [12] J. Gonzalez, A. Stentz and A. Ollero. A mobile robot iconic position estimator using a radial laser scanner. *J. of Intelligent and Robotic Systems*, vol.13, pp 161-79, 1995.
- [13] J. Gonzalez and R. Gutierrez. Direct motion estimation from a range scan sequence, *J. of Robotics Systems*, vol. 16(2), pp 73-80, 1999.
- [14] D.F. Huber, O. Carmichael, and M. Hebert. 3D map reconstruction from range data. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*, vol. 1, pp. 891-97, April, 2000.
- [15] K. Konolige and K. Chou. Markov localization using correlation. IJCAI, 1999.
- [16] J.J. Leonard and H.F. Durrant-Whyte. Simultaneous map building and localization for an autonomous mobile robot. *IEEE/RSJ Int. Workshop on Intelligent Robots and Systems*, vol.3, pp. 1442-47, 1991.
- [17] F. Lu and E. E. Milios. Robot pose estimation in unknown environments by matching 2D range scans. *Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, Seattle, WA, pp 935-938, June 1994.
- [18] F. Lu and E. E. Milios. Globally consistent range scan alignment for environment mapping. *Autonomous Robots*, vol.4, pp. 333-49, 1997.
- [19] R. Madhavan, M.W.M.G. Dissanayake, H.F. Durrant-Whyte. Autonomous underground navigation of an LHD using a combined ICP-EKF approach. *IEEE Int. Conf. on Robotics and Automation*, vol.4, pp 3703-08, 1998.
- [20] M. Hebert. Active and passive range sensing for robotics. In *Proc. of the IEEE Int. Conf. on Robotics and Automation*, Vol. 1, pp. 102 – 10, 2000.
- [21] C. Mertz, S. McNeil, C. Thorpe. Side Collision Warning Systems for Transit Buses. In *Proc. Of IEEE Intelligent Vehicle Symposium*, October, 2000.
- [22] S. A. Nene and S. K. Nayar. Closet Point Search in High Dimensions. *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, June 1996.
- [23] S. Rusinkiewicz and M. Levoy. Efficient Variants of the ICP algorithms. *Proc. of the Third Int. Conf. on 3-D Digital Imaging and Modeling*, pp. 145-52, May, 2001.
- [24] A. Scott, L. E. Parker and C. Touzet. Quantitative and qualitative comparison of three laser-range mapping algorithms using two types of laser scanner data. *IEEE Int. Conf. on Systems, Man, and Cybernetics*, vol.2, pp. 1422-27, 2000.
- [25] D.A. Simon, M. Hebert and T. Kanade. Real-time 3-D pose estimation using a high-speed range sensor. *IEEE Int. Conf. on Robotics and Automation*, vol.3, pp. 2235-41, 1994.
- [26] G. Shaffer, *Two-Dimensional Mapping of Expansive Unknown Areas*, PhD dissertation, Carnegie Mellon University, Pittsburgh, 1995.
- [27] S. Sukkarieh, H. Durrant-Whyte. Towards the Development of Simultaneous Localization and Map Building for an Unmanned Air Vehicle. In *Proc. Int. Conf. On Field and Service Robotics*, pp. 193-200, 2001.
- [28] S. Thrun, D. Fox, W. Burgard, and F. Dellaert. Robust Monte Carlo Localization for Mobile Robots. *Artificial Intelligence J.*, 2001.
- [29] S.B. Williams, P. Newman, G. Dissanayake and H. Durrant-Whyte. Autonomous underwater simultaneous localisation and map building. *IEEE Int. Conf. on Robotics and Automation*, vol. 2, pp. 1793-98, 2000.
- [30] Z. Zhang. Iterative Point Matching for Registration of Free-Form Curves and Surfaces. *Int. J. of Computer Vision*, 12(2), pp. 119-52, 1994.
- [31] H. Zhao and R. Shibasaki. Reconstructing Urban 3D Model using Vehicle-borne Laser Range Scanners. *Proc. of the Third Int. Conf. on 3-D Digital Imaging and Modeling*, pp. 349-56, May, 2001.
- [32] L. Zhao and C. Thorpe. Qualitative and Quantitative Car Tracking from a Range Image Sequence. *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 496-501, 1998.