TOWARDS ROBUST LOCALIZATION IN HIGHLY DYNAMIC ENVIRONMENTS

Shao-Wen Yang

Proposal for Doctoral Thesis

Department of Computer Science and Information Engineering
College of Electrical Engineering and Computer Science
National Taiwan University
Taipei, 10617 Taiwan

January 2009

Submitted in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy

Thesis Committee:
Chieh-Chih Wang (Chair)
Li-Chen Fu
Jane Yung-Jen Hsu
Han-Pang Huang
Ta-Te Lin
John J. Leonard, MIT

© SHAO-WEN YANG, MMIX
ABSTRACT

Localization in urban environments is a key prerequisite for making a robot truly autonomous, as well as an important issue in collective and cooperative robotics. It is not easily achievable when moving objects are involved or environment changes. Ego-motion estimation is the problem of determining the pose of a robot relative to its previous location without an absolute frame of reference. Mobile robot localization is the problem of determining the pose of a robot relative to a given map of the environment. The performance of ego-motion estimation completely depends on the consistency between sensor information at successive time steps, whereas the performance of global localization highly depends on the consistency between the sensor information and the a priori environment knowledge. The inconsistencies make a robot unable to robustly localize itself in real environments. Explicitly taking into account the inconsistencies serves as the basis for mobile robot localization.

In this thesis, we explore the problem of mobile robot localization in highly dynamic environments. We proposed a multiple-model approach to solve the problems of ego-motion estimation and moving object detection jointly in a random sample consensus (RANSAC) paradigm. We show that accurate identification of static environments can help classification of moving objects, whereas discrimination of moving objects also yields better ego-motion estimation, particularly in environments containing a significant percentage of moving objects.

It is believed that a solution to the moving object detection problem can provide a bridge between the simultaneous localization and mapping (SLAM) and the detection and tracking of moving objects (DATMO) problems. Based on the ego-motion estimation framework, to provide reliable moving object detection, data association can still be problematic due to merge and split of objects and temporal occlusion. We propose the use of discriminative models to reason about the joint association between measurements. Scaling such a system to solve the global localization problem will increase the reliability for mobile robots to perform autonomous tasks in crowded urban scenes. We propose to use a multiple-model approach based on the probabilistic mobile robot localization framework and formulate an extension to the global localization problem. Besides, detecting objects of small sizes at low speeds, such as pedestrians, is difficult, but of particular interest in mobile robotics. We propose the use of prior knowledge from the mobile robot localization framework to deal with the problem of pedestrian detection, and formalize the localization-by-detection and detection-by-localization framework. The proposed approach will be demonstrated using experimental testing with real data.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>ix</td>
</tr>
<tr>
<td><strong>CHAPTER 1. Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1. Problem Formulation</td>
<td>2</td>
</tr>
<tr>
<td>1.2. Proposed Solution</td>
<td>4</td>
</tr>
<tr>
<td>1.3. Thesis Statement</td>
<td>5</td>
</tr>
<tr>
<td>1.4. Document Outline</td>
<td>6</td>
</tr>
<tr>
<td><strong>CHAPTER 2. Related Work</strong></td>
<td>7</td>
</tr>
<tr>
<td>2.1. Dealing with Laser Scanner Failure</td>
<td>7</td>
</tr>
<tr>
<td>2.2. Ego-motion Estimation</td>
<td>11</td>
</tr>
<tr>
<td>2.3. Global Localization</td>
<td>14</td>
</tr>
<tr>
<td><strong>CHAPTER 3. Work to Date</strong></td>
<td>17</td>
</tr>
<tr>
<td>3.1. Ego-motion Estimation</td>
<td>17</td>
</tr>
<tr>
<td>3.1.1. Random Sample Consensus</td>
<td>18</td>
</tr>
<tr>
<td>3.1.2. RANSAC-based Segment Matching</td>
<td>19</td>
</tr>
<tr>
<td>3.1.3. Multiple-Model RANSAC</td>
<td>24</td>
</tr>
<tr>
<td>3.1.4. Experimental Results</td>
<td>26</td>
</tr>
<tr>
<td>3.1.5. Discussion</td>
<td>31</td>
</tr>
<tr>
<td>3.2. Global Localization</td>
<td>31</td>
</tr>
<tr>
<td><strong>CHAPTER 4. Proposed Work</strong></td>
<td>35</td>
</tr>
<tr>
<td>4.1. Increasing Robustness</td>
<td>35</td>
</tr>
<tr>
<td>4.1.1. Out-of-plane Motion Detection</td>
<td>35</td>
</tr>
<tr>
<td>4.1.2. Joint Data Association</td>
<td>37</td>
</tr>
<tr>
<td>4.2. Scaling to Global Localization</td>
<td>37</td>
</tr>
<tr>
<td>4.3. Pedestrian Detection</td>
<td>38</td>
</tr>
<tr>
<td>4.4. Timeline</td>
<td>39</td>
</tr>
<tr>
<td><strong>CHAPTER 5. Expected Contributions</strong></td>
<td>41</td>
</tr>
<tr>
<td><strong>BIBLIOGRAPHY</strong></td>
<td>43</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

1.1 Inconsistency in localization .................................................. 3
1.2 System overview ................................................................. 5

2.1 Occupancy grid maps in environments with mirrors and windows ...... 10
2.2 Mapping in the environment with mirrors .................................. 11
2.3 Mapping in the environment with mirrors and windows .................. 11
2.4 Mapping in the environment with a mirror pillar .......................... 12

3.1 Ego-motion estimation results in highly dynamic environments ........ 22
3.2 Ego-motion estimation results in highly dynamic environments ......... 23
3.3 Moving object detection results using MM-RANSAC ....................... 28
3.4 Moving object detection results using MM-RANSAC ....................... 29
3.5 Moving object detection results using MM-RANSAC ....................... 30
LIST OF TABLES

3.1 Quantitative comparisons between RANSAC and MM-RANSAC . . . . . . . . 31
THE simultaneous localization and mapping (SLAM) problem asks if it is possible for a mobile robot to build a consistent map of the environment and at the same time determine its location within this map (Durrant-Whyte, 1988; Smith et al., 1990). The solution to the SLAM problem has been seen as the fundamental in making a robot truly autonomous (Thrun, 2002; Durrant-Whyte & Bailey, 2006). Most researchers on SLAM assume that the unknown environment is static containing only rigid, stationary objects. Non-rigid or moving objects are processed as outliers and filtered out.

The detection and tracking of moving objects (DATMO) problem has also been extensively studied for several decades (Bar-Shalom et al., 2002). In surveillance applications, even though the sensors are mounted on stationary platforms, changes of the environment still make the DATMO problem difficult. Solving the DATMO problem in urban environments from a moving vehicle is much harder. One of the most important, yet difficult, issues of the DATMO problem is to discriminate moving objects from stationary objects. A common approach is to identify moving objects from the portion of an observation that differs significantly from a reference model. Background subtraction is a widely used approach for detecting moving objects from static sensors by comparing each observation to a predetermined model of the scene background (Ohta, 2001; Cucchiara et al., 2003).

There exist almost no static environments. Even in office environments, furniture is moved, doors are opened and closed, etc. The solution to the moving object detection problem provides a bridge between the SLAM and the DATMO problems (Wang, 2004). However, moving object detection from moving sensors is much more difficult from both
CHAPTER 1. INTRODUCTION

theoretical and practical perspectives. Theoretically, the wide variety of objects in urban scenes make appearance-based approaches infeasible. Obtaining an a priori background model is necessary for robust moving object detection. Determining a background model, a stationary object map, in highly dynamic environments is usually intractable for online application. Practically, there are many challenges in developing a good moving object detection algorithm as well as an ego-motion estimator. First, it must be robust against changes in robot pose. Second, it should avoid detecting stationary objects which are partially or even fully occluded. Finally, its internal model should be capable of tackling environments dominated by moving objects, especially for robots at high speeds.

Both the SLAM problem and the DATMO problem has been widely studied. The performance of a SLAM algorithm can be improved if moving object information is filtered out. On the contrary, the performance of a DATMO algorithm can be more robust if the environment map is available and the robot is capable of determining its location using the map. For robots acting in crowded urban environments containing a variety of objects, solving the SLAM problem and the DATMO problem concurrently is essential, particularly for safe driving (Wang et al., 2003). Aside from the moving object detection problem, modeling objects of light-refracted, light-reflected, light-absorbed, and transparent materials, which can be problematic in laser sensing, is also of essence for mobile robotics (Dioso & Kleeman, 2004; Yang & Wang, 2008).

1.1. Problem Formulation

Mobile robot localization is the problem of determining a robot’s pose, location and orientation, from sensor data, given an a priori environment map. Localization has been dubbed “the most fundamental problem to providing a mobile robot with autonomous capabilities” by Cox (1991). Given an accurate stationary object map, the probabilistic localization algorithms can perform nicely if the map and the environment are consistent (Fox et al., 1999; Olson, 2000; Thrun et al., 2000). However, it is not easily achievable when moving objects are involved or environment changes, especially in urban environments. Ego-motion estimation performance completely depends on the consistency between sensor information at successive time steps, whereas global localization performance highly depends on the consistency between the sensor information and the a priori environment knowledge. Besides, there are still sorts of objects which are stationary but can make the
map inconsistent with the environment in optical sensing – light-refracted, light-reflected, light-absorbed, and transparent materials (Dioso & Kleeman, 2004; Wang, 2004; Yang & Wang, 2008; Petrovskaya & Thrun, 2008). The variety of inconsistencies, as depicted in Figure 1.1, make a robot unable to robustly localize itself in real environments. Explicitly taking into account the inconsistencies serves as the basis for robust mobile robot localization.

Future robots will be required to act autonomously in environments where people are involved. Robotic vehicles should be capable of autonomous driving or driver assistance. Service robots are asked to interact with people in a variety of environments where people are usually moving. Detecting and handling changes of environments are essential for the successful achievement of autonomous tasks (Wang & Thorpe, 2002). The DARPA Urban Challenge also aims at dealing with dynamic urban scenes. Maneuvers that were specifically required for the Urban Challenge included merging into fast-moving traffic, left turns across oncoming traffic and the execution of U-turns in situations in which a road is completely blocked. These tasks are about to deal with environments that change or contain non-static entities. Boss (Urmson et al., 2008), Junior (Montemerlo et al., 2008), and Talos (Leonard et al., 2008), the autonomous vehicles in the DARPA Urban Challenge,
achieved these tasks by analyzing dense 3D scans from a number of costly 2D and 3D laser scanners for situational awareness in urban scenes.

1.2. Proposed Solution

The thesis explores the problem of mobile robot localization in highly dynamic environments using planar range scans. The robustness of the solution is considered in developing an ego-motion estimator capable of representing multiple-model distributions of stationary and moving entities. Extensions are proposed to tackle the problem of global localization in highly dynamic environments.

The problem of ego-motion estimation, as well as moving object detection, is first addressed by formulating the problem in a random sample consensus (RANSAC) paradigm. We proposed a RANSAC-based ego-motion estimator to deal with highly dynamic environments using one planar laser scanner. Instead of directly sampling on individual measurements, the RANSAC process performs at a high level abstraction for systematic sampling and computational efficiency. Moving objects are successfully detected without incorporating any grid maps, that are inherently time and space consuming. Multiple-model RANSAC (MM-RANSAC) is introduced to simultaneously deal with multiple models – a static environment model for ego-motion estimation and a moving object model for moving object detection. We show that accurate identification of static environments can help classification of moving objects, whereas discrimination of moving objects also yields better ego-motion estimation, particularly in environments containing a significant percentage of moving objects. The SLAM with DATMO framework derived by Wang et al. (2003) assumes that an observation can be decomposed into measurements of stationary and moving objects. The proposed work also aims at providing a theoretical foundation for the problem of simultaneous localization, mapping, and moving object tracking (SLAMMOT) (Wang, 2004; Wang et al., 2007), as illustrated in Figure 1.2.

Furthermore, in order to adequately address the problem of global localization in highly dynamic environments, we propose to extend the multiple-model approach to tackle these difficulties to obtain more accurate global estimates by incorporating negative information. Negative information pertains to the absence of a feature (Thrun et al., 2005, Chap. 7.8) and thus can be exploited when an object is supposed to but does not actually
1.3 Thesis Statement

Localization is of essence for mobile robots to act autonomously in real environments. Detecting and handling changes of environments are of most importance for the successful achievement of tasks in environments where people involved. Furthermore, negative information can be exploited for mobile robot localization in highly dynamic environments in which lackness of stationary landmarks can make a robot unable to determine its location with respect to a predetermined environment map. The thesis seeks to answer the following question:

Can a robot effectively improve its capability of self-localization by explicitly taking into account negative information – what the robot is supposed to see but actually does not – in the presence of moving objects in highly dynamic environments?
1.4. Document Outline

The proposal document is organized as follows. First, in Chapter 2, a literature survey of the related work in the fields of ego-motion estimation and mobile robot localization in highly dynamic environments is presented. In Chapter 3, we discuss the thesis work completed to date. Algorithms for ego-motion estimation and moving object detection in highly dynamic environments are presented as a preliminary step toward developing a formal global localization algorithm. In Chapter 4, the proposed work necessary to complete this thesis is described in which the potential research directions for localization in highly dynamic environments is explored. Finally, Chapter 5 presents a summary of the expected contributions of this thesis upon its completion.
CHAPTER 2

Related Work

This thesis draws on the work from three major bodies of study – dealing with laser scanner failure, ego-motion estimation, and global localization – which are about to deal with three sorts of inconsistencies respectively – the inconsistency between the map and the environment, the inconsistency between observations, and the inconsistency between observations and the map. The chapter provides an overview and examines the state-of-the-art approaches. Through a comparison of previous methods with those proposed in this thesis, we will show that no prior work provides a comprehensive framework necessary to solve the problem of localization in highly dynamic environments.

2.1. Dealing with Laser Scanner Failure

Mirrors and glasses are quite common objects that appear in our daily lives. While laser scanners play an important role nowadays in the field of robotics, there is very few literature addresses the related issues such as mirror reflection and glass transparency. As light can be reflected off the mirrors and penetrates the windows, mobile robots equipped with only laser scanners are not capable of dealing with real environments. It is straightforward to deal with mirrors and windows by fusing sensors of heterogeneous characteristics (Dioso & Kleeman, 2004). This should be fine for mobile robots to perform mapping and localization tasks simultaneously. However, indistinguishability between mirror images and true objects makes the map inconsistent with the true environment. The planning and control of a robot can still be problematic without explicitly modeling these kinds of objects. We also proposed a unified framework to tackle the problem of mirror reflection which
can be seamlessly integrated into the state-of-the-art mobile robot localization framework (Yang & Wang, 2008).

While there are a number of sensors that could be used, laser scanners are currently the standard sensors, which are also somehow of most reliability, for both indoor and outdoor environments. The data from a laser scanner include the angle and the distance to objects in the field of view. In comparison, vision sensors require complicated and error-prone processing before obtaining the depth information and range sensors such as sonar sensors and infrared sensors are not capable of fine angular resolution or realtime behavior. A laser scanner is capable of fine angular and distance resolution, realtime data retrieval, and low false rates.

As the light can be reflected off the mirrors and penetrate the windows, mobile robots equipped with only laser scanners might not be capable to deal with real environments. The sonar, oppositely, is capable of detecting those objects that the laser scanner fails. The main drawbacks in sonar sensing are specularity, wide beam width, and frequent misreadings due to either external ultrasound sources or crosstalk (Jörg, 1995). The sonar aids the laser scanner by detecting mirrors and windows not seen by the laser scanner, while the laser scanner aids the sonar by retrieving more detailed environment information.

A couple of new laser scanners are introduced recently. SwissRanger SR-3000 (Oggier et al., 2005) is a 3D laser scanner which can provide immediate depth images. It enables a diverse set of emerging medical and biometric applications, as well as robotics. Hokuyo URG-04LX is a laser scanner for indoor use which comes with a reasonable price and low power consumption. It provides data at high data rates with roughly millimeter resolution. SICK also conducts new models S200/S300 and LMS100 which are smaller than LMS291 with an affordable price. Konolige et al. (2008) proposed a low-cost laser distance sensor with reasonable accuracy and a very low price. The build cost of the device could be under $30. As the development of laser scanners is getting more and more mature, prices are greatly reduced nowadays. Robots also rely more and more on laser sensing. Nevertheless, laser scanners are versatile, but never omnipotent. As a matter of fact, the aforementioned new laser scanners also suffer from the problems of mirror reflection and glass transparency.

To make robots fully autonomous in environments with mirrors and windows, detection and modeling of these objects are critical. Jörg (1995) proposed to use laser scanner
measurements to filter out spurious sonar measurements. The objective is the extraction of sonar range readings which are complementary to corresponding laser range information in the sense that they provide additional environmental information. The laser scanner information is used to verify corresponding sonar range information. Extracting of complementary sonar information is accomplished by applying an ideal sonar sensor model to compute a hypothetical sonar scan from the laser scan. The superposition of both a hypothetical sonar scan and its corresponding real sonar scan allows to identify those real sonar range readings which come along with additional environmental information. The spatial information is counted and stored in an accumulated grid map. Dudek et al. (1996) introduced an approach to extract line segments in laser scans and sonar readings. The robot is commanded to rotate in place. A collection of sonar measurements is acquired to obtain a dense range map. The laser sensing is used to complement the sonar sensing by accurately pinpoint the corners and the borders of objects, where the sonar data is ambiguous. The exploration strategy for the fusion of sonar and laser data is a form of wall following. The world is modeled in terms of line segments with marked endpoints. Both Jörg (1995) and Dudek et al. (1996) proposed to extract complementary sonar readings to detect those objects not seen by laser scanners. However, the indistinguishability between mirrors and windows still makes mobile robot exploration problematic.

To visualize the indistinguishability in sensor fusion, we implemented a sensor fusion algorithm (Yang & Wang, 2008) using the data collected in an environment with mirrors and windows. In Figure 2.1, the upper rectangle indicates the mirror location and the lower rectangle shows the location of a French window. Figure 2.1(a) and 2.1(b) depict the occupancy grid maps (Elfes, 1989) built using data from a laser scanner and sonar sensors, respectively. It is observed that mirrors and windows are objects which are probable to be seen by the sonar sensor, but less likely by the laser scanner. Figure 2.1(c) shows the sensor fusion map. It is clear that most of the mirror and window locations are successfully identified in comparison with the laser scanner map. With fusion of heterogeneous sensors, navigation of mobile robots can be collision-free.

However, in the sensor fusion map, mirrors and windows, as potential obstacles, make no difference. This should be fine for mobile robots to perform mapping and localization tasks simultaneously. Indistinguishability between mirror images and true objects does make the map inconsistent with the true environment. In Figure 2.1, rectangles show
the potential obstacles not seen by a laser scanner which are identified with the use of the sensor fusion map. Robots still cannot know whether the area behind a potential obstacle is real or not.

We addressed the problem of mirror reflection and solved the problem by accumulating the positive information from laser scanner and eliminate the fake counterpart by using the property of mirror symmetry (Yang & Wang, 2008). To our best knowledge, the solution to the problem of mirrors and windows has not yet been addressed before. The method does not require an additional heterogeneous sensor such as sonar or infrared and explicitly models the phenomenon of mirror reflection to identify potential mirror locations. Furthermore, the proposed method is able to be integrated seamlessly into the occupancy grid map representation and the mobile robot localization framework.

Figures 2.2, 2.3 and 2.4 illustrate the process of the proposed approach. The occupancy grid maps of the environment are shown, where the rectangles filled with blue indicate the robot locations, the lines in red are the line models of the mirrors, the ellipses in green show the $2\sigma$ covariances of the line models, and the thick lines in red indicate the mirror locations. In Figures 2.2(a), 2.3(a) and 2.4(a), the maps before post-processing are presented. In Figures 2.2(b), 2.3(b) and 2.4(b), the resulting maps after post-processing are shown where the grid cells located behind mirrors are corrected. The false estimates are also successfully eliminated. The less certain mirror estimates are discarded in the process and the post-processing process only take into account the confident mirror estimates.
2.2. Ego-motion Estimation

Ego-motion estimation is the process by which a mobile robot can recover its motion, or location relative to a previous time step, from sensor data. Schulz et al. (2003)
and Montemerlo et al. (2002) proposed to localize a robot and track dynamic objects using a previously generated map. This is reliable when robots act in environments that do not change. Moving object detection can be done by taking the differential of each scan with the environment map. However, maps are usually not affordable, particularly in urban scenes in which environments change as time goes by. Wang et al. (2003) applied a motion-based moving object detector in a divide-and-conquer manner. Spatio-temporal information is accumulated using a stationary object map and a moving object map (Wang et al., 2007). Inconsistent parts are divide into two categories: approaching and leaving. A stationary object map and a moving object map are accumulated and used for moving object detection. Zhao et al. (2008) applied a delayed mapping and tracking approach in which geometric features are employed. Mendes et al. (2004) proposed a voting scheme in which geometric features including shape, size, and other geometric properties are also used.

In the robotics literature, the last decade has seen more and more researchers take moving object information into account and solve the simultaneous localization and mapping (SLAM) and the detection and tracking of moving objects (DATMO) problems concurrently. Wang et al. (2003) proposed a consistency-based moving object detector and provided a joint framework to solve the SLAM and the DATMO problems simultaneously.

![Figure 2.4: Mapping in the environment with a mirror pillar](image)
The multiple hypothesis tracking (MHT) method is applied to accomplish data association among moving objects. Bibby & Reid (2007) proposed a method that combines sliding window optimization and least-squares together with offline expectation maximization (EM) to do reversible model selection and data association that allows dynamic objects to be included directly into the SLAM estimate. Zhao et al. (2008) utilizes GPS data and control inputs to achieve global consistency in dynamic environments.

In the computer vision literature, random sample consensus (RANSAC), which has been widely used for outlier rejection (Fischler & Bolles, 1981), is one of the most effective algorithm for model fitting to data containing a significant percentage of gross errors. RANSAC is a robust multiple hypothesis estimator in the presence of many outliers. A theoretical investigation of scoring under a simple inlier-outlier model is performed to discriminate outliers from the inlier model. Nistér (2003) proposed an ego-motion estimator for perspective cameras in a rigid scene. Some structure of the scene is also estimated which is highly related to structure from motion (SFM). Sharma et al. (2006) proposed the use of RANSAC for change detection in remote sensing. The transformation in the dynamic range of the images is estimated and classify pixels not satisfying this shift as changes. Both of the works reject moving objects as outliers which are inconsistent with the inlier model. RANSAC is advanced in its effectiveness and efficiency, but unable to extract multiple models due to its exclusivity. We introduce a multiple-model extension for RANSAC to deal with highly dynamic environments which enables RANSAC to fit data of multiple models at the same time.

The proposed work in this thesis will address the challenges encountered by the current state-of-the-art localization algorithms, including dealing with environments containing a significant percentage of moving objects. Ego-motion estimation is necessary for mobile robots to ensure that estimates of moving objects remain bounded as they explore the environment. The solution to the moving object detection problem provides a bridge between the SLAM and the DATMO problems. Establishing the spatial and temporal relationships among the robot, stationary objects and moving objects in the environment serves as the basic for scene understanding.
CHAPTER 2. RELATED WORK

2.3. Global Localization

Probabilistic localization techniques have been developed for decades and demonstrated to be a robust approach to solve the problem of mobile robot localization. In global localization, the initial pose of the robot is unknown. The robot is initially placed somewhere in its environment, but lacks knowledge of its whereabouts. Thrun et al. (2005, Chap. 7.1) gives a comprehensive review of the taxonomy dividing localization problems along a number of important dimensions.

Leonard & Durrant-Whyte (1991) proposed the use of extended Kalman filters (EKFs) for feature-based localization through artificial beacons. However, EKF localization is infeasible for global localization due to its uni-modality. Roumeliotis & Bekey (2000) showed that multiple-hypothesis EKFs are capable of global localization. Fox et al. (1999) proposed a version of Markov localization which provides accurate position estimates and is tailored towards dynamic environments. The key idea of Markov localization is to maintain a probability density over the space of all locations of a robot in its environment. Markov localization comes at the expense of high computational cost and thus is improper for applications in large-scale, dynamic environments. Fox et al. (1999) developed Monte Carlo localization (MCL) using the condensation algorithm (Isard & Blake, 1998), also known as particle filters, which is motivated by adding random samples. Both Markov localization and MCL are applicable to the global localization problem and have already become the most popular localization algorithms.

One of the key challenges in the context of probabilistic localization lies in the design of the motion model and the observation model. Specifically, the motion model is a function specifying the probability of a robot pose given a control input and the robot’s previous location, and the observation model is a function specifying the likelihood of an observation given the robot pose in a given map. In particular, the proper design of the observation model is essential for probabilistic localization. Too optimistic observation models make the robot estimate overly confident, whereas too conservative observation models yield a high uncertainty in the robot pose (Dellaert et al., 1999). Recently, sophisticated observation models have been developed for probabilistic approaches to mobile robot localization. Thrun (2001) proposed the use of multi-modal likelihood functions to deal with discontinuities in the map. Pfaff et al. (2008) developed a place-dependent observation model using
Gaussian mixture models to cope with cluttered environments in which the likelihood of an observation depends on the location of the observer, or the robot.

We propose to extend the proposed multiple-model approach to deal with the problem of global localization in highly dynamic environments. Apart from the place-dependent characteristic of an observation, there are still other variants ignored in the observation model. We plan to formulate a motion-dependent extension of the observation model based on the occupancy grid map representation. The likelihood of an observation depends on not only the location of the robot, but also the location and the motion of the observed object. We seek to model possible multi-modalities in the observation model to improve the robustness of probabilistic localization algorithms, especially in environments that are dominated by moving objects.
CHAPTER 3

Work to Date

Robust localization in urban environments is not easily achievable as there are two motions involved: the motions of moving objects and the motion of the robot itself. In this thesis, the problems of ego-motion estimation, moving object detection and global localization are explored. We proposed a random sample consensus (RANSAC) based ego-motion estimator to deal with highly dynamic environments using one planar laser scanner. Instead of directly sampling on individual measurements, the RANSAC process performs at a high level abstraction for systematic sampling and computational efficiency. We proposed a multiple-model approach to solve the problems of ego-motion estimation and moving object detection jointly in a RANSAC paradigm. While the proposed multiple-model RANSAC (MM-RANSAC) approach offers a solution to the ego-motion estimation problem, we plan to extend the multiple-model approach to solve the global localization problem in highly dynamic environments.

3.1. Ego-motion Estimation

The solutions to the simultaneous localization and mapping (SLAM) and the detection and tracking of moving objects (DATMO) problems are known to be at the core for mobile robots to act autonomously in real environments. Moving object detection serves as the foundation for solving the SLAM and the DATMO problems simultaneously (Wang et al., 2003). We developed an online ego-motion estimation algorithm in a RANSAC paradigm using planar laser scans. It is applied for robustly matching successive observations obtained by a laser scanner. The RANSAC paradigm which is capable of smoothing data containing a significant percentage of gross errors is particularly applicable to scene analysis (Fischler & Bolles, 1981). The RANSAC process is applied at the segment level
CHAPTER 3. WORK TO DATE

in which sampling of consensus sets is performed systematically and the computational complexity can also be reduced. However, due to the exclusivity, RANSAC is not able to present multiple models for ego-motion estimation and moving object detection at the same time. A multiple-model extension is thus introduced for RANSAC to fit multiple models at the same time. Moving object information is extracted and seamlessly integrated into the RANSAC process such that the robustness of ego-motion estimation can be considerably improved, particularly in highly dynamic environments where surroundings of robots are dominated by non-stationary objects. The proposed algorithm does not employ any geometric features which are often environment dependent. It is also a non-delayed algorithm without incorporating any grid maps, that are inherently time and space consuming. Experimental results show that our algorithm works robustly in highly dynamic environments even when more than 50% of the environment seen by the robot are dynamic.

3.1.1. Random Sample Consensus

In this section, we review the foundation and probabilistic formulation of RANSAC. Classical techniques for parameter estimation optimize the fit of a functional description to all of the presented data. The RANSAC procedure is opposite to that of conventional smoothing technique. Rather than using as much of the data as possible to obtain an initial solution and then attempting to eliminate the invalid data points, RANSAC uses as small an initial data set as feasible and enlarges this set with consistent data when possible (Fischler & Bolles, 1981).

RANSAC uses the geometric distribution in statistics which models the discrete distribution – the probability distribution of the number $X$ of Bernoulli trials needed to get one success, supported on natural numbers $\mathbb{N}$. If the probability of success on each trial is $b$, then the probability that the $k$-th trial is the first success is

$$P_r(X = k) = (1 - b)^{k-1}b$$

(3.1)

$$= (1 - w^n)^{k-1}w^n$$

(3.2)

where $w$ is the probability that any selected data point is within the error tolerance of the model, and $n$ is the number of good data points required to determine the model, for all $k \in \mathbb{N}$. If we want to ensure with probability $p$ that at least one of the random selections is
an error-free set of \( n \) data points, we must expect to make \( k \) selections, where

\[
(1 - b)^k \leq (1 - p), \quad (3.3)
\]

\[
k \geq \log(1 - p) / \log(1 - b), \quad (3.4)
\]

RANSAC is effective for model fitting, particularly when a significant percentage of data are outliers. It is ideally suited for applications in range image analysis. The RANSAC formulation contains two remaining unspecific parameters \( n \) and \( w \) which are highly relevant to characteristics of data. Fischler & Bolles (1981) gives a detailed derivation and a comprehensive description for RANSAC. In the next section, we will derive the parameters for the ego-motion estimation problem where non-static objects can be rejected as outliers.

3.1.2. RANSAC-based Segment Matching

Ego-motion estimation can be performed using range image registration algorithms in the computer vision literature. To ensure against the possibility of the final consensus set being compatible with an incorrect model, the size of data points per selection should be greater than or equal to three for determining the pose transformation including translation and rotation. However, one of the most difficult problem in laser sensing is data association. Every single laser measurement is featureless. Researchers usually apply the closest point association rule to associate data points with unknown data association, such as the ICP algorithm (Besl & McKay, 1992; Lu & Milios, 1997).

3.1.2.1. Scan Segmentation. In the RANSAC paradigm, a number of random samples consisting of small sets are taken from the observation. A first attempt is to generate random samples directly from all measurements of an observation. Closest point association often yields good estimate for data containing a mass of points. However, registration performs very poorly on data containing few points and often results in ambiguity. Sampling directly on all measurements also requires comparatively large size of data points to preserve sufficient shape information for registration. For example, if \( w = 0.5 \) and \( n = 5 \), then \( b = 0.03125 \). To obtain a 99% assurance of making at least one error-free selection, by Equation 3.4, we have \( k \geq 146 \). It is time-consuming and computationally intractable for online applications, even though five measurements are still insufficient for obtaining a good registration result.
CHAPTER 3. WORK TO DATE

Instead of direct sampling on measurements, we propose to use a higher level data representation – a segment – to achieve reliable registration and online applications. The observation is segmented and further split into sets of measurements representing objects. Specifically, objects are extracted by segmenting the scan into densely sampled parts. Here, we use a distance criterion to segment measurements into objects. Although the distance thresholding method cannot produce perfect segmentation, it is plausible to use such distance gaps to find distinct objects and perform moving object detection. More precise segmentation can be accomplished using spatial and temporal information from the map or a multi-scale representation (Pfister & Burdick, 2006). However, experimental results show that the proposed algorithm works well even if the segmentation is not perfect.

3.1.2.2. Segment Matching. In the classical RANSAC paradigm, letting \( o \) be a feature and \( h \) be some hypothesis, the effectiveness of each \((o, h)\) is examined and represented using a binary score. Specifically, if \((o, h)\) is an inlier pair, the score \( s_h \) of the hypothesis \( h \) is incremented. As segments might be of significantly different sizes, a binary score is insufficient to describe the quantity of an association between two segments. Let \( z \) be an observation and \( z^i \) be the \( i \)-th segment in \( z \). Compared to the classical RANSAC process, the score \( s_h^i \) of each segment \( z^i \) is supported on \( \mathbb{N} \) and the effectiveness of the pair \((z^i, h)\) is represented by a natural number.

To build consensus sets, \( z \) is segmented and represented as a collection of segments \( z = \{z^i\}_i \). The system randomly permutes the segments firstly. A hypothesis \( h \) is generated from randomly selected \( n \) segments with probabilities proportional to the sizes \( |z^i| \) of the segments \( z^i \) by matching the \( n \) segments with the reference model \( \bar{z} \), which is the scan obtained at the previous time step.

To obtain the score \( s_h^i \) of a segment \( z^i \), the effectiveness of \((z^i, h)\) should be examined and check if \((y, h)\) is an inlier pair where \( y \in z^i \). The score \( s_h^i \) of a segment \( z^i \) is defined as the number of measurements \( y \in z^i \) which are located within neighborhoods of measurements in the reference model \( \bar{z} \). Here, \((y, h)\) is judged as an inlier pair if and only if the measurement \( y \) transformed to the global coordinate by the hypothesis \( h \) is located within a neighborhood \( d \) of some measurement in the reference model \( \bar{z} \). Specifically, the score \( s_h^i \)
is incremented if the pair \((y, h)\) is judged as an inlier pair. Therefore, we have

\[
s_h = \sum_{\{i | z_i \in z\}} s_{ih}^i \tag{3.5}
\]

\[
= \sum_{z^i \in z} \sum_{y \in z^i} 1_{h}(y) \tag{3.6}
\]

where \(1_{h}(y)\) is an indicator function indicating whether or not \((y, h)\) is an inlier pair. When the process finishes, a hypotheses with the highest score is output as the best transformation \(\psi\). In this work, \(d\) is 1.5 meter, which is the same as the segmentation threshold.

The parameter \(n\) should be carefully determined and take into account the tradeoff between efficiency and reliability, and the characteristics of the data. For matching segments with the reference model, one segment is usually sufficient to preserve the shape information of the environment unless an environment is composed of line segments which result in ambiguity. It is clear that the higher the value \(n\), the higher the probability at least one hypotheses is an inlier, and thus the reliability increases. Letting \(n = 2\) and \(w = 0.5\), according to Equation 3.4, to obtain a 99% assurance of making at least one error-free selection, the number \(k\) of selections must be greater than or equal to 17, which is computationally sufficient for online and realtime applications.

Figures 3.1 and 3.2 show that RANSAC outperforms ICP in urban environments. In Figures 3.1(a) and 3.2(a), it is clear that ICP converges to local minimums as the environments seen by the laser scanner are dominated by moving objects. Figures 3.1(b) and 3.2(b) demonstrate the effectiveness of the proposed RANSAC-based segment matching approach. The ego-motion estimates are more accurate as outliers are filtered out in the RANSAC process. The static parts of the environment are aligned nicely.

Here, we assume at least \(w = 50\%\) of the measurements from the laser scanner are stationary objects. However, in urban environments, it is often implausible to make this assumption. It is possible for a robot to be almost fully surrounded by moving objects. As illustrated in Figure 3.1(b), for robots in highly dynamic environments, RANSAC still fails. To obtain better ego-motion estimates, instead of filtering out moving objects as outliers, taking into account the moving object information is necessary. In the next section, we will propose a multiple-model extension for RANSAC to solve the problems of ego-motion estimation and moving object detection simultaneously.
CHAPTER 3. WORK TO DATE

![Graphs and images showing ego-motion estimation results in highly dynamic environments](image)

Figure 3.1. Ego-motion estimation results in highly dynamic environments
Figure 3.2. Ego-motion estimation results in highly dynamic environments
CHAPTER 3. WORK TO DATE

3.1.3. Multiple-Model RANSAC

RANSAC always finds the most consistent hypothesis and rejects inconsistent parts as outliers because of its exclusivity in nature. In environments where robots cannot collect sufficient static environment information, the output hypotheses can be far from the ground truth. To accommodate RANSAC to multiple models – a static environment model for ego-motion estimation and a moving object model for moving object detection, a compact representation models moving object information implicitly is proposed. It is seamlessly integrated into the RANSAC process. In addition, a segment classifier is introduced to discriminate moving objects from static environments. By introducing MM-RANSAC, ego-motion estimation can be performed robustly to highly dynamic urban scenes. The false positive rate can also be reduced as multiple models are fitted at the same time, rather than filtered out.

3.1.3.1. Data Representation. To represent multiple models simultaneously in a RANSAC paradigm, we construct a scan maintaining spatio-temporal information – a virtual scan. Virtual scans simplify data access by compressing moving object information into one single scan. Constructed in this manner, a virtual scan provides a compact representation of moving objects around the robot. A virtual scan is generated in accordance of the results from the segment classifier while the segments are also classified using the virtual scan. The proposed MM-RANSAC process alternates the two stages – segment classification and virtual scan generation – at each time step.

3.1.3.2. Segment Classification. To integrate virtual scans into the RANSAC process, segments of an observation are classified into three categories: static, unknown, and moving. To clarify, at each time step, there are three scans available, the observation $z$, the reference model $\bar{z}$, and the virtual scan $\tilde{z}$. The observation $z$ and the reference model $\bar{z}$ are the scans collected at the current time step and the previous time step, respectively. The virtual scan $\tilde{z}$ is generated in accordance of the reference model and the temporal information at the previous time step. Initially, the virtual scan is the same as the reference scan.

Each segments $z^i$ in the observation $z$ is associated with segments in both the reference model $\bar{z}$ and the virtual scan $\tilde{z}$. Let $\omega(z^i)$ and $\tilde{\omega}(z^i)$ be the percentage of measurements in
3.1 EGO-MOTION ESTIMATION

which are within neighborhoods of measurements in $\tilde{z}$ and $\tilde{z}$, respectively, and $\tilde{z}^i \in \tilde{z}$ and $\tilde{z}^i \in \tilde{z}$ be the associated segments of $z^i \in z$.

The classification of $z^i$ can be expressed as

$$\varphi(z^i) = \begin{cases} 
\text{moving} & \text{if } \tilde{\omega}(z^i) \geq \bar{\phi}, \varphi(\tilde{z}^i) = \text{moving} \\
& \text{or } \tilde{\omega}(z^i) < \bar{\phi}, \varphi(\tilde{z}^i) = \text{unknown} \\
\text{unknown} & \text{if } \tilde{\omega}(z^i) < \bar{\phi}, \varphi(\tilde{z}^i) = \text{static} \\
\text{static} & \text{if otherwise}
\end{cases} \quad (3.7)$$

where $\varphi(\tilde{z}^i)$ indicates the class of the associated segment $\tilde{z}^i$ of $z^i$, and $\bar{\phi}$ and $\tilde{\phi}$ are pre-defined parameters for determining the effectiveness of an association between segments, which is the only parameters should be chosen for the proposed ego-motion estimator, in addition to RANSAC parameters. Specifically, if a segment $z^i$ is static, it is probably associated with some segment in the reference model $\bar{z}$ in a relatively great proportion $\bar{\phi}$, unless it be either moving or occluded. In the case that $\tilde{\omega}(z^i)$ is less than some proportion, it is probably moving and firstly marked as unknown. Later on, consistency between the observation $z$ and the virtual scan $\tilde{z}$ is further verified. If the associated segment $\tilde{z}^i$ of a segment $z^i$ was classified as moving previously and $\tilde{\omega}(z^i)$ is greater than or equal to some proportion $\tilde{\phi}$, it is then classified as moving. As virtual scans maintain empirical moving object information, to have MM-RANSAC free of uncertainties of these estimates, $\tilde{\phi}$ should be far less than $\bar{\phi}$. In our implementation, the values of $\bar{\phi}$ and $\tilde{\phi}$ are 70% and 30%, respectively.

3.1.3.3. Virtual Scan Generation. To generate the virtual scan for ego-motion estimation in the upcoming time step, for each segment $z^i$ with $\varphi(z^i) \neq \text{moving}$, the virtual segment $\tilde{z}^i$ is the same as the segment $z^i$. Conversely, the transformation $\psi_i$ from $\bar{z}^i$ to $z^i$ is calculated by matching these two segments using the ICP algorithm. The linear and the angular velocities $\nu_i$ for each segment $z^i$ are estimated accordingly. Hence, assuming a constant velocity model, $z^i$ is further transformed with the estimated linear and angular velocities and assigned to $\tilde{z}^i$. The virtual scan generation process can be expressed as

$$\varphi(z^i) \neq \text{moving}, \forall y \in z^i \Rightarrow y = \tilde{y} \in \tilde{z}^i \quad (3.8)$$

$$\varphi(z^i) = \text{moving}, \forall y \in z^i \Rightarrow y + \nu_i \Delta t = \tilde{y} \in \tilde{z}^i \quad (3.9)$$

where $y$ is a measurement and $\nu_i$ is the estimated linear and angular velocities of the segment $z^i$. Then, we can obtain the virtual scan $\tilde{z} = \{\tilde{z}^i\}_i$ for the next time step. With the use of the virtual scan technique, motion modeling can be naturally integrated into the RANSAC process. MM-RANSAC builds consensus sets on the observation $z$ and scores
hypotheses with respect to the virtual scan $\tilde{z}$ in which multiple models are implicitly maintained.

In the MM-RANSAC process, the virtual scan $\tilde{z}$ is utilized, instead of using the reference model $\bar{z}$ directly. Figures 3.1(d) and 3.2(d) visualize the virtual scans. Comparing to RANSAC, we do not assume at least 50% of the measurements from the laser scanner are stationary objects anymore. The meaning of the parameter $w$ changes as virtual scans are involved, in which multiple models are fitted simultaneously. As virtual scans are employed and used by the MM-RANSAC process, $w = 0.5$ stands for at least 50% of the measurements of an observation are properly modeled, in which both stationary objects and moving objects are included. As a result, moving object information can help ego-motion estimation while multiple models are taken into account at the same time. It is particularly critical for mobile robots to act autonomously in highly dynamic environments.

Figures 3.1 and 3.2 show that MM-RANSAC outperforms RANSAC in urban scenes where environments are highly dynamic. In Figures 3.1(b) and 3.2(b), it is clear that the best hypotheses selected by RANSAC are still inconsistent with the real environments. The exclusivity of RANSAC make it unable to obtain good ego-motion estimates in such circumstances. Figures 3.1(c) and 3.2(c) demonstrate the superiority of MM-RANSAC which utilizes moving object information for ego-motion estimation. By modeling motions of moving objects implicitly with virtual scans, both the results of ego-motion estimation and moving object detection are much more accurate.

3.1.4. Experimental Results

The proposed approach is demonstrated using real data collected in a crowded urban scene. The average processing time of the proposed ego-motion estimator is 63ms, implemented using MATLAB, running on a desktop PC with Intel Core2 Quad CPU 2.40GHz and 4.0GB RAM, which is sufficient for realtime applications. Figures 3.1 and 3.2 depict the results of ego-motion estimation using ICP, RANSAC, and MM-RANSAC, respectively. Though RANSAC outperforms ICP, it fails when environments change significantly from scan to scan. As can be seen that MM-RANSAC which takes into account moving object information is robust to highly dynamic environments. Figures 3.1(d) and 3.2(d) depict the observations overlaid with the virtual scans for the MM-RANSAC results given in Figures 3.1(c) and 3.2(c), respectively. In these experiments, RANSAC and MM-RANSAC apply
3.1 EGO-MOTION ESTIMATION

...a common sampling stage at each time step, for a fair comparison. As addressed previously, the problem of occlusion is one of the most challenging problem in moving object detection. In Figure 3.2(c), it can be seen that the segment classifier of MM-RANSAC is robust to the occlusion problem, for which the false-positive rate is reduced, by observing the segments which are rejected as outliers due to occlusion in Figure 3.2(b) but classified as static by MM-RANSAC. A sequence of MM-RANSAC results are shown in Figures 3.3, 3.4, and 3.5. In these figures, grey dots are reference models, red dots are static objects, green dots are unknown objects, and blue dots and rectangles show moving objects. The ample experimental results show that the proposed algorithm performs robustly for ego-motion estimation and moving object detection in urban environments. The effectiveness in dealing with the occlusion problem is demonstrated as well. The issue of imperfect segmentation, as addressed in Section 3.1.2.1, is also presented here. In the MM-RANSAC paradigm, the problem is resolved naturally by the segment classifier. The values of $\bar{\phi}$ and $\tilde{\phi}$ are essential in presence of segmentation error, as described in Section 3.1.3.2.

Table 3.1 gives a quantitative comparison presenting the scores of the hypotheses applied by RANSAC and MM-RANSAC in Figures 3.1 and 3.2. Letting $\bar{\psi}$ and $\tilde{\psi}$ be the hypotheses output by RANSAC and MM-RANSAC, respectively, to demonstrate the robustness of MM-RANSAC, we show the scores of $(\bar{z}, \bar{\psi})$, $(\tilde{z}, \bar{\psi})$, $(\bar{z}, \tilde{\psi})$, and $(\tilde{z}, \tilde{\psi})$. Going through the table row-by-row, the reference model $\bar{z}$ contributes similar scores for both hypotheses. As a result, RANSAC cannot give the inlier hypothesis $\tilde{\psi}$ the highest score due to the presence of moving objects. Conversely, MM-RANSAC outperforms RANSAC as it fits multiple models simultaneously using the virtual scan $\bar{z}$. By introducing the virtual scan $\tilde{z}$, the score of the hypothesis $\tilde{\psi}$, which is more consistent with the environment, shows significant difference from other hypotheses. We also note that in our experiments there are 1750 out of 7554 time steps in which RANSAC and MM-RANSAC select different hypotheses. The travel distance of the data set is approximately 5 kilometer. RANSAC and MM-RANSAC apply the same hypotheses 76.83% of the time. On the other hand, MM-RANSAC select another hypothesis 23.17% of the time, in which the surroundings of the robot might be highly dynamic and RANSAC does not work well. While RANSAC and MM-RANSAC output different hypotheses, MM-RANSAC often provides better ego-motion estimates, as illustrated in Figures 3.1 and 3.2, and sometimes other hypotheses which are very close to that of RANSAC are output.
CHAPTER 3. WORK TO DATE

Figure 3.3. Moving object detection results using MM-RANSAC
3.1 EGO-MOTION ESTIMATION

Figure 3.4. Moving object detection results using MM-RANSAC
Figure 3.5. Moving object detection results using MM-RANSAC
3.2 GLOBAL LOCALIZATION

Table 3.1. Quantitative comparisons between RANSAC and MM-RANSAC

<table>
<thead>
<tr>
<th>Score</th>
<th>Scan 111</th>
<th>Scan 12366</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{\psi}$</td>
<td>191</td>
<td>153</td>
</tr>
<tr>
<td>$\tilde{\psi}$</td>
<td>181</td>
<td>147</td>
</tr>
<tr>
<td>$\bar{z}$</td>
<td>160</td>
<td>172</td>
</tr>
<tr>
<td>$\tilde{z}$</td>
<td>183</td>
<td>198</td>
</tr>
</tbody>
</table>

3.1.5. Discussion

We address the problem of ego-motion estimation in highly dynamic environments which is of essence for mobile robots to act autonomously in real environments. As moving object detection serves as the basic for solving the SLAM and the DATMO problems simultaneously, detecting and handling changes of environments are of most importance for the successful achievement of tasks in environments where people involved. Consensus sets are built at the segment level in which measurements can be sampled systematically to achieve reliable registration. A higher level data representation also make it feasible for realtime applications, especially for robots at high speeds. Though RANSAC is robust to data containing a significant percentage of outliers, it is still infeasible for data of multiple models due to its exclusivity. MM-RANSAC, a multiple-model extension of RANSAC, is thus introduced, in which the problems of ego-motion estimation and moving object detection are solved jointly in a RANSAC paradigm. The proposed algorithm does not employ any geometric properties which are unreliable in urban scenes. It is also a non-delayed algorithm without incorporating any grid maps which are inherently time and space consuming. The ample experimental results show that accurate identification of static environments can help classification of moving objects, whereas discrimination of moving objects also yields better ego-motion estimation. The feasibility and effectiveness of the proposed approaches have been demonstrated using data from Wang et al. (2004) without incorporating odometry.

3.2. Global Localization

Mobile robot localization is the problem of determining a robot’s pose $x$, from sensor data $z$, given an a priori environment map $m$. The general formula for the localization problem can be formalized in the probabilistic form

$$p(x_t|z_0, z_1, \ldots, z_t, u_1, u_2, \ldots, u_t, m)$$ (3.10)
CHAPTER 3. WORK TO DATE

where \( x_t \) is the pose of the robot at time \( t \), \( z_k \) is the observation from perceptual sensors, such as a laser scanner, at time \( k \), \( u_k \) is the control input from motion sensors, such as an odometer, and \( m \) is the environment map. First of all, we define the following set to refer data leading up to time \( t \).

\[
\mathcal{Z}_t \triangleq \{ z_0, z_1, \ldots, z_t \} \tag{3.11}
\]

\[
\mathcal{U}_t \triangleq \{ u_1, u_2, \ldots, u_t \} \tag{3.12}
\]

The key idea of probabilistic localization algorithms is to maintain a probability density function \( p(x_t | \mathcal{Z}_t, \mathcal{U}_t, m) \) of the location \( x_t \) of the robot at time \( t \) given all observations \( \mathcal{Z}_t \) and all control inputs \( \mathcal{U}_t \) leading up to time \( t \) and the map \( m \). The probability can be derived recursively as

\[
p(x_t | \mathcal{Z}_t, \mathcal{U}_t, m) \propto p(z_t | x_t, m) \int p(x_t | u_t, x_{t-1}) p(x_{t-1} | \mathcal{Z}_{t-1}, \mathcal{U}_{t-1}, m) \, dx_{t-1} \tag{3.13}
\]

where \( p(x_t | \mathcal{Z}_t, \mathcal{U}_t, m) \) is the posterior probability at time \( t \), \( p(x_{t-1} | \mathcal{Z}_{t-1}, \mathcal{U}_{t-1}, m) \) is the posterior probability at time \( t - 1 \), \( p(x_t | u_t, x_{t-1}) \) is the motion model, or the prediction model, and \( p(z_t | x_t, m) \) is the observation model, or the sensor model.

The proposed ego-motion estimator can decompose an observation into stationary objects and moving objects, which can be expressed as

\[
z_t = z^s_t \cup z^m_t \tag{3.14}
\]

\[
\phi = z^s_t \cap z^m_t \tag{3.15}
\]

where \( z^s_t \) and \( z^m_t \) consist of the stationary segments and the moving segments in the observation \( z_t \), respectively, at time \( t \), and form a partition of the observation \( z_t \). Hence, each segment \( z^s_i \) in \( z_t \) is in either \( z^s_t \) or \( z^m_t \). Specifically, this decomposition implies the following conditional independence

\[
p(z_t | x_t, m) = p(z^s_t | x_t, m) p(z^m_t | x_t, m) \tag{3.16}
\]

where \( p(z^s_t | x_t, m) \) and \( p(z^m_t | x_t, m) \) are likelihood functions associating stationary objects and moving objects, respectively, with the map and calculating the corresponding likelihoods accordingly. Furthermore, we can rewrite Equation 3.13 using the conditional independence into the following form

\[
p(x_t | \mathcal{Z}_t, \mathcal{U}_t, m) \propto p(z^s_t | x_t, m) p(z^m_t | x_t, m) \int p(x_t | u_t, x_{t-1}) p(x_{t-1} | \mathcal{Z}_{t-1}, \mathcal{U}_{t-1}, m) \, dx_{t-1} \tag{3.17}
\]
where the observation model is decomposed into two stages calculating likelihoods for stationary objects and moving objects respectively. In the proposed observation model, 
\[ p(z^s_t| x_t, m) \] 
corresponds to the standard observation model for probabilistic localization, whereas 
\[ p(z^m_t| x_t, m) \] 
exploits the information from non-static objects to give rewards or penalties to the observation model. The key insight of the proposed motion-dependent observation model lies in the design of a discriminative observation model. For example, assume a segment \( z^i_t \) is observed in some free space and not within neighborhoods of any stationary landmarks in the map \( m \). The calculation of the likelihood \( p(z^i_t| x_t, m) \) simply treats \( z^i_t \) as sensor noise or unmodeled objects if \( z^i_t \in z^s_t \). Otherwise, it certainly implies that the presence of \( z^i_t \) makes some stationary landmarks occluded, which is supposed to be observed, if \( z^i_t \in z^m_t \). Instead of using a generative observation model, we propose the use of discriminative observation models to better describe the likelihood of an observation towards robust localization.
CHAPTER 4

Proposed Work

My work to date utilizes a multiple-model approach in a random sample consensus (RANSAC) paradigm. The proposed multiple-model approach reliably models moving objects and stationary objects in nature and provides a joint framework to solve the problems of ego-motion estimation and moving object detection. However, the data association problem can be problematic due to merge and split of objects and temporal occlusion. We propose the use of discriminative models to reason about the joint association between measurements. Scaling such a system to solve the global localization problem will increase the reliability for mobile robots to perform autonomous tasks in crowded urban scenes. We propose to formulate a multiple-model extension based on the proposed multiple-model RANSAC (MM-RANSAC) approach for the global localization problem in the probabilistic mobile robot localization framework. Besides, detecting objects of small sizes at low speeds, such as pedestrians, is difficult, but of particular interest in mobile robotics. We propose to use prior knowledge on spatial and temporal coherency from the map to deal with the problem of pedestrian detection. The proposed approach will be demonstrated using experimental testing with real data.

4.1. Increasing Robustness

4.1.1. Out-of-plane Motion Detection

In urban environments, there are mainly two sources of outliers. First, robots do not know whether surrounding objects are stationary or not. Thus, while robots navigate in
unmapped areas, moving object information should be discriminated for obtaining reliable ego-motion estimates. Second, ground terrains are usually not flat. Pitch motions as well as uphill environments result in false positive estimates and will severely affect the accuracy of ego-motion estimation (Singh & Keller, 1991; Batavia & Singh, 2002; Wang, 2004). The solution to the first one is the main contribution of our work to date and we propose to explicitly model pitch motions to tackle the second one using the proposed MM-RANSAC approach. Virtual scans can be naturally generalized into 3D Cartesian coordinate by applying an assumption that environments are composed of vertical planes, and used to provide ego-motion estimates in the pitch motion of a robot.

Prior work in ego-motion estimation suggests that it is impossible to accurately represent pitch motions in planar laser scanner estimation problem. Singh & Keller (1991) indicated that planar laser scanners parallel to the ground plane are not sufficient to detect out-of-plane motions for a vehicle traveling at high speeds. A laser scanner is angled down toward the ground in front of the vehicle to scan the road surface. This method selects to process only parts of the scans based on vehicle pitch and vehicle speed. Batavia & Singh (2002) also proposed the use of a two-axis range scanner which is adapted from a low-cost single line laser scanner. Thus, the laser scanner scans vertically instead of horizontally and provides horizontal coverage. Wang (2004) demonstrated the use of multiple-hypothesis tracking (MHT) (Reid, 1979; Cox & Hingorani, 1996) to minimize the effect of 2D environment assumption in 3D environments using shape and motion inconsistency. Upon divergences of tracks that differ considerably in shape or motion, the corresponding measurements are considered as false measurements. One particular drawback of using a multiple-hypothesis tracker is that the process by which a hypothesis is added to or removed from the filter is heuristic in nature. Whether a hypothesis is pruned away or not is determined by a predefined threshold of the relative likelihood which is determined heuristically. In contrast, the proposed MM-RANSAC approach maintains a direct subspace of the robot’s configuration space (C-space), which facilitates the straightforward use of sophisticated motion models. Extensions of the MM-RANSAC approach could provide benefits to the problem of out-of-plane motion detection while retaining the ease of performing motion propagation. Such an extension will also remove the need for the multiple-hypothesis tracker.
4.2 SCALING TO GLOBAL LOCALIZATION

4.1.2. Joint Data Association

In this thesis, we propose to evaluate consensus sets by examining each internal measurement of the segment. The closest point association rule is applied for each measurement of a data point – a segment. The score of each segment is thus an integer presenting the number of measurements which are located within neighborhoods of measurements in the reference model. As can be seen from Figures 3.1, 3.2, 3.3, 3.4, and 3.5, MM-RANSAC is robust to highly dynamic environments and performs nicely in both ego-motion estimation and moving object detection. The experimental results also show that the proposed algorithm yields a low false-positive rate for moving object detection and the accuracy of both ego-motion estimation and moving object detection are improved.

Yet another concern is the use of nearest neighbor association among segments in the MM-RANSAC approach. Data association can be problematic due to merge and split of objects and temporal occlusion. Comparing to the data association problem in computer vision, the poverty of laser scanner information make data association difficult. Though nearest neighbor association performs well in many circumstances such as for the iterative closest points (ICP) (Besl & McKay, 1992; Lu & Milios, 1997) algorithm, it fails when initial estimates are considerably inaccurate or environments change significantly. Ramos et al. (2007) proposed feature-based scan matching using conditional random fields (CRFs) (Lafferty et al., 2001) to tackle the data association problem. Camera data is also integrated to help the data association. Principal components analysis (PCA) (Jolliffe, 2002) is incorporated for dimensionality reduction of image features. We propose to incorporate discriminative models to reason about the joint association between objects, rather than measurements. Instead of using a distance threshold or defining shape and appearance features for range scans manually, we plan to solve the data association problem at a higher level abstraction.

4.2. Scaling to Global Localization

Probabilistic mobile robot localization, such as extended Kalman filter (EKF) localization (Leonard & Durrant-Whyte, 1991), Markov localization (Fox et al., 1999) and Monte-Carlo localization (MCL) (Fox et al., 1999; Thrun et al., 2000), outperforms other approaches as it models explicitly noises from actuators and perceptual sensors. These approaches heavily rely on the proper definition of underlying probabilistic models – the observation
CHAPTER 4. PROPOSED WORK

model, and are very sensitive to changes in the environment and discontinuities in the map (Thrun, 2001; Pfaff et al., 2008). A key limitation of all localization algorithms arises from the static world assumption, or Markov assumption. However, unmodeled dynamics induce effects on the sensor measurements over multiple time steps. When such effects are paramount, probabilistic localization algorithms that rely on the static world assumption may fail (Thrun et al., 2005, Chap. 8.4).

Prior work in global localization suggests to perform outlier rejection to minimize the effects from unmodeled environment dynamics. Hoffman et al. (2005) proposed to utilize the negative information – what the robot does not see. On the contrary, we propose to make use of the negative information from non-static objects – what the robot is not supposed to see – by explicitly taking into account moving object information for making localization more robust. Scaling the proposed multiple-model approach to solve the global localization problem can increase the reliability for mobile robots to perform autonomous tasks in crowded urban scenes. We propose to use a multiple-model approach based on the probabilistic mobile robot localization framework and formulate an extension to the global localization problem using a motion-dependent observation model in which a likelihood function of an observation depends not only on its location with respect to the map but also its velocity with respect to its prior location.

4.3. Pedestrian Detection

Pedestrians are difficult to detect in terms of their sizes and speeds. The variable density of laser scans make measurements from distant objects considerably sparse. The sparsity of the measurements from a distant object makes its shape insufficient for obtaining a 2D pose from scan matching. Moving at a low speed also result in ambiguity with sensor noise. Wang (2004) proposed to incorporate a moving object map accumulating moving object information temporally and identify that if a segment is in an area previously occupied by moving objects, it can be recognized as a potential moving object.

We propose to combine the proposed ego-motion estimator and probabilistic localization in a single framework. Moving objects are detected at each time step based on the proposed MM-RANSAC approach. Prior knowledge on spatial and temporal coherency of the map are modeled using the occupancy grid map representation (Elfes, 1989) and the probabilistic localization algorithm. The proposed multiple-model extension for probabilistic
localization utilize the results from the proposed ego-motion estimator and moving object
detector. We further propose the use of prior knowledge from the mobile robot localization
framework to improve ego-motion estimation and moving object detection. We will show
how the localization-by-detection and detection-by-localization framework improves the
overall performance of mobile robot localization.

4.4. Timeline

The timeline for the proposed work is given below, which proposes an year of work
leading up to the thesis defense in Spring 2010.

- Spring 2009
  - Formulate the out-of-plane motion in a RANSAC paradigm.
  - Derive a principled multiple-model extension for probabilistic localization.

- Summer 2009
  - Formalize the discriminative learning strategy for joint data association.

- Fall 2009
  - Formalize the localization-by-detection and detection-by-location framework.
  - Set up the experimental testbed.

- Winter 2009
  - Perform experimental testing and performance analysis.
  - Write and defend thesis.
CHAPTER 5

Expected Contributions

The thesis expects to make the following contributions.

1. A principled multiple-model approach in a random sample consensus (RANSAC) paradigm is designed to solve the problems of ego-motion estimation and moving object detection jointly.
2. Robustness of the proposed approach can be increased by taking into account out-of-plane motions and learning joint data association.
3. The first multiple-model extension for global localization using a motion-dependent observation model is proposed to improve the localization performance.
4. The localization-by-detection and detection-by-localization framework is formalized to deal with objects of small sizes at considerably low speeds.


BIBLIOGRAPHY


