Multiple-Model RANSAC for Ego-motion Estimation in Highly Dynamic Environments

Shao-Wen Yang and Chieh-Chih Wang

Abstract— **Robust ego-motion estimation in urban environments is a key prerequisite for making a robot truly autonomous, but is not easily achievable as there are two motions involved: the motions of moving objects and the motion of the robot itself. 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 is performed at a higher 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. 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. Moving objects are successfully detected without incorporating any grid maps, that are inherently time and space consuming. The 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, particularly in environments containing a significant percentage of moving objects.**

I. Introduction

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 [1]. The solution to the SLAM problem has been seen as the fundamental in making a robot truly autonomous [2]. 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 [3]. 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 [4]. There are many challenges in developing a good moving object detector 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.

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 [5]. 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 [6] and Junior [7], 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.

In the robotics literature, the last decade has seen more and more researchers take moving object information into account and solve the SLAM and the DATMO problems concurrently. Wang et al. [8] proposed a consistency-based moving object detector and provided a joint framework to solve the SLAM and the DATMO problems simultaneously. The multiple hypothesis tracking (MHT) method is applied to accomplish data association among moving objects. Bibby and Reid [9] proposed a method that combines sliding window optimization and least-squares together with 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. [10] uses GPS data and control inputs to achieve global consistency in dynamic environments. As a result, establishing the spatial and temporal relationships among the robot, stationary objects and moving objects in the environment serves as the basic for scene understanding. The solutions to the SLAM and the DATMO problems are known to be at the core for mobile robots to act autonomously in real environments. The solution to the moving object detection problem provides a

Shao-Wen Yang is with the Department of Computer Science and Information Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617 Taiwan any@robotics.csie.ntu.edu.tw

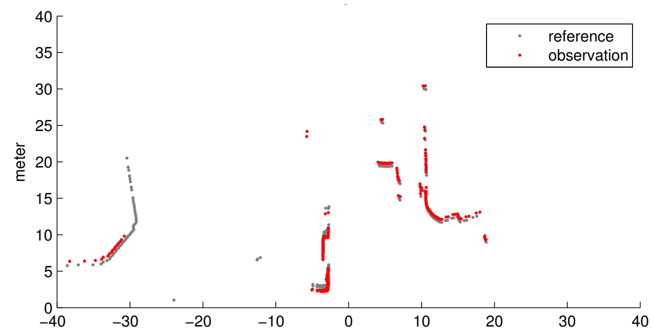
Chieh-Chih Wang is with Faculty of the Department of Computer Science and Information Engineering, and the Graduate Institute of Networking and Multimedia, National Taiwan University bobwang@ntu.edu.tw

bridge between the SLAM and the DATMO problems.

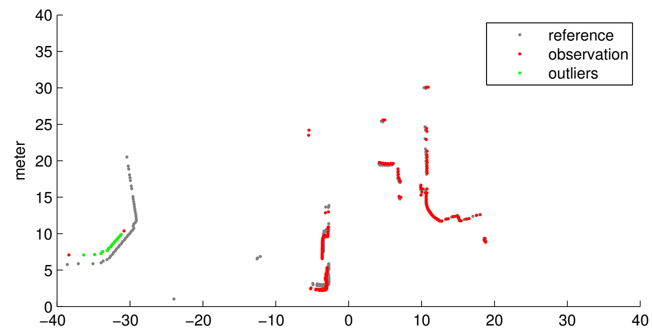
In the computer vision literature, random sample consensus (RANSAC) [11] is one of the most effective algorithm for model fitting to data containing a significant percentage of gross errors. It is an iterative method to estimate parameters of a mathematical model from a set of observed data which contains outliers. A theoretical investigation of scoring under a simple inlier-outlier model is performed to discriminate outliers from the inlier model. The RANSAC paradigm which is capable of smoothing data containing a significant percentage of gross errors is particularly applicable to scene analysis. RANSAC is advanced in its effectiveness and efficiency, but unable to extract multiple models due to its exclusivity in nature.

This paper focuses on the problems of ego-motion estimation and moving object detection in urban environments. We developed an online ego-motion estimation algorithm in a RANSAC paradigm using planar laser scans. The RANSAC process is applied at a higher level abstraction such that sampling of consensus sets can be performed systematically and the computational complexity can also be reduced. A multiple-model extension is 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.

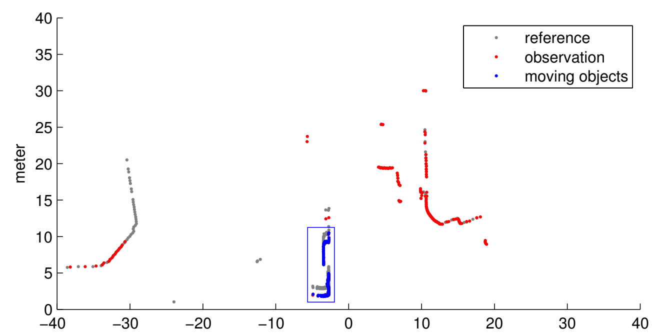
The rest of this paper is organized as follows. In Section II, we present the recent literature regarding ego-motion estimation in urban environments. In Section III, we briefly review the theoretical foundation of RANSAC. In Section IV, a RANSAC-based ego-motion estimator is described. In Section V, to deal with real urban environments, a motion modeling technique is proposed which enables RANSAC to fit data of multiple models simultaneously. The feasibility and tractability of the proposed approach are demonstrated in Section VI. Finally, we conclude with a summary of this work and suggest future extensions in Section VII.



(a) The ICP result at scan 111



(b) The RANSAC result at scan 111



(c) The MM-RANSAC result at scan 111


(d) The visual image at scan 111

Fig. 1. Ego-motion estimation results in which the environments seen by the laser scanner is dominated by moving objects

II. Related Work

Both the SLAM problem and the DATMO problem have been widely studied for decades. 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 accurate if the environment map is available and the robot is capable of localize itself using the map. However, 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 [8].

Schulz et al. [12] and Montemerlo et al. [13] 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 as environments always change, particularly in urban environments. Wang et al. [8] 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 [14]. Inconsistent parts from SLAM 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. [10] applied a delayed mapping and tracking approach in which geometric features are employed. Mendes et al. [15] proposed a voting scheme in which geometric features including shape, size, and other geometric properties are also used.

RANSAC, which has been widely used for outlier rejection in the computer vision literature, is a robust multiple hypothesis estimator in the presence of many outliers. Nistér [16] 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. [17] uses RANSAC for change detection in remote sensing. The transformation in the dynamic range of the images is estimated. Pixels not satisfying this transformation are classified as changes. Both of the works reject moving objects as outliers which are inconsistent with the inlier model.

In this work, the RANSAC paradigm is utilized to extract inlier measurements from planar laser scans. To build consensus sets systematically and achieve tractability for online applications, a RANSAC-based segment matching algorithm is proposed. In case of environments lack of stationary objects for ego-motion estimation, a motion modeling technique for RANSAC to apply multiple models is introduced. To facilitate the access to data of multiple models, a compact representation is used to maintain spatio-temporal information between successive observations. As a result, the proposed multiple-model extension for RANSAC can be performed without any modification of the conventional RANSAC paradigm and seamlessly integrated into the RANSAC process.

III. 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 [11].

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

$$\begin{aligned} \Pr(X = k) &= (1-b)^{k-1}b \quad (1) \\ &= (1-w^n)^{k-1}w^n \quad (2) \end{aligned}$$

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

$$\begin{aligned} (1-b)^k &\leq (1-p), \quad (3) \\ k &\geq \log(1-p)/\log(1-b). \quad (4) \end{aligned}$$

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. A detailed derivation and a comprehensive description can be found in Fischler et al. [11]. In the next section, we will derive the parameters for the ego-motion estimation problem where non-static objects can be rejected as outliers.

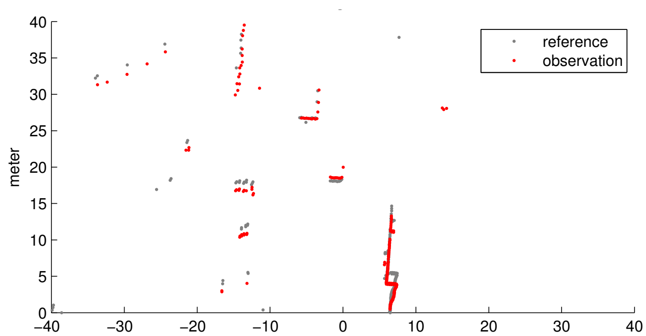
IV. RANSAC-based Segment Matching

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 iterative closest points (ICP) algorithm [18].

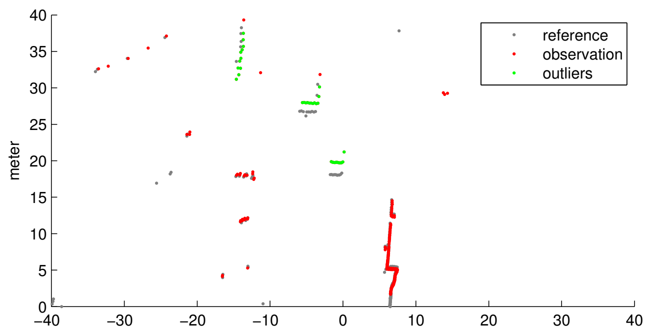
A. Scan Segmentation

In the RANSAC paradigm, a number of random samples consisting of small sets are taken from an 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 4 we have $k \geq 146$. It is time-consuming and computationally intractable for online applications, even though five points are still insufficient for obtaining a good registration result.

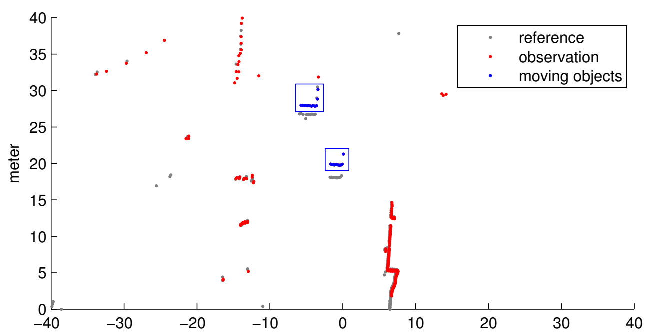
Instead of direct sampling on measurements, we propose to use a higher level data representation – a *segment* – to achieve reliable registration and online applications. An observation is segmented and further split into sets of measurements representing objects. Specifically, objects are extracted by segmenting the scan into densely sampled parts. Here, we use a distance criterion to segment measurements into objects. Although the distance thresholding method



(a) The ICP result at scan 12366



(b) The RANSAC result at scan 12366



(c) The MM-RANSAC result at scan 12366

(d) The visual image at scan 12366

Fig. 2. Ego-motion estimation results in which the environments seen by the laser scanner is highly dynamic

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 [19]. However, experimental results show that the proposed algorithm works well even if the segmentation is not perfect.

B. 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 a 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.

1) Sampling: 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.

2) Scoring: To obtain the score s_h^i of a segment z^i, the effectiveness of (z^i, h) should be examined by checking if (y, h) is an inlier pair for all $y \in z^i$. 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

$$s_h = \sum_{\{i|z^i \in z\}} s_h^i \tag{5}$$

$$= \sum_{z^i \in z} \sum_{y \in z^i} \mathbf{1}_h(y) \tag{6}$$

where $\mathbf{1}_h(y)$ 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 ψ. 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 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 1 and 2 show that RANSAC outperforms ICP in urban environments. In Figures 1(a) and 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 1(b) and 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 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.

V. 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 multiple-model RANSAC (MM-RANSAC), ego-motion estimation can be performed robustly in 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.

A. 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 description of moving objects around the robot. The proposed MM-RANSAC process performs at each time step the additional two stages – segment classification and virtual scan generation.

B. 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 $\bar{\omega}(z^i)$ and $\tilde{\omega}(z^i)$ be the percentage of measurements in z^i which are within neighborhoods of measurements in $\bar{z}$ and $\tilde{z}$, respectively, and $\bar{z}^i \in \bar{z}$ 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

$$\varphi(z^i) = \begin{cases} \text{moving} & \text{if} \quad \tilde{\omega}(z^i) \geq \tilde{\phi},\ \varphi(\tilde{z}^i) = \text{moving} \\ & \text{or} \quad \bar{\omega}(z^i) < \bar{\phi},\ \varphi(\bar{z}^i) = \text{unknown} \\ \text{unknown} & \text{if} \quad \bar{\omega}(z^i) < \bar{\phi},\ \varphi(\bar{z}^i) = \text{static} \\ \text{static} & \text{if} \quad \text{otherwise} \end{cases} \tag{7}$$

where $\varphi(\bar{z}^i)$ indicates the class of the associated segment $\bar{z}^i$ of z^i, and $\bar{\phi}$ and $\tilde{\phi}$ are predefined parameters for determining the effectiveness of an association between segments, which is the only parameters have to be chosen in this work, 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 $\bar{\omega}(z^i)$ 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 is also 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 represented estimated 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.

C. Virtual Scan Generation

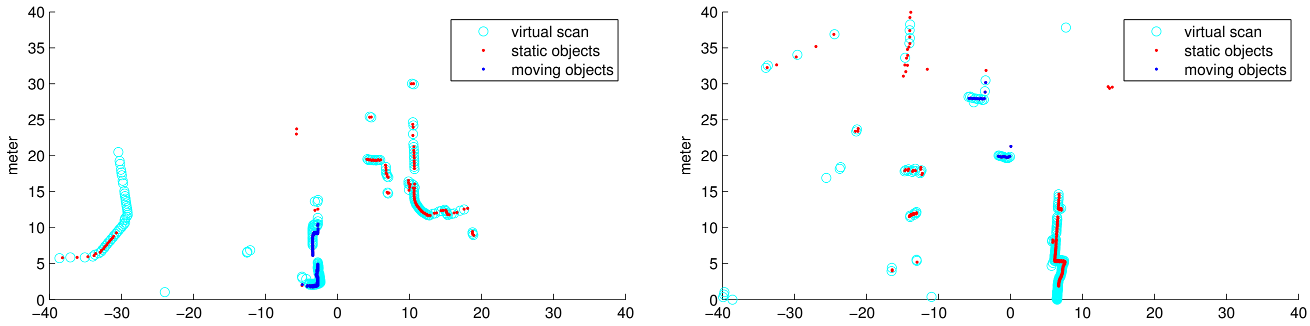
To generate the virtual scan for ego-motion estimation in the upcoming time step, for each segment z^i with $\varphi(z^i) \neq$ moving, the virtual segment $\tilde{z}^i$ is the same as the segment z^i. Conversely, the transformation ψ^i from $\bar{z}^i$ to z^i is calculated by matching these two segments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

$$\varphi(z^i) \neq \text{moving},\ \forall y \in z^i \quad \Rightarrow \quad y = \tilde{y} \in \tilde{z}^i \tag{8}$$

$$\varphi(z^i) = \text{moving},\ \forall y \in z^i \quad \Rightarrow \quad y + \upsilon^i \Delta t = \tilde{y} \in \tilde{z}^i \tag{9}$$

where y is a measurement and υ^i is the estimated linear and angular velocity 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. Comparing to Section IV, we do not assume at least 50% of the measurements in a laser scan are stationary objects anymore. The meaning of the parameter w changes as virtual scans are



(a) The virtual scan at scan 111

(b) The virtual scan at scan 12366

Fig. 3. Virtual scans overlaid with observations

introduced, for 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 within 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 1 and 2 show that MM-RANSAC outperforms RANSAC in urban scenes where environments are highly dynamic. In Figures 1(b) and 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 1(c) and 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.

VI. Experimental Results

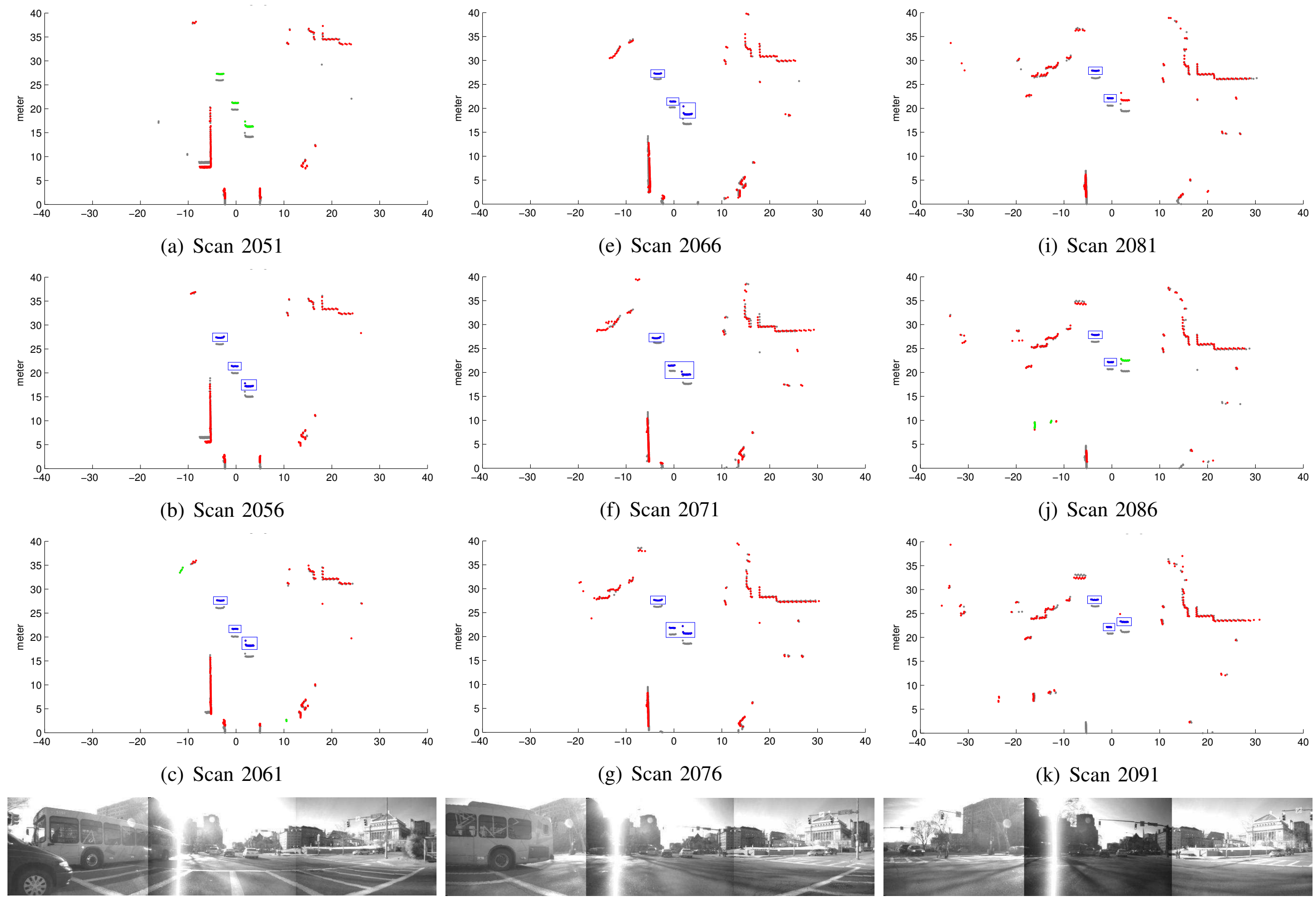
The proposed approach is demonstrated using data from Wang et al. [20]. The travel distance of the data set is approximately 5 kilometer. The average processing time of the proposed ego-motion estimator is 63ms, implemented using MATLAB, running on a desktop PC with Intel Core2 Quad CPU 2.40GHz and 4.0GB RAM, which is sufficient for realtime applications. Figures 1 and 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. Figure 3 visualizes the virtual scans. Figures 3(a) and 3(b) depict the observations overlaid with the virtual scans for the MM-RANSAC results given in Figures 1(c) and 2(c), respectively. In these experiments, RANSAC and MM-RANSAC apply a common sampling stage at each time step, for a fair comparison. A sequence of MM-RANSAC results are shown in Figure 4. The ample experimental results show that the proposed algorithm performs robustly for ego-motion estimation and moving object detection in urban environments. The issue of imperfect segmentation, as addressed in Section IV-A, is also presented here. In the MM-RANSAC paradigm, the problem is resolved naturally by the segment classifier. The values of $\tilde{\phi}$ and $\bar{\phi}$ are essential in the presence of segmentation error, as described in Section V-B.

TABLE I

A quantitative comparison for results in Figures 1 and 2

Score	Scan 111			Scan 12366		
	$\bar{\psi}$		$\tilde{\psi}$	$\bar{\psi}$		$\tilde{\psi}$
$\bar{z}$	191	$>$	181	153	$>$	147
$\tilde{z}$	160	$\ll$	183	172	$\ll$	198

Table I gives a quantitative comparison presenting the scores of the hypotheses applied by RANSAC and MM-RANSAC in Figures 1 and 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 $\tilde{z}$. 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. That is to say, RANSAC and MM-RANSAC give different hypotheses the highest scores 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 1 and 2, and sometimes other hypotheses which are very similar to that of RANSAC are output.



(a) Scan 2051 (e) Scan 2066 (i) Scan 2081

(b) Scan 2056 (f) Scan 2071 (j) Scan 2086

(c) Scan 2061 (g) Scan 2076 (k) Scan 2091

(d) The visual image at scan 2061 (h) The visual image at scan 2076 (l) The visual image at scan 2091

Fig. 4. MM-RANSAC results from scan 2051 to scan 2091. 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.

VII. Conclusion and Future Work

A. Conclusion

In this paper, we address the problems of ego-motion estimation and moving object detection in highly dynamic environments which is of essence for mobile robots to act autonomously in real environments. The main contribution of this paper is to propose a robust ego-motion estimation algorithm which handles moving object information implicitly in a RANSAC paradigm. Consensus sets are built at the segment level such that 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 can be 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 approach has been demonstrated using real data collected in a crowded urban scene without incorporating odometry.

B. Future Work

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 to obtain reliable ego-motion estimation. Second, ground terrains are usually not flat. Pitch motions result in false positive estimates and will severely affect the accuracy of ego-motion estimation. The solution to the first one is the main contribution of this paper and we believe that the second one can also be tackled using only one planar laser scanner. 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.

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 applications such as the ICP algorithm, it fails when initial estimates are considerably inaccurate or environments change significantly. As can be seen from Figure 4, MM-RANSAC can misdetect splitting moving objects or objects undergoing different motions. Future work will also include incorporating 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 manually, we plan to solve the data association problem at a higher level abstraction.

VIII. Acknowledgments

We gratefully acknowledge the helpful discussion with Kuo-Hwei Lin, Chun-Hua Chang and Chung-Han Weng, and the suggestions from anonymous reviewers. This work was partially supported by grants from Taiwan NSC (#96-2628-E-002-251-MY3, #96-2218-E-002-035, #97-2218-E-002-017, #98-2623-E-002-017-D), Taiwan DOIT TDPA Program (#95-EC-17-A-04-S1-054), Taiwan MSI (#97-S-C26) and Taiwan CCI (#97-S-C25).

References

[1] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," *Autonomous Robot Vehicles*, vol. 1, pp. 167–193, 1990.

[2] H. Durrant-Whyte and T. Bailey, "Simultaneous Localisation and Mapping (SLAM): Part I The Essential Algorithms," *Robotics and Automation Magazine*, vol. 13, pp. 99–110, 2006.

[3] Y. Bar-Shalom, T. Kirubarajan, and X.-R. Li, *Estimation with Applications to Tracking and Navigation*. New York, NY, USA: John Wiley & Sons, Inc., 2002.

[4] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting Moving Objects, Ghosts, and Shadows in Video Streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1337–1342, 2003.

[5] C.-C. Wang and C. Thorpe, "Simultaneous Localization and Mapping with Detection and Tracking of Moving Objects," in *Proceedings of the IEEE International Conference on Robotics and Automation*, Washington, DC, May 2002.

[6] C. Urmson, J. Anhalt, D. Bagnell, C. Baker, R. Bittner, M. Clark, J. Dolan, D. Duggins, T. Galatali, C. Geyer *et al.*, "Autonomous Driving in Urban Environments: Boss and the Urban Challenge," *Journal of Field Robotics*, vol. 25, no. 8, pp. 425–466, July 2008.

[7] M. Montemerlo, J. Becker, S. Bhat, H. Dahlkamp, D. Dolgov, S. Ettinger, D. Haehnel, T. Hilden, G. Hoffmann, B. Huhnke *et al.*, "Junior: The Stanford entry in the Urban Challenge," *Journal of Field Robotics*, vol. 25, no. 9, pp. 569–597, August 2008.

[8] C.-C. Wang, C. Thorpe, and S. Thrun, "Online Simultaneous Localization And Mapping with Detection And Tracking of Moving Objects: Theory and Results from a Ground Vehicle in Crowded Urban Areas," in *Proceedings of the IEEE International Conference on Robotics and Automation*, Taipei, Taiwan, September 2003.

[9] C. Bibby and I. Reid, "Simultaneous Localisation and Mapping in Dynamic Environments (SLAMIDE) with Reversible Data Associa," in *Proceedings of Robotics: Science and Systems III*, Atlanta, GA, USA, June 2007.

[10] H. Zhao, M. Chiba, R. Shibasaki, X. Shao, J. Cui, and H. Zha, "SLAM in a Dynamic Large Outdoor Environment using a Laser Scanner," in *Proceedings of the IEEE International Conference on Robotics and Automation*, Pasadena, California, May 2008.

[11] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[12] D. Schulz, W. Burgard, D. Fox, and A. B. Cremers, "People Tracking with Mobile Robots Using Sample-based Joint Probabilistic Data Association Filters," *The International Journal of Robotics Research*, vol. 22, no. 2, pp. 99–116, March 2003.

[13] M. Montemerlo, S. Thrun, and W. Whittaker, "Conditional particle filters for simultaneous mobile robot localization and people-tracking," in *Proceedings of the IEEE International Conference on Robotics and Automation*, Washington, DC, May 2002.

[14] C.-C. Wang, C. Thorpe, S. Thrun, M. Hebert, and H. Durrant-Whyte, "Simultaneous Localization, Mapping and Moving Object Tracking," *The International Journal of Robotics Research*, vol. 26, no. 9, pp. 889–916, September 2007.

[15] A. Mendes, L. C. Bento, and U. Nunes, "Multi-target Detection and Tracking with a Laserscanner," in *IEEE Intelligent Vehicles Symposium*, Parma, Italy, June 2004, pp. 796–801.

[16] D. Nistér, "Preemptive RANSAC for live structure and motion estimation," in *Proceedings of the IEEE International Conference on Computer Vision*, Nice, France, October 2003.

[17] B. Sharma, I. Rishabh, and S. Rakshit, "Unsupervised Change Detection using RANSAC," in *IET International Conference on Visual Information Engineering*, Bangalore, India, September 2006.

[18] P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 2, pp. 239–256, 1992.

[19] S. T. Pfister and J. W. Burdick, "Multi-scale Point and Line Range Data Algorithms for Mapping and Localization," in *Proceedings of the IEEE International Conference on Robotics and Automation*, Orlando, Florida, May 2006.

[20] C.-C. Wang, D. Duggins, J. Gowdy, J. Kozar, R. MacLachlan, C. Mertz, A. Suppe, and C. Thorpe, "Navlab SLAMMOT Datasets," www.csie.ntu.edu.tw/˜bobwang/datasets.html, May 2004, Carnegie Mellon University.