Robust SLAM in Dynamic Environments

Lingzhu Xiang, Zhile Ren, Mengrui Ni
Department of Computer Science
Brown University
Providence, RI 02912
{xlz, ren, mni}@cs.brown.edu

Abstract

Recent developments in human-robot interaction bring about higher requirements for robot navigation. Existing Simultaneous Localization and Mapping (SLAM) algorithms have become unsuited for navigation in complex dynamic environments due to presumptions of static environments or exceeding computational limitations. In this paper we propose a new graph SLAM framework exploring incremental EM algorithms to maintain a reference frame while establishing the estimations of robot trajectory and the map in real-time. We plan to evaluate the performance of existing robust SLAM algorithms as baselines, and validate the improvement of our new framework against datasets of dynamic environments with moving objects.

1 Introduction

Simultaneous localization and mapping (SLAM) is the central problem of autonomous robot navigation, also relates to 3-D reconstruction and augmented reality. Graph SLAM formulate it as an inference problem on a factor graph, where landmark locations and robot poses are the hidden variables nodes to be mapped and localized, and spatial measurements are the observed factor nodes as constraints between variable nodes. Then the goal of the inference problem is to obtain the maximum likelihood estimate of the joint probability of the graph, which becomes the geometrically consistent estimate of robot trajectory and the map. This maximum likelihood estimate on factor graphs can be solved by belief propagation, or more recently, by numerical methods after converting into a non-linear square optimization.

Problems arise when factors incorrectly link unrelated variable nodes, effectively creating wormholes between spatially distant locations thus distorting the map geometry. Various circumstances can result in wrong factors. For pose only graph, the robot continously acquires odometric measurements, which represent factors between the previous and next robot poses, and occasionally acquires loop closures, specifying spatial localities between arbitrary poses. Loop closures are provided by the front-end using algorithms based on certain similarities and inevitably make mistakes. For landmark-based graph, the data association process can also easily obtain wrong visual feature correspondances, that is, connecting landmarks to the wrong poses. Hence, accurate SLAM necessitates robust handling of front-end outliers.

Another issue beyond robust SLAM is dynamism in the environment. Typical SLAM approaches in the literature are designed for unmanned navigation in uninhabited areas, and thus mostly assume a static environment with stationary landmarks or loop closures. However, recent human-robot interaction research has seen more applications of navigation in populated, crowded, or social environments where people and furniture moving around is the major characterstics. If the landmarks are moving, then by the current methods, localization is either carried away by the movement, resulting in the kidnapped robot problem, or more inconsistency happens and the map gets distorted.

In this paper we plan to devise new graph representations and algorithms to address the issue of front-end outliers and the issue of environmental dynamism within the factor graph framework. We will evaluate existing approaches of robust SLAM and conduct experiments on real world datasets to validate the performance of our approach.

2 Related Work

There are multiple recent successful optimization techniques for graph SLAM based on similar formulations. iSAM[1] simultaneously acquires optimal smoothing estimates of the whole trajectory and the map by converting the graph SLAM maximum likelihood estimate into a non-linear least squares optimization problem, which is then incrementally solved by numerical methods, obtaining real-time performance and smoother accuracy. These successful techniques show the effectiveness of the factor graph formulation of the SLAM problem, and we base our formulation in similar forms.

A known solution to moving objects in the environment is combining SLAM with object detection and tracking. In [5], the authors developed a Bayesian framework to solve the SLAM together with detection and tracking of moving objects. They employed sophisticated object detection and tracking and data association algorithms to model object motion. This approach is suitable for the kind of data such as dense point-clouds produced by laser scanners for the purpose of obtaining fast and accurate estimation of motion. However, in the case of sparse landmarks or features generated by visual sensors, there is usually less than sufficient data to achieve the same level of performance in object detection and tracking. Moreover, object detection and tracking is not a part of the SLAM framework per se, instead it is used as a preprocessing filter to prevent moving objects from corrupting the input of the SLAM framework. If the goal is just to estimate the robot trajectory and the still part of the map for future localization, a large part of the work maintaining motion models of moving objects will not be essential to the SLAM problem. In our approach, we do not explicitly model the motion of individual objects, instead we maintain a coherent reference frame in order to obtain only the trajectory and the map and discard irrelevant or potentially moving landmarks. This simplifies the pipeline and removes a chunk of computation for increased efficiency.

In terms of front-end outliers, several robust SLAM approaches have been proposed so far without relying on pre-filtering. Some use robust objective functions or robust representation of graph factors, such as Max-Mixture[3] which represents factors not as a single spatial Gaussian distribution, but as a mixture of Gaussians, which can handle occasional outliers efficiently. The problem with this kind of approaches is that they mainly considers perceptual aliasing errors stemmed from wrong loop closures, and they assume landmarks to be immovable in static environments. Unfortunately such assumptions do not hold for the application of robot navigation in social environments, where there will be people and objects moving around. If a whole block of presumed “landmarks” change their locations, Max-Mixture, for example, will not be able to handle the plausible errors.

However, these methods based on the EM algorithm suffer from the limitation of the EM algorithm in which they do not have the guarantee of convergence in specified number of iterations, making them unsuitable for real-time applications where the optimization must complete before the next measurement arrives. Our approach will address this issue by designing incremental algorithms that is reasonable in performance while fast enough to compute.

3 Graphical Model Formulation

Following [1] we formulate the SLAM problem in graphical models in Figure 1. Specifically, the robot states are denoted by $X = \{x_i\}$ with $i \in 0, \ldots T$, the landmarks by $L = \{l_j\}$ with $j \in 1, \ldots, N$, the control inputs by $U = \{u_i\}$ for $i \in 1, \ldots, T$ and the landmard measurements by $Z = \{z_k\}$ with $k \in 1, \ldots, K$. In addition to the traditional SLAM formulation, we also add latent parameters $W = \{w_j\}$ with $j \in 1, \ldots, N$ The joint probability of all variables and measurements are given by

$$P(X, L, U, Z, W) = \prod_i P(x_i | x_{i-1}, u_i) \prod_k P(z_k | x_{i_k}, l_{j_k}, w_k) \qquad (1)$$

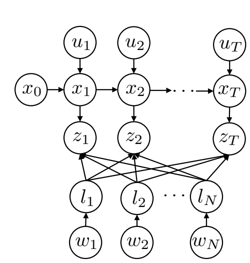

Figure 1: Graphical Model Formulation

Using a Gaussian representation for the sensor model, the process model and measurement equation follows

$$\begin{aligned} x_i &= f_i(x_{i-1}, u_i) + w_i \\ z_k &= h_k(x_{i_k}, l_{j_k}) + v_k \end{aligned} \tag{2}$$

where w_i and v_k follow zero-mean normal distribution with covariance matrices Γ_i and Σ_k. With this formulation, the second part of the equation 1 is

$$P(z_k|x_{i_k}, l_{j_k}, w_k) \propto \exp\{-w_k((z_k - h(x_{i_k}, l_{j_k})^T)\Sigma^{-1}(z_k - h(x_{i_k}, l_{j_k}))\} \tag{3}$$

w_k is the likelihood that the measurement comes from static landmark. In our case, we need to infer what those values are and ideally the process could be online or folows an incremental fashion. Possible solutions could be

- Using visual cues to cluster the landmarks to "moving"/"static". Challenges mainly lies in whether the features we use is reliable or not in terms of this clustering task.
- Instead of using an expensive EM algorithm in [4], we design particle filters to incrementally update parameter w_k, making the whole process online.

4 Preliminary Results

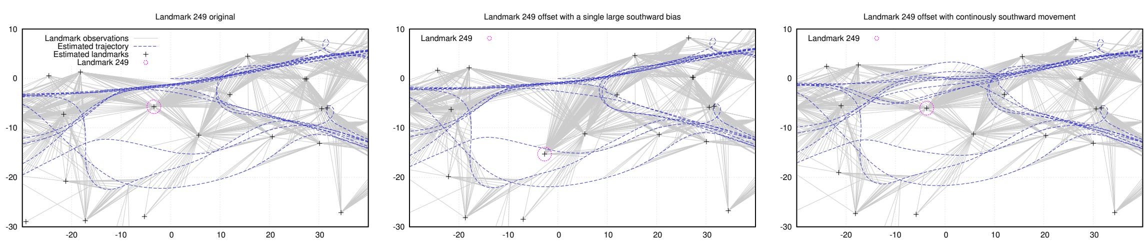

Figure 2: Left: convergence chi-square is 17287. Middle: observations of landmark 249 are displaced with a 10 meter southward bias; convergence chi-square is 231898 indicating high uncertainty. Half of observations of 249 are rejected by max-mixture, while the trajectory suffers from slight distortion with landmark 249 displaced by correct amount. Right: observation of 249 are added with southward moving offsets with the total movement of 7 meter. Convergence chi-square is 17807.

In this section, we evaluate the max-mixture algorithm well-known for its capability of handling a large amount of incorrect loop closures. Since the max-mixture algorithm is a representative framework in the robust SLAM literature, its performance against dynamic environments will have border implication on later work. The evaluation is based on the Victoria Park dataset with different crafted noises introduced on a single landmark (landmark 249). The dataset contains 10607 measurements, 151 landmarks, and 6969 poses. Figure 2 shows a local part of the trajectory and map estimation obtained by the Max-Mixture algorithm on three instances of the dataset with all observations of landmark 249 converted to unimodal max-mixture type. A single large bias introduced in all observations of one landmark makes Max-Mixture results slightly distorted but mostly recovered from the errors. But the simulated errors of slowly moving landmark has largely distorted the result of Max-Mixture.

This demonstrates how the max-mixture algorithm is unable to handle a locally consistent moving landmark while capable of handling noise of large displacement or spurious loop closures. An explanation for this is that each clique of landmarks with coherent motion forms a plausible reference frame for related observations. Inference based on each independent reference frame will reach plausible estimate of robot trjectory and the map, however averaging over a sum of different reference frames would lead to wrong conclusions.

5 Discussion

Inference challenges need to be solved: 1) Not fast enough, how to make EM fast? How to decompose the EM problem into incremental parts? 2) Not robust enough, need testing with real world and see failure modes.

An experimental evaluation protocol: most data related to SLAM will or should have ground truth. To evaluate, compare with the ground truth to determine if the result is accurate enough. Examine the internals of EM, log-likelihood of each iteration, chi-square test to verify the internal correctness.

6 Timeline

11/10 - 11/16 Learning and confirm the exact data structure we'll be handling for each variables in our SLAM model, including parsing the .g2o files[2] used by many SLAM algorithms.

11/17 - 11/23 Try using visual cues to cluster the landmarks. Try to design particle filter algorithms or other incremental algorithms to estimate w_k

11/24 - 11/30 Implementing the algorithms and experiment on datasets.

12/1 - 12/7 Continue implementing the algorithms and experiment on datasets.

12/8 - 12/14 Finish up the write-up and experiment.

References

[1] M. Kaess, A. Ranganathan, and F. Dellaert. iSAM: Incremental smoothing and mapping. *IEEE Trans. on Robotics (TRO)*, 24(6):1365–1378, December 2008.

[2] Rainer Kummerle, Giorgio Grisetti, Hauke Strasdat, Kurt Konolige, and Wolfram Burgard. g2o: A general framework for graph optimization. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 3607–3613. IEEE, 2011.

[3] Edwin Olson and Pratik Agarwal. Inference on networks of mixtures for robust robot mapping. In *Proceedings of Robotics: Science and Systems*, Sydney, Australia, July 2012.

[4] John G Rogers, Alexander JB Trevor, Carlos Nieto-Granda, and Henrik Iskov Christensen. Slam with expectation maximization for moveable object tracking. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pages 2077–2082. IEEE, 2010.

[5] Chieh-Chih Wang, Charles Thorpe, and Sebastian 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 *Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on*, volume 1, pages 842–849. IEEE, 2003.