Collaborative exploration for the construction of visual maps

Ioannis Rekleitis[1], Robert Sim[1], Gregory Dudek[1], and Evangelos Milios[2]

[1]Centre for Intelligent Machines, McGill University
3480 University St, Montréal, Canada H3A 2A7
[2]Faculty of Computer Science, Dalhousie University
Halifax, Nova Scotia, Canada

Abstract

We examine the problem of learning a visual map of the environment while maintaining an accurate pose estimate. Our approach is based on using two robots in a simple collaborative scheme; in practice, one of these robots can be much less capable than the other. In many mapping contexts, a robot moves about collecting data (images, in particular) which are later used to assemble a map; we can think of map construction as a training process. Without outside information, as a robot collects training images, its position estimate accumulates errors, thus corrupting its knowledge of the positions from which observations are taken. We address this problem by deploying a second robot to observe the first one as it explores, thereby establishing a virtual tether*, and enabling an accurate estimate of the robot's position while it constructs the map. We refer to this process as* cooperative localization. *The images collected during this process are assembled into a representation that allows vision-based position estimation from a single image at a later date. In addition to developing a formalism and concept, we validate our results experimentally and present quantitative results demonstrating the performance of the method in over 90 trials.*

1 Introduction

Map construction is essential to robot autonomy. The tasks of path-planning, localization, and reasoning about the environment depend highly on an accurate and robust representation of the world. Furthermore, a representation of the robot's environment is essential to the tasks of teleoperation and debugging remote systems. Examples of useful representations include measures of radiation hot-spots, magnetic declination, sonar and other range-based representations, and visual maps[1, 2, 3]. Of these representations, visual maps offer significant advantages in terms of the richness of the sensor output, the potential for constructing low-cost systems and the utility of the map for application to human-oriented problems such as virtual environment representation. We are interested, in particular, in constructing a vision-based representation of the environment that allows us to subsequently estimate the robot's position accurately from the appearance of a single image. To do this, however, we need to initially collect training data which is the key focus of this paper.

A significant issue faced by many map-building schemes is the accumulation of positional error as the robot collects observations from the environment. That is, as the robot undertakes successive actions, it becomes more uncertain of where it is– leading to a corrupted, or even useless map. In some cases, one can exploit correlation between observations either incrementally or in a batch mode, for example using expectation-maximization (EM), to correct the observation poses [4]. However, it is often the case that either there is insufficient geometric constraint in the observations to produce confident pose estimates even *post hoc*, or, especially in the case of a monocular vision sensor, the computational cost of making the appropriate inferences is infeasible. Other methods such as Kalman filtering can reduce the severity of the problem, but certainly do not eliminate it.

Our visual mapping paradigm develops an implicit representation of visual features, or *landmarks* [5]. Landmark-based representations allow simultaneously for a reduction in processing requirements and robustness in dynamic environments, such as those populated with other agents. One of the requirements of this representation is that a relatively dense sampling of robot positions must be taken in order to reliably track and model the visual features. As such, the robot performs a large number of noise-prone incremental operations, which lead to a rapid degradation of its pose estimate, as illustrated in Figure 1.

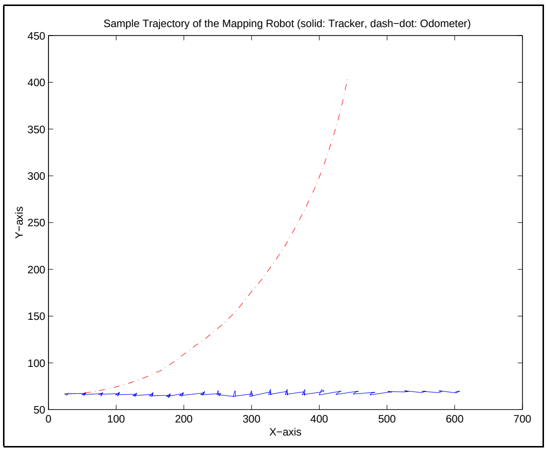

Figure 1: Dead-reckoning estimate versus actual robot trajectory.

Accurate pose estimates are important to the collection of calibrated measurements, which may be useful to tasks beyond that of mapping. Our approach to this problem entails the use of multiple robots working together (Figure 2). Several authors have considered the use of marsupial robots or robot teams either in theory or practice [6, 7]. Others have examined the problem of distributing exploration tasks for efficient coverage of space and time [8]. Our work seeks to exploit robot collaboration for an explicitly quantitative mapping task, where one robot provides a local frame of reference during exploration.

In this framework, the robot collecting measurements for the map operates in concert with a second robot that acts as an active observer. In our *cooperative localization* scheme, this second robot tracks the motions of the first as it collects data and provides it with the information required to prevent odometric error from accumulating. In this sense, a *virtual tether* is established between the two robots that enables the task of mapping to be accomplished without significant error. In principle, more than one of these active observers could be used simultaneously, although this is not elaborated in this paper. Note that, in the case of homogeneous robotic teams, once the map is constructed it can be used by each robot for independent navigation and localization.

This paper provides a quantitative evaluation validating the effectiveness of the methodology outlined above. The remainder of this paper is structured as follows: Section 2 discusses related work that addresses the problem of minimizing localization error during exploration and Section 3 describes our approach to the pose-correction problem. We then discuss the application of this approach to the problem of visual landmark learning in Section 4 and experimental results are presented in Section 5. Finally, we discuss open questions and future directions in Section 6.

2 Related Work

The problem that we have described is closely related to the problem of simultaneous localization and map-building, wherein the robot is tasked to explore its environment and construct a map [9]. In the context of terrain coverage in particular, Balch and Arkin were among the first to quantitatively evaluate the utility of inter-robot communication [10]. Mataric was another pioneer in considering utility of inter-robot communication in space coverage [11]. Dudek, Jenkin, Milios and Wilkes proposed a multi-robot mapping strategy akin to that proposed here, but they only considered certain theoretical aspects of the approach as it applied to very large groups of robots. Several authors have also surveyed the range of possible approaches for collaborative robot interactions [12, 13, 14, 15].

A number of authors have considered pragmatic multi-robot map-making in particular. Most existing approaches operate in the sonar domain, where it is relatively straightforward to transform observations from a given position to expected observations from nearby positions, thereby exploiting structural relationships in the data [16, 8, 7]. These approaches successfully apply the probabilistic *expectation maximization* (EM) paradigm to the task by iteratively refining the map and the estimates of the observation points. In addition, these multi-robot approaches focus on the efficient division of labour.

In other work, Rekleitis, Dudek and Milios have demonstrated the utility of introducing a second robot to aid in the tracking of the exploratory robot's position[17]. In that work, the robots exchange roles from time to time during the exploration of a polygon-shaped world, thus serving to minimize the accumulation of odometry error. The authors refer to this procedure as *cooperative localization.* This paper builds on these results by Rekleitis *et al* by considering the task of exploring the visual domain. In particular, in this paper we deal with the acquisition of a visual representation of the environment. We also assume different roles for the two robots, which would permit the use of robots with very different capabilities. In the following section, we describe the method employed for tracking the position of the robot as it explores.

3 Robot Tracker

We have constructed a tracking device that can estimate the position and orientation of a mobile robot relative to a base robot equipped with a laser rangefinder. The motion planning strategy is such that at any time one of the robots is stationary while the other robot is moving. The stationary robot acts as an artificial landmark in order for the moving robot to recover its pose with respect to the stationary one. Therefore, a detectable landmark is provided without any modification of the environment. We call this approach *Cooperative Localization.* Different types of sensors could be used depending on the required precision of the specific task. In earlier work a visual tracker with a helical pattern on the target robot was used, resulting into a 3-5cm accuracy in the position and a $3-7^{\circ}$ accuracy in the orientation [17]. Currently we employ an *Accu-Range* laser range-finder mounted on one robot and a three plane target mounted on the observed robot (see Figure 2). The target is a set of three vertical planes extending from the center of the target at three distinct angles (approximately $100^{\circ}, 120^{\circ}, 140^{\circ}$). At any given orientation of the target robot at least two vertical planes are "visible". The intersection of the two planes define a unique point in a fixed position with reference to the observed robot. Further on, the angle between the two planes combined with their orientations provides an estimate for the orientation of the robot.

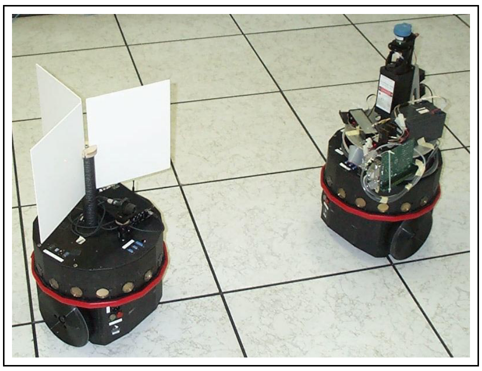

Figure 2: The two robots during the exploration of our laboratory.

The precision of the laser rangefinder subsystem is much higher than the precision of the visual tracker. The position estimation is accurate to half a centimeter and the orientation estimation error is below one degree.
The Robot Tracker returns a triplet of $T =< \rho\ \phi\ \theta >$ that represent: ρ the distance between the two robots, ϕ the angle at which the observing robot sees the observable robot (eg. in Figure 3 the angle the stationary robot sees the moving robot), and θ the orientation of the observed robot as observed by the observing robot (eg. the orientation of the moving robot in Figure 3). As can be seen in Figure 3 both configurations are feasible (any of the two robot could observe). If the stationary robot is equipped with the laser then the Pose ($\mathbf{X_m}$) of the moving robot is given by equation 1, where $< x_s, y_s, \theta_s >$ is the pose of the stationary robot. If the moving robot is equipped with the laser then its Pose ($\mathbf{X'_m}$) is given by equation 2.

$$\mathbf{X_m} = \begin{pmatrix} x_m \\ y_m \\ \hat{\theta_m} \end{pmatrix} = \begin{pmatrix} x_s - \rho * \cos(\hat{\theta}_s + \hat{\phi}) \\ y_s - \rho * \sin(\hat{\theta}_s + \hat{\phi}) \\ \pi + \hat{\phi} + \hat{\theta}_s + \hat{\theta} \end{pmatrix} \quad (1)$$

$$\mathbf{X'_m} = \begin{pmatrix} x'm \\ y'm \\ \hat{\theta}'_m \end{pmatrix} = \begin{pmatrix} x_s - \rho * \cos(\hat{\phi} - \hat{\theta}_s) \\ y_s - \rho * \sin(\hat{\phi} - \hat{\theta}_s) \\ \pi + \hat{\phi} + \hat{\theta}_s - \hat{\theta} \end{pmatrix} \quad (2)$$

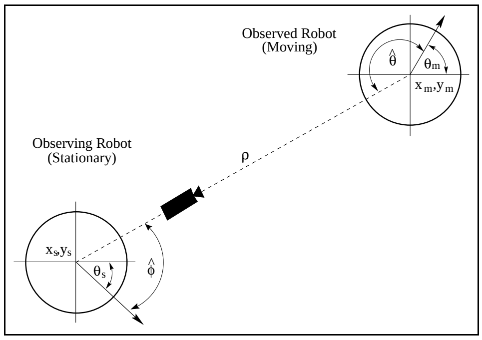

Figure 3: Observation of the Moving Robot by the Stationary Robot. Note that the "camera" indicates the robot with the *Robot Tracker.*

4 Visual Map Learning

We now consider the application of our approach to the task of learning a visual representation of the environment. This demonstrates the performance of the method over a large number of sample locations in the map construction process, as well as a large set of subsequent test cases. The tracker is employed to provide "ground truth" positions while the robot collects training images. We employ the landmark learning framework described in [5] and [18], which tracks a set of salient features using an arbitrary model of visual attention and attempts to construct a set of generative representations of the landmark attributes as a function of the pose of the robot.

The generative nature of each landmark model is useful for the task of *probabilistic robot localization.* That is, we can construct a likelihood function which allows us to measure the likelihood of an observation $\mathbf{z}$, assuming knowledge of the robot's pose $\mathbf{q}$, $p(\mathbf{z}|\mathbf{q})$. Such a likelihood function can be employed in a *Bayesian framework* to infer the probability distribution of $\mathbf{q}$ given the observation $\mathbf{z}$:

$$p(\mathbf{q}|\mathbf{z}) = \frac{\mathbf{p(z|q)p(q)}}{\mathbf{p(z)}} \tag{3}$$

where $p(\mathbf{q})$ represents the prior information about $\mathbf{q}$ and $p(\mathbf{z})$ is a constant relative to the independent variable $\mathbf{q}$. Note that this framework is more generic than a Kalman filter in that it allows for a multimodal representation of the pose likelihood.

When the robot requires a pose estimate without the aid of the tracker, it can obtain a camera image and locate the learned landmarks in the image using the predictive model. The difference in appearance and position between the prediction and the observation is used to compute the likelihood of the observation in the Bayesian framework. The maximum likelihood pose estimate can be recovered by gradient ascent over the likelihood as a function of pose. An example likelihood function is plotted at a coarse scale in Figure 4. Note that the pose likelihood is a useful measure of confidence in the final estimate.

5 Experimental Results

In this section we present the results of deploying the tracking method for the task of landmark learning. Our operating environment was an open laboratory workspace. At the outset, one robot remained stationary while the other used a seed-spreader exploration procedure [19] across the floor, taking image samples at 25cm intervals, and in four orthogonal viewing directions, two of which are illustrated in Figure 5.

The trajectory of the exploratory robot was defined at the outset by a user. However, as the robot explored,

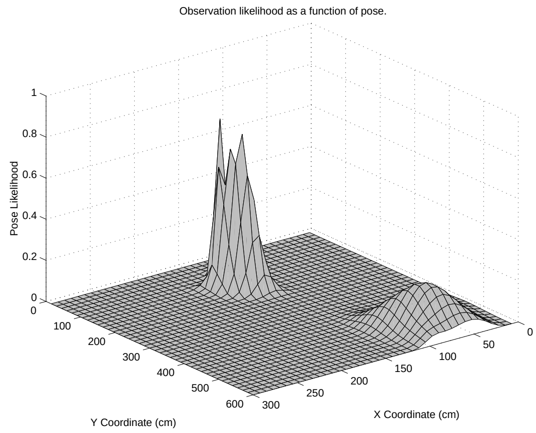

Figure 4: The likelihood of an observation as a function of pose.

accumulated error in odometry resulted in the robot straying from the desired path. The differential drive configuration of the exploratory robot, coupled with frequent rotations to capture the four viewing directions led to a rapid, and somewhat systematic degradation in dead reckoning, as illustrated in Figure 6, where the uncorrected odometric trajectory is plotted as a dash-dotted line, and the actual trajectory of the robot, as observed by the tracker, is plotted as a solid line. The accumulated odometric error is plotted versus total distance traveled in Figure 7.

Once image samples were obtained using the tracker estimates as ground truth position estimates, it was possible to apply our landmark learning framework to the image samples in order to learn a mapping between appearance-based landmarks and the pose of the robot. Training was applied separately to each of the four viewing directions, developing a set of tracked landmark observations.

Our final experiment involved navigating the robot to a series of 93 random positions and acquiring images along the four orthogonal viewing directions. Image- and tracker-based maximum likelihood pose estimates were then generated for one of the viewing directions, and outliers removed on the basis of a likelihood threshold. Of the 93 observations, 29 estimates were rejected. In general, these outliers corresponded to observations where the robot was very close to the wall it was facing. One would expect that an observation from a different viewing direction would return an estimate with higher confidence. We have omitted this application for the sake of brevity.

The remaining 64 image-based estimates of high confidence are plotted with their associated tracker-based

Figure 5: Opposing views of the lab as seen by the robot.

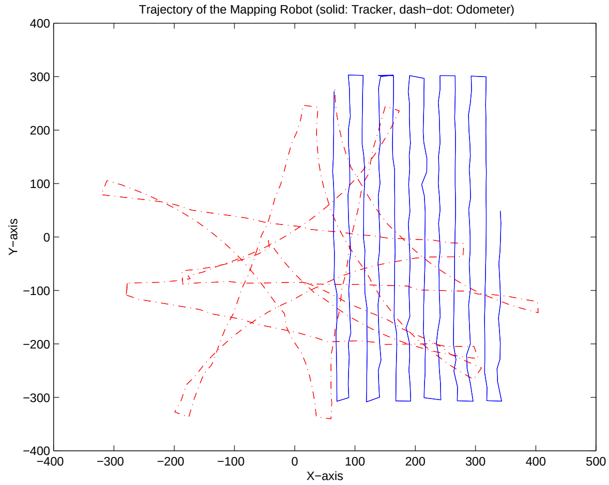

Figure 6: Odometric (x) vs Tracker-corrected (o) trajectories of the robot.

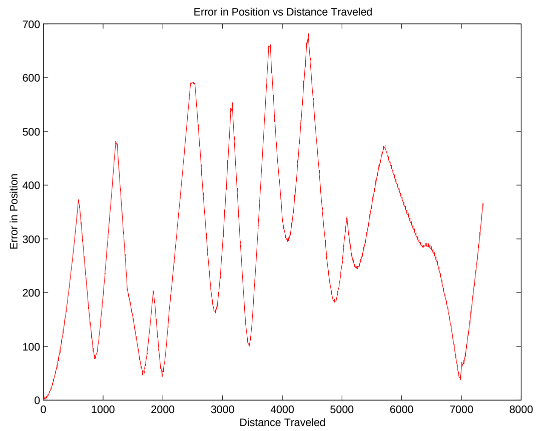

Figure 7: Odometric error versus distance traveled.

estimates in Figure 8. Assuming that the tracker-based position is correct, the mean error in the image-based estimate was 17.3cm.

In another experiment (reported elsewhere [18]) the two robots travelled through two rooms. By exchanging roles of moving and providing a fixed reference point they succesfully mapped two areas 3m by 4m each. The localization results showed an average error of 13.3cm.

6 Conclusions

We have presented a method for the automatic mapping of a visual arbitrary environment which utilizes *cooperative localization* in order to maintain a *virtual tether* between two robots as one explores the environment and the other tracks its pose. The method relies on a mounted target whose pose can be estimated using a laser range-finder. The need for such an approach to maintaining a “ground truth” estimate of the exploring robot is validated by the magnitude of the accrued odometric error in our experimental results. Furthermore, we have validated the utility of a set of learned landmarks for localization when the second robot cannot be deployed.

Since our approach (unlike [17]) assumes one robot does almost all the terrain coverage while the observer robot remains largely stationary, one could posit the use of two rather different robots in this context. While the moving robot needs to cover a substantial amount of terrain rapidly, it does not need to have a very good internal odometry estimate since that is taken care of by the collaborative observation strategy. The stationary robot, on the other hand, only has to move between observation posts and, as such, it can be slower; in principle, it might move only with the assistance of the other robot.

Our results demonstrate that the virtual tether strategy can be used to provide an accurate visual map of the environment. This representation can then be used to accurately estimate the position of a single

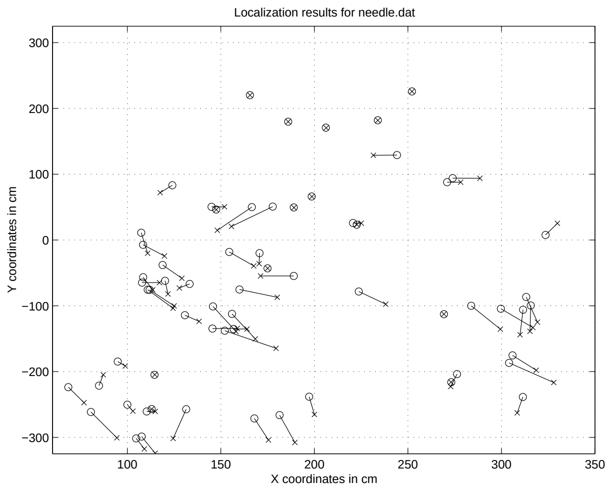

Figure 8: Tracker estimates vs Image-based estimates for a set of 98 random positions.

robot from a single image, as demonstrated on over 90 independent trials. In addition, it would be possible to construct a non-visual map (such as a range map) using the same type of virtual tether methodology. Finally, it is worth noting that the map being constructed could be used for incremental position estimation in combination with the virtual tether for even greater accuracy.

References

[1] B. Kuipers and Y.-T. Byun, "A robot exploration and mapping strategy based on a semantic hierachy of spatial representations", *Robotics and Autonomous Systems*, vol. 8, pp. 46–63, 1991.

[2] Howie Choset and Joel Burdick, "Sensor based planning, part 11: Incremental construction of the generalized voroni graph", in *Proc. of IEEE Conference on Robotics and Automation*, Nagoya, Japan, May 1995, pp. 1643 – 1648, IEEE Press.

[3] F. Dellaert, W. Burgard, D. Fox, and S. Thrun, "Using the condensation algorithm for robust, vision-based mobile robot localization", in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. June 1999, IEEE Press.

[4] Feng Lu and Evangelos E. Milios, "Optimal global pose estimation for consistent sensor data registration", in *Proc. of IEEE Conference on Robotics and Automation*, Nagoya, Japan, May 1995, pp. 93 – 100, IEEE Press.

[5] R. Sim and G. Dudek, "Mobile robot localization from learned landmarks", in *Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems(IROS)*, Victoria, Canada, October 1998, IEEE Press.

[6] Robin R. Murphy, Michelle Ausmus, Magda Bugajska, Tanya Ellis, Tonia Johnson, Nia Kelley Jodi Kiefer, and Lisa Pollock, "Marsupial-like mobile robot societies", in *Agents*, 1999, pp. 364–365.

[7] R. Kurazume and S. Hirose, "Study on cooperative positioning system - optimum moving strategies for cps-iii", in *Proc. IEEE Int. Conf. on Robotics and Automation*, IEEE, Ed., 1998, vol. 4, pp. 2896–2903.

[8] Wolfram Burgard, Dieter Fox, Mark Moors, Reid Simmons, and Sebastian Thrun, "Collaborative multi-robot exploration", in *Proceedings of the IEEE International Conference on Robotics and Automation*, San Francisco, CA, May 2000, IEEE Press.

[9] A. Schultz and W. Adams, "Continuous localization using evidence grids", in *Proc. IEEE International Conference on Robotics and Automation*, Leuven, Belgium, May 1998, pp. 2833–2839, IEEE Press.

[10] Tucker Balch and Ronald C. Arkin, "Communication in reactive multiagent robotic systems", *Autonomous Robots*, vol. 1, no. 1, pp. 27–52, 1994.

[11] M.J Mataric, "Environmental learning using a distributed representation", *IEEE International Conference on Robotics and Automation*, vol. 1, pp. 402–406, 1990.

[12] Gregory Dudek, Michael Jenkin, Evangelos Milios, and David Wilkes, "A taxonomy for multi-agent robotics", *Autonomous Robots*, vol. 3, pp. 375–397, 1996.

[13] Y. U. Cao, A. S. Fukunaga, A. B. Kahng, and F. Meng, "Cooperative mobile robotics: Antecedents and directions", in *Proc. IEEE/RSJ IROS*, Pittsburgh, PA, 1995, vol. 1, pp. 226–234.

[14] Gregory Dudek, Michael Jenkin, Evangelos Milios, and David Wilkes, "A taxonomy for multi-agent robotics", *Autonomous Robots*, vol. 3, pp. 375–397, 1996.

[15] Gregory Dudek and Michael Jenkin, *Computational Principles of Mobile Robotics*, Number ISBN: 0521568765. Cambridge University Press, May 2000.

[16] John J. Leonard and Hugh F. Durrant-Whyte, "Mobile robot localization by tracking geometric beacons", *IEEE Transactions on Robotics and Automation*, vol. 7, no. 3, pp. 376–382, June 1991.

[17] Ioannis M. Rekleitis, Gregory Dudek, and Evangelos Milios, "Multi-robot collaboration for robust exploration", *Annals of Mathematics and Artificial Intelligence*, vol. 31, no. 1-4, pp. 7–40, 2001.

[18] Ioannis Rekleitis, Robert Sim, Gregory Dudek, and Evangelos Milios, "Collaborative exploration for map construction", in *2001 IEEE International Symposium on Computational Intelligence in Robotics and Automation*, July 2001.

[19] V. Lumelsky, S. Mukhopadhyay, and Kang Sun, "Sensor-based terrain acquisition: The 'sightseer' strategy.", in *Proceedings of the IEEE Conference on Decision and Control Including The Symposium on Adaptive Processes*, IEEE Service Center, Piscataway, NJ, USA, 1989, IEEE, vol. 2, pp. 1157–1161, IEEE.