

Chapter 1

Cooperative Exploration, Localization, and Visual Map Construction

Ioannis M. Rekleitis, Robert Sim, and Gregory Dudek

Abstract We examine the problem of learning a visual map of the environment based on discrete landmarks. While making this map we seek to maintain an accurate pose estimate for the mapping robots. Our approach is based on using a team of at least two (heterogeneous) mobile robots in a simple collaborative scheme. In many mapping contexts, a robot moves about the environment collecting data (images, in particular) which are later used to assemble a map; we view the map construction as both a knowledge acquisition and a training process. Without reference to the environment, as a robot collects training images, its position estimate accumulates errors, thus corrupting its estimate 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 time. In addition to developing a formalism and concept, we validate our approach experimentally and present quantitative results demonstrating the performance of the method in over 90 trials.

1.1 Introduction

Many robotic tasks require that the robot learn a representation, or map, of some property of the environment. Examples of such maps include measures of radiation hot-spots, magnetic declination, sonar representations and visual maps [4, 6, 16].

I. Rekleitis and G. Dudek
Centre for Intelligent Machines and the School of Computer Science, McGill University
e-mail: [yiannis,dudek]@cim.mcgill.ca

R. Sim
Microsoft Corporation
e-mail: rsim@microsoft.com

The specific problem we consider is mapping a property of interest over an unknown environment. A significant issue faced by many map-building schemes is the management and estimation of positional (or pose) errors as the robot collects observations from the environment. That is, as a robot or a team of robots collects successive measurements from different positions and orientations (poses), the certainty of their pose estimates decreases with each new measurement. In some cases where the observations lie on a high-dimensional manifold, correlation between dimensions allows for globally consistent alignment of the observations via an expectation-maximization or iterative optimization approach to correcting the observation poses [20, 37]. That is, we can recover the spatial distribution of the measurements. 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 that the computational cost of making the appropriate inferences is infeasible. Uncertainty modeling methods such as Kalman filtering can reduce the severity of the problem, but certainly do not eliminate it.

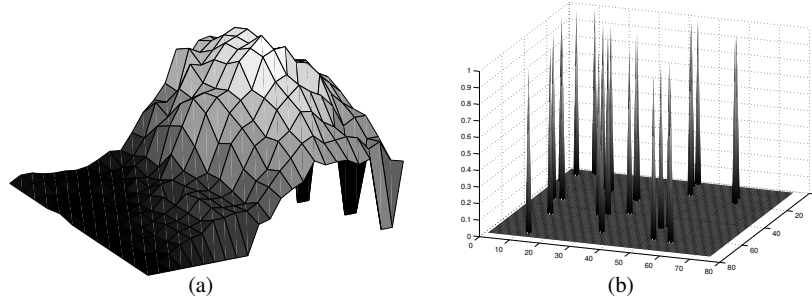


Fig. 1.1 Mapping: **(a)** Continuous function such as: Radiation, Visual appearance, Elevation, Magnetic field, Temperature, etc. **(b)** Discrete function such as: Mine detection, Lost objects, Holes, Electrical outlets, etc.

Our approach to the pose estimation problem for map building involves the use of two or more robots working in close cooperation. Several authors have also considered the use of marsupial robots or robot teams either in theory or practice [13, 24, 40].

This paper addresses the problem of establishing accurate pose estimates in the context of robotic mapping. The pose estimates can be used to collect accurately localized measurements in their own right, or as a precursor to a system that builds a map. The robot collecting measurements for the map operates in concert with a second robot that acts as an active observer. In our *cooperative localization* [31] 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. We can view the robots as being “connected” by a *virtual tether* which is established

between the two robots and which enables the task of mapping to be accomplished without significant error and independent of the ground surface conditions and the quality of the odometry estimate. In principle, more than one of these active observers could be used simultaneously, although this is not elaborated on in this paper. Beyond presenting the details of the approach and its implementation, this paper provides a quantitative evaluation validating the effectiveness of this methodology.

The remainder of this paper is structured as follows: Section 1.2 discusses the general framework in which our approach applies. Section 1.3 presents related work that addresses the problem of minimizing localization error during exploration. The *cooperative localization* strategy is introduced in Section 1.4. We then discuss a particular application of our approach to the task of visual landmark learning in Section 1.5 and experimental results are presented in Section 1.6. Finally, we discuss open questions and future directions in Section 1.7.

1.2 Motivation

The work presented here is motivated by the need to use a mobile robot in order to accurately map a spatially varying property of an unknown, possibly hazardous, environment. Such a property could be a continuous function, see Fig. 1.1a, over the accessible area such as radiation, temperature, magnetic field variation, elevation, or visual appearance, or the property could be a discrete function, see Fig. 1.1b, such as presence of mines, lost objects, holes/anomalies on the ground, or electrical outlets. In addition, the property could be a scalar or a vector-valued function. In most cases the sensor used to map arbitrary properties such as those noted above is not itself suitable for the accurate localization of the exploring robot – for example, a radiation meter cannot readily be used to accurately recover the pose of the exploring robot (except in very special cases). In practice, there are two different issues to consider. First the property of interest may vary too slowly to accurately assist in localization of the robot; second the values of the property of interest may be identical in many places, thus making the task of distinguishing two places impossible. Thus, the self-localization ability of a single robot using only the measurements of the function of interest may be poor in the absence of additional sensory apparatus. Furthermore, the terrain being explored may be uneven or otherwise problematic, resulting in wheel slippage, and rendering the odometry unreliable. Our approach employs cooperative localization [29] in order to recover the pose of the exploring robot with high accuracy, independent of the ground surface properties and the reliability of the odometry.

Another motivation for using more than one mobile robot is that several applications require the exploration or inspection of hazardous environments with an attendant risk to the robot doing the work. Such applications include but are not limited to: de-mining rural areas, inspecting nuclear facilities or marking/mapping of chemical spills. In order to improve robustness or reduce the potential cost in such a scenario we can deploy a team of heterogeneous robots consisting of a “base” robot which is equipped with the main computer, a communication module and the

robot tracker sensor, and a team of lower-cost “exploring” robots that are equipped with only the mapping sensor (and the target for the robot tracker). In particular, our scheme obviates the need for accurate odometry on the exploring robots. The base robot is always located at a safe area keeping visual and radio contact with the exploring robots. If any of the exploring robots is destroyed the expense is limited, and the mission can continue with the surviving robots. The main advantage of this approach is that the (expensive) base robot is not endangered by moving to an unexplored hazardous environment. In the experiments presented in this paper there is one base robot equipped with the robot tracker sensor and one mapping robot equipped with the target and a measuring device (a camera).

1.3 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 [19]. The advantages of collaborative behavior have been examined extensively in the context of biological systems [38].

In the context of terrain coverage in particular, Balch and Arkin were among the first to quantitatively evaluate the utility of inter-robot communication [1]. Mataric was another pioneer in considering the utility of inter-robot communication and teamwork in space coverage [22]. 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 a range of possible approaches for collaborative robot interactions [3, 8, 9].

More recently, teams of mobile robots are used to reduce the localization error [17, 33]. In most cases the robots use each other to localize only if they meet by chance. Different estimation techniques have been employed to combine the information from the different robots: particle filters have gained popularity [11, 30] together with more traditional Kalman filter estimation [10, 34] or more recently maximum likelihood estimation [15].

A number of authors have considered pragmatic map-making in particular. Most existing approaches operate in the range-sensing 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 [2, 12, 18]. Such approaches usually differ in when the map is constructed. Off-line approaches post-process the data, usually by applying the *expectation maximization* (EM) paradigm [7] to the task by iteratively refining the map and the estimates of the observation points. On-line methods compute the maximum-likelihood map and robot pose as it explores, and are usually based on Kalman filtering or particle filtering and their extensions [23, 32, 33].

Several authors have investigated vision-based pose estimation and map construction [14, 39]. Se *et al* construct a map by extracting point features from images and localizing them in the world with a stereo camera [35]. Davison performs

SLAM using a monocular camera by applying techniques from the structure-from-motion literature [5]. Other authors have considered generic features described implicitly by computing the principal components of sensor observations [25, 26]. Our work is similar to the earlier localization techniques in that it applies probabilistic methods to localization from feature observations. However, it is more similar to the latter techniques in that feature and camera geometry are not modeled explicitly but rather the (possibly complex) interaction of feature and sensor is learned as a function of pose.

1.4 Cooperative Localization

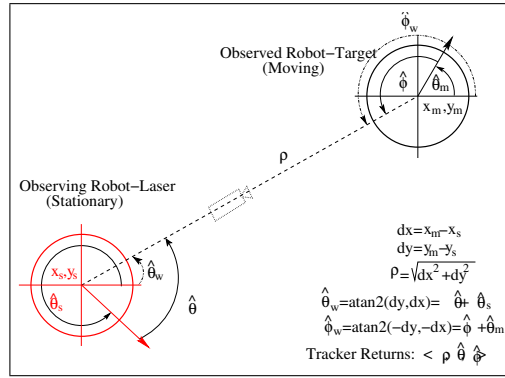


Fig. 1.2 Pose Estimation via Robot Tracker: Observation of the Moving Robot by the Stationary Robot.

In previous 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 [28]. 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*. We have constructed a tracking device that can estimate the position and orientation of a mobile robot relative to a base robot equipped with the robot tracker sensor. 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.

This paper builds on the results by Rekleitis *et al.* [27] by considering the task of exploring the visual domain. In the following section, we describe the method employed for tracking the position of the robot as it explores. Different types of sensors could be used depending on the required precision of the specific task.

1.4.1 Tracker-based Pose Estimation

The Robot Tracker sensor returns three measurements, the triplet $T = [\rho \ \phi \ \theta]$ where ρ is the distance between the two robots, ϕ is the angle at which the observing robot sees the observed robot relative to the heading of the observing robot, and θ is the heading of the observed robot as measured by the observing robot relative to the heading of the observing robot; see Fig. 1.2. If the stationary robot is equipped with the Robot Tracker then the Pose ($\mathbf{X}_m$) of the moving robot is given by eq.(1.1), where $[x_s, y_s, \theta_s]^T$ is the pose of the stationary robot.

$$\mathbf{X}_m = \begin{pmatrix} x_m \\ y_m \\ \hat{\theta}_m \end{pmatrix} = \begin{pmatrix} x_s + \rho * \cos(\hat{\theta}_s + \hat{\theta}) \\ y_s + \rho * \sin(\hat{\theta}_s + \hat{\theta}) \\ \pi + \hat{\theta}_s + \hat{\theta} - \hat{\phi} \end{pmatrix} \quad (1.1)$$

We have implemented two trackers based on this paradigm. The first operates in the visual domain using a helical target, whereas the second employs a laser range-finder and a geometric target. In the following sections we outline their implementations.

1.4.2 Implementation 1: Visual Robot Tracker

The first implementation of a visual robot tracker involved the mounting of a helical target pattern on the observed robot, while the observing robot was equipped with a camera.

The bottom part of the target pattern is a series of horizontal circles (in fact, these are cylinders and they project into a linear pattern in the image). This allows the robot to be easily discriminated from background objects: the ratio of spacing between the circles is extremely unlikely to occur in the background by chance. Thus, the presence of the robot is established by a set of lines (curves) with the appropriate length-to-width ratio, and the appropriate inter-line ratios. Figure 1.3a shows the gray-scale image and Fig. 1.3b presents the identifying stripes highlighted.

By mounting the observing camera above (or below) the striped pattern of the observed robot, the distance from one robot to the other can be inferred from the height of the stripe pattern in the image ¹, due to perspective projection; scaling of the pattern could also be used. The difference in height between the observing camera and the target can be selected to provide the desired trade-off between range of operation and accuracy.

The second component of the target pattern is a helix that wraps once around the observed robot. The elevation of the center of the helix allows the relative orientation of the observed robot to be inferred; see Fig. 1.3.

In practice, the above process allows the observed robot's pose to be inferred with an accuracy of a few centimeters in position and 3 to 5 degrees in heading [29]. The

¹ After an initial calibration process a look-up table is constructed that relates y-coordinates in the image with distance.

range of the visual tracker depends on the height of the robots, in our experimental setup the effective range was between 1.5 m to 4.5 m. Experimental results showed the system to be sensitive to partial occlusions and especially to uneven floors. In the next section we present a robot tracker sensor based on a laser range finder. The experiments presented in this paper utilized the laser-based robot tracker.

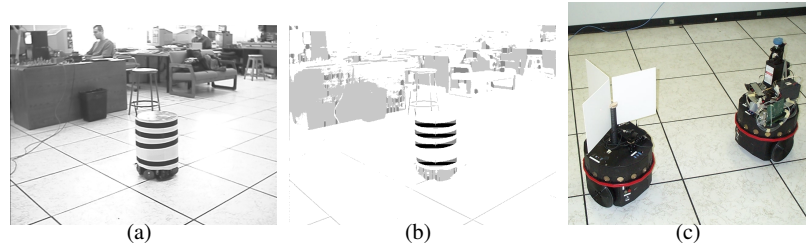


Fig. 1.3 Robot Tracker: (a) The raw image of the moving robot as observed by the robot tracker. (b) The helical and cylindrical pattern detected in the image. (c) The laser-based tracker and target.

1.4.3 Implementation 2: Laser Robot Tracker

The second and more accurate implementation employs an *AccuRange*² laser range-finder mounted on the observing robot and a three-plane target mounted on the observed robot; see Fig. 1.3c. The *AccuRange* laser range-finder produces a range scan in the horizontal plane. The effective range of the *AccuRange* scanner is up to 12 m with an angular resolution of 11.3 points per degree (4096 points for 360 degrees). The manufacturer specifications claim sub-millimeter accuracy per point; in practice we observed less than a centimeter variation regardless of the range. The target is a set of three vertical planes extending from the center of the target at three distinct angles (100°, 120°, 140°) with length of 25cm and height of 30cm. From any direction around the observed robot at least two vertical planes are “visible”. We employ the laser range-finder in order to detect the two planes. The intersection of the two planes defines 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 heading of the robot. When the approximate position of the observed robot is available (most of the time) then the laser points are filtered around that position and few lines have the correct size to select. At maximum range (12m) between 8 and 12 points are returned per target plane.

² *AccuRange* 4000 LineScanner laser range-finder from ACUITY RESEARCH Inc. (see <http://www.acuityresearch.com/acculine.html>)

The precision of the laser range-finder subsystem is much higher than the precision of the visual tracker. The position estimation is accurate to half a centimeter and the heading estimation error is below one degree.

1.5 Application: Landmark Learning and Landmark-based Pose Estimation

In this section we demonstrate the effectiveness of our approach as it applies to the problem of constructing a visual map of an unknown environment. The visual map entails learning a set of visual landmarks which are useful for the task of estimating the pose of a single robot equipped with a monocular camera. We employ the tracker described in the previous section to properly register the landmark observations in the map, i.e. to provide “ground truth” positions while the robot explores the visual environment. We employ the visual mapping framework described in [36], which tracks the set of points output by a model of visual attention and attempts to construct representations of the landmarks as functions of the pose of the robot. The landmark representations do not recover 3D information about the landmark, nor do they depend on pre-defined camera calibration parameters. Instead, the representations learn the generating function $\mathbf{z} = \mathbf{f}(\mathbf{X}_m)$ that maps robot poses to landmark observations $\mathbf{z}$. This is accomplished by training a set of interpolation networks on a set of landmark training observations and ground truth poses. The remainder of this section will describe the visual mapping approach, and in subsequent sections we will present experimental results illustrating how the tracker can be employed to construct the map.

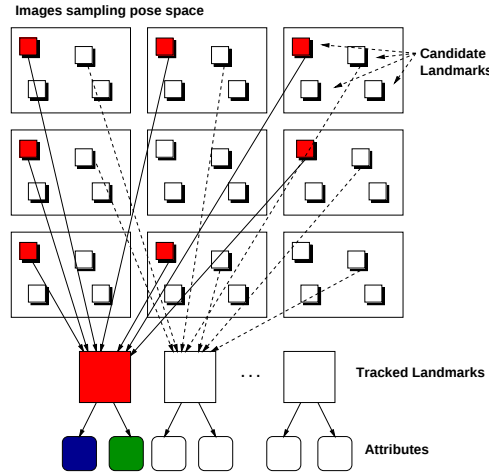


Fig. 1.4 The off-line training method. Images (large rectangles) are collected sampling the pose space. Landmarks are extracted from the images and matched across the samples. The *tracked landmarks* are parameterized as a function of pose and saved for future pose estimation.

The learning method is depicted in Fig. 1.4 and operates as follows (refer to [36] for further details):

1. **Exploration:** One robot tracks the other as it collects images sampling a range of poses in the environment. The pose at which each image is taken is recorded as the estimate given by the tracker.
2. **Detection:** *Landmark candidates* are extracted from each image using a model of visual attention. In this work, we identify candidates as local maxima of edge density in the image.
3. **Matching:** *Tracked landmarks* are extracted by tracking visually similar candidate landmarks over the configuration space. The measure of visual similarity we employ is the normalized correlation between candidate image regions.
4. **Parameterization:** The tracked landmarks are parameterized by computing the behavior of a set of computed landmark attributes (for example, position in the image, intensity distribution, edge distribution, etc), versus the “ground truth” pose provided by the tracker. The resulting models are then measured in terms of their *a priori utility* for pose estimation using cross-validation. The cross-validation covariance C , along with the inferred generating function $f(\mathbf{X}_m)$ for each landmark will subsequently be employed for modeling the feature likelihood distribution $p(\mathbf{z}|\mathbf{X}_m)$.
5. The set of sufficiently useful landmark models is stored for future retrieval.

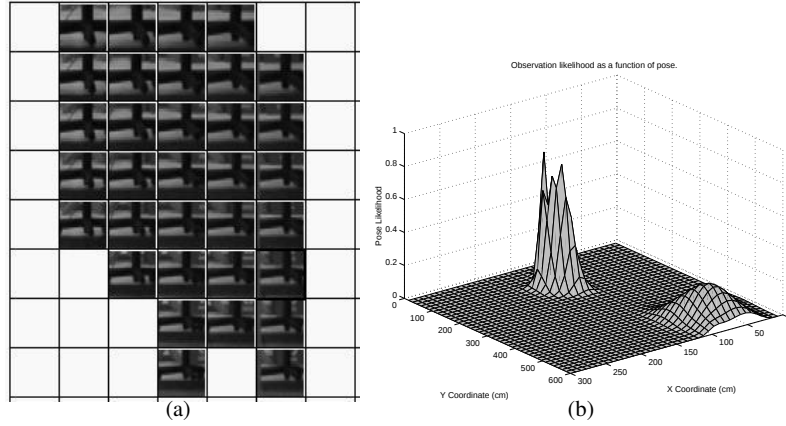


Fig. 1.5 (a) A set of observations of an extracted scene feature. The grid represents an overhead view of the pose space of the camera, and feature observations are placed at the pose corresponding to where they were observed. Note that the observations capture variation in feature appearance. (b) The likelihood of an observation as a function of pose.

For the purposes of our experiments, the visual landmarks are initially selected from a subset of the training images using an attention operator that responds to

local maxima of edge density in the image. The selected landmark candidates are then tracked over the remaining images along the robot’s trajectory by maximizing correlation with the local appearance of the initially detected landmark. The set of matches to a given candidate constitute a *tracked landmark*, and is stored for parameterization and evaluation. Figure 1.5a depicts an example of a tracked landmark. Each thumbnail corresponds to the landmark as it was observed from the pose corresponding to its position in the grid (representing an overhead view of the pose space).

The parameterization of each landmark feature f_i is accomplished by employing a radial basis function regularization framework to model the observation generating function

$$\mathbf{z} = \mathbf{f}_i(\mathbf{X}_m), \quad (1.2)$$

where $\mathbf{z}$ is a low-dimensional vector-valued representation of the landmark attributes and $\mathbf{X}_m$ is the pose of the robot. In other words, $\mathbf{f}_i(\cdot)$ is the function that predicts the attributes of the landmark i as a function of the pose of the robot. Furthermore, the landmark is evaluated for its utility by computing the covariance C of a randomly sampled subset of leave-one-out cross-validation residuals over the training set. The cross-validation error provides an *a priori* estimate of the utility of the landmark and landmarks with large error can be discarded.

The parameterization of each landmark affords a maximum likelihood prediction of an observation, given an *a priori* pose estimate $\mathbf{X}_m$, as well as a measure of the uncertainty (C) of that prediction. As such, the landmark models are useful for the task of probabilistic robot localization. That is, we define a likelihood function $p(\mathbf{z}|\mathbf{X}_m)$ which allows us to measure the likelihood of an observation $\mathbf{z}$, assuming knowledge of the robot’s pose $\mathbf{X}_m$:

$$p(\mathbf{z}_i|\mathbf{X}_m) = k \exp(-0.5(\mathbf{z}_i - \mathbf{f}_i(\mathbf{X}_m))^T C^{-1}(\mathbf{z}_i - \mathbf{f}_i(\mathbf{X}_m)))$$

where k is a normalizing constant. This likelihood distribution can be employed in a *Bayesian framework* to infer the probability distribution of $\mathbf{X}_m$ given the observation $\mathbf{z}$:

$$p(\mathbf{X}_m|\mathbf{z}) = \frac{p(\mathbf{z}|\mathbf{X}_m)p(\mathbf{X}_m)}{p(\mathbf{z})} \quad (1.3)$$

where $p(\mathbf{X}_m)$ represents the prior information about $\mathbf{X}_m$ and $p(\mathbf{z})$ is a constant relative to the independent variable $\mathbf{X}_m$. Several such probability distributions can be generated—one for each observed landmark—and can be combined to obtain a full posterior pose distribution. Note that this framework is more generic than a Kalman filter in that it allows for a multi-modal representation of the pose likelihood.

The set of landmarks observed and computed over the environment during the mapping stage constitutes the visual map and can later be used for accurate single-robot pose estimation. When the robot requires a pose estimate without the aid of the robot tracker, it obtains a camera image and locates the learned landmarks in the image using the predictive model and the tracking mechanism. The differences in appearance and position between the prediction and the observation of each landmark are combined to compute the likelihood of the observation in the Bayesian

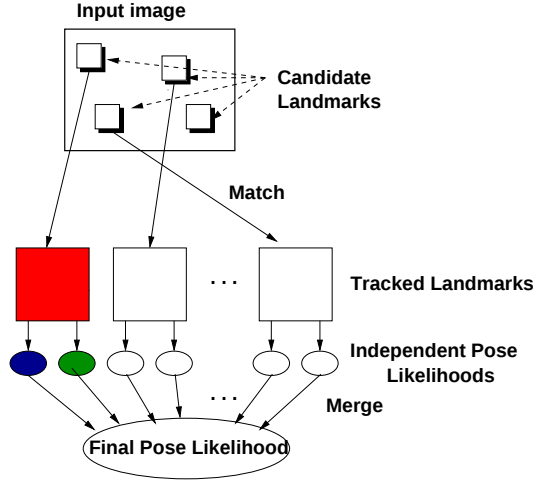


Fig. 1.6 Pose estimation based on learned visual landmarks. Landmarks (small squares) are extracted from the current camera observation and matched to the previously learned tracked landmarks. Each match generates a pose estimate, all the matches are filtered and combined to generate a final pose estimate.

framework. This process is illustrated in Fig. 1.6. The maximum posterior pose estimate is recovered by gradient ascent over the observation likelihood as a function of pose. An example likelihood function is plotted at a coarse scale in Fig. 1.5b. Note that the pose likelihood is a useful measure of confidence in the final estimate allowing for the rejection of outlier pose estimates on the basis of a user-defined threshold.

The cost of computing the posterior distribution will be dependent on any prior pose information that is available. Uninformative priors will imply that every learned landmark will be considered for matching, and successfully matched landmarks will subsequently be evaluated over the search region of the pose space using their generative models. Assuming that landmark matching constitutes the greater computational cost of pose estimation (which is generally the case), the pose estimation process is linear in the number of landmarks in the map. This computational cost can be reduced by exploiting prior pose information to pre-select for matching only those landmarks that are *a priori* likely to be visible.

While this work computes only the mode of the posterior pose distribution, it should be noted that eq.(1.3) is well suited to alternative representations, such as computation of the posterior over a discretization of the pose space, or computation at the loci of a set of points in a particle filter. Furthermore, the parameterization of the landmark models is generative in nature, eq.(1.2), so that the representation can be employed in a more approximate approach, such as a Kalman filter or other hypothesis tracker. Such approaches are beyond the scope of this work, but are straightforward to apply in practice.



Fig. 1.7 Views of the two “rooms” as seen by the robot, and the floor plan of the two “rooms”.

1.6 Experimental Results

In this section we present the results of deploying the tracking method for the task of landmark learning using a team of two robots. Two different experiments were conducted in our laboratory. In the first experiment we tested the ability of the team to move through two consecutive rooms while in the second experiment we maximized the area covered using all the available space.

1.6.1 Experiment 1:

Our environment consisted of a laboratory partitioned into two “rooms” by room dividers, with an open doorway connecting them. The first two pictures in Fig. 1.7 are the robot’s-eye-view of the two rooms, and the third picture presents the top view of the floor plan. At the outset, one robot remained stationary while the other used a seed-spreader exploration procedure [21] across the floor, taking image samples at 40cm intervals. When the robot had completed the first room, it moved beyond the door and the stationary robot followed it to the threshold, where it remained stationary while tracking the exploratory robot as it continued its exploration of the second room. In this experiment the robots first map one room, then move to the next room. The accumulated uncertainty is very small due to the sort path taken by the observing (base) robot. More significant is the odometric error that occurred when the observed (mapping) robot ran over a cable. The cooperative localization approach, though, corrected the pose estimate and the visual map was constructed accurately.

a) Odometry versus tracking: The trajectory of the exploratory robot was defined at the outset by a user. However, as the robot explored, accumulated error in odometry resulted in the robot straying from the desired path. The tracking estimate of the stationary robot was provided to the moving robot in order to correct this accumulated error. During the exploration the pose of the robot was corrected based on the observations of the robot tracker. During the experiment the pure odometry estimates were kept for comparison. Figure 1.8a plots the uncorrected odometric trajectory (plotted as ‘x’) and the actual trajectory of the robot, as measured by the tracker (plotted as ‘o’). For the sake of clarity, the odometric error was reset to zero

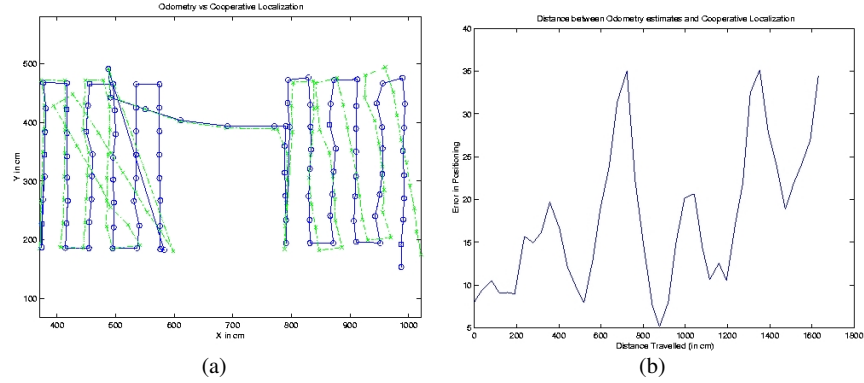


Fig. 1.8 (a) Odometric (x) vs. Tracker-corrected (o) trajectories of the robot. (b) Odometric error versus distance traveled. between the first and second rooms. Figure 1.8b displays the accumulated odometric error in the second room versus total distance traveled (after it was reset to zero)

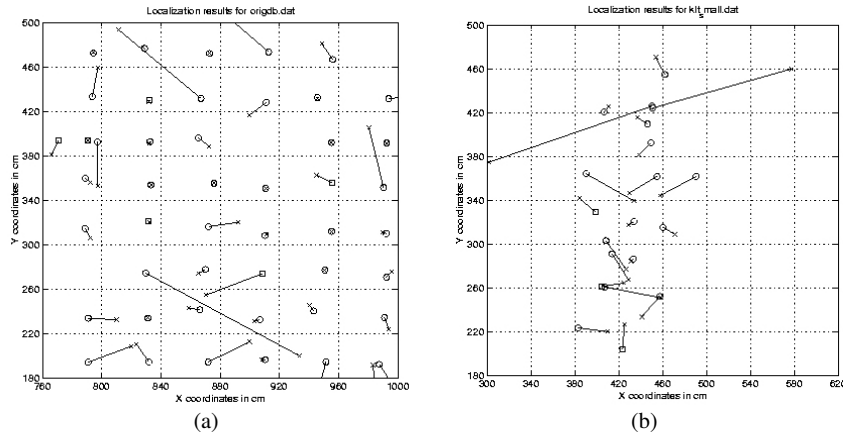


Fig. 1.9 (a) Tracker estimates vs. Vision-based estimates for training images. (b) Tracker estimates vs. Image-based estimates for a set of 21 random positions.

b) Tracking versus vision-based pose estimation: 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. Figure 1.9a shows the discrepancies between the pose estimates from the tracker (marked as circles) and the landmark-based vision pose estimator (marked as x) in

Room 2. At each position, the two 2-D projections of the alternative pose estimates are joined by a line. While the tracker is clearly more accurate, the quality of the landmark-based pose estimates is sufficient for situations where only one robot is present. There are a few large outliers that can be easily eliminated using prior pose estimates.

Our final stage of this experiment involved navigating the robot to a series of random positions and acquiring image- and tracker-based pose estimates, which are plotted together in Fig. 1.9b. This final experiment illustrates the use of a multi-sensor estimator in removing outliers. Assuming that the tracker-based position is correct, the mean error in the image-based estimate was 33cm, a large part of which can be attributed to the two significant outliers from nearly the same position. The purpose of this experiment is to verify the accuracy of pure vision based localization. Clearly, combining probabilistically the vision based estimated and an odometry based prior estimate results in a more accurate pose estimation.

1.6.2 Experiment 2

A second experiment was performed where the two robots explored a single large room. At the outset, one robot remained stationary while the other used a seed-spreader exploration procedure [21] across the floor, taking image samples at 25cm intervals, and in four orthogonal viewing directions, two of which are illustrated in Fig. 1.10. In this experiment the goal was to map a larger open area and to obtain maps in different orientations.



Fig. 1.10 Opposing views of the lab as seen by the exploring robot.

As before, the trajectory of the exploratory robot was defined at the outset by a user. However, as the robot explored, 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 Fig. 1.11a, 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 Fig. 1.11b.

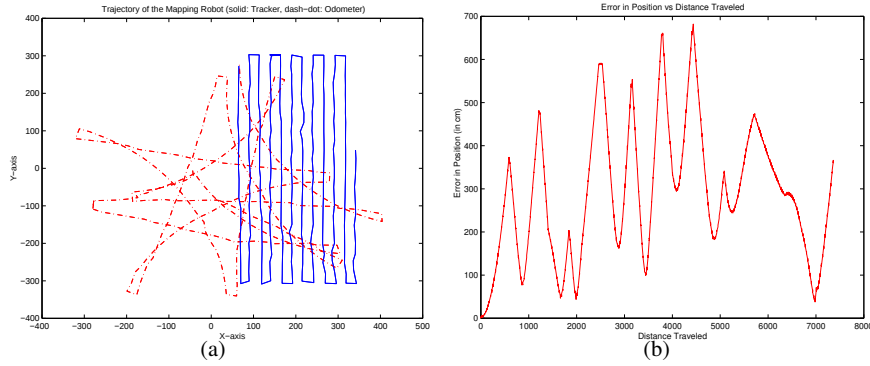


Fig. 1.11 In this experiment the robot took pictures in four orientation; the higher number of rotations increased non-linearly the odometric error. (a) Odometric (denoted by dash-dotted line) vs. Tracker-corrected (denoted by a solid line) trajectories of the robot. (b) Odometric error versus distance traveled.

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. Each thumbnail corresponds to an observation of the landmark. The relative position of the thumbnail in the image corresponds to the physical location from which it was acquired (that is, the positions of the thumbnails constitute an overhead view of locations visited by the robot). Note that in some locations there is no observation, as the tracker was not able to locate a visual match to the landmark.

Again, the final stage of our 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, 4 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 89 image-based estimates of high confidence are plotted with their associated tracker-based estimates in Fig. 1.12a. Assuming that the tracker-based position is correct, the mean error in the image-based estimate was 17cm,

(7.7cm in the x direction vs. 15cm in the y direction). The larger error in the y direction corresponds to the fact that the camera was pointed parallel to the positive y axis, and changes in observations due to forward motion are not as pronounced as changes due to side-to-side motion. The smallest absolute error was 0.49cm, which is insignificant compared to the “ground truth” error, and the largest error was 76cm. Note that most of the larger errors occur for large values of y . This is due to the fact that the camera was closest to the wall it was facing at these positions y , and as has been mentioned, tracking scene features over 25cm pose intervals became difficult.

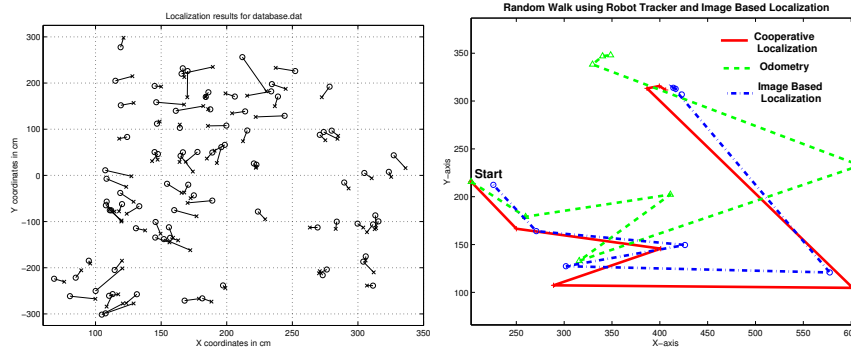


Fig. 1.12 (a) Tracker estimates vs. Image-based estimates for a set of 93 random positions. (b) The trajectory of the moving robot based on odometry estimates (triangles connected with a dashed line), the robot tracker cooperative localization ('+' connected with a solid line) and the image based localization ('o' connected with a dash-dotted line).

1.6.2.1 Random Walk

Figure 1.12b presents a random walk of the exploring robot through the mapped environment. The robot starts at a random location (marked as a “*”). Initially the odometry estimate is set to the value of the robot tracker estimate at that starting position, the pose estimate from the vision based system is approximately 30cm to the right of the robot tracker estimate. The robot took seven random steps and the three estimated trajectories are presented in Fig. 1.12b. First the odometer estimate (marked as triangles connected with a dashed line) is plotted; second, the robot tracker estimate (marked as “+” connected by a solid line), and third the visual pose estimator results (marked as “o” connected with a dash-dotted line). The robot tracker estimate provides a close approximation to ground truth at the end of the random walk the disparity between the robot tracker and the visual pose estimator is 17.5cm and between the robot tracker and the odometer is 68cm. The much higher disparity is a result of an increase in the accumulated error in orientation. The goal of this experiment was to evaluate the accuracy of the vision based localization; as

noted earlier the fusion of odometry data and the visual pose estimation resulted in higher positioning accuracy.

1.7 Conclusions

We introduced a method for the automatic mapping of an 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 implementation we presented 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 validate the utility of a set of learned landmarks for localization when the second robot cannot be deployed. This demonstrates conclusively that the virtual tether provides more accurate pose estimates, and hence a more accurate appearance-based map, than could be achieved with the robots operating independently.

Our approach does not eliminate positional error for the mapping robot but significantly reduces it. While the base robot does not move, the pose uncertainty of the mapping robot is equal to the uncertainty of the robot tracker sensor (a few centimeters) for an area as large as $225m^2$ for a robot tracker with range $12m$. Every time the base robot moves, its positional uncertainty increases by the positional uncertainty of the mapper plus the uncertainty of the robot tracker sensor. An analysis of the accumulation of the uncertainty is beyond the scope of this paper. In many environments, environmental features may be deemed accurate enough to be used to assist in localization. Such a SLAM formulation would further improve performance.

Our work demonstrates how co-operative localization can be employed as a mechanism for constructing a new map representation of the environment. One aspect of the problem that we did not consider is that of including feedback from the visual representation to further augment the robot’s pose estimate as it explores. This approach would be an example of simultaneous localization and mapping (SLAM) using multiple sources of information. We will consider this problem in future work.

The particular map we produce, an appearance-based representation of the environment, allows a single robot to accurately estimate its position on subsequent visits to the same area. While such single-robot pose estimates are not as accurate as when two robots are used, their accuracy is substantially ameliorated by the fact that two robots were used in the initial mapping stage. The use of an appearance-based model obviates most dependences on the particular geometry of the local environment. In principle, if a pair of robots was used in this subsequent stage the accuracy of the estimates could be further improved, but the extent of this advantage remains to be determined.

It would appear that these advantages become even more profound if more than two robots are used for position estimation and mapping. In the particular algorithmic scheme the use of many more robots would be an issue, but it seems that several

feasible solutions can be formulated; we hope to examine this problem in the future.

References

1. Balch, T., Arkin, R.C.: Communication in reactive multiagent robotic systems. *Autonomous Robots* **1**(1), 27–52 (1994)
2. Burgard, W., Fox, D., Moors, M., Simmons, R., Thrun, S.: Collaborative multi-robot exploration. In: *Proc. IEEE Int. Conf. on Robotics and Automation*, pp. 476–481 (2000)
3. Cao, Y.U., Fukunaga, A.S., Kahng, A.B., Meng, F.: Cooperative mobile robotics: Antecedents and directions. In: *Proc. IEEE/RSJ IROS*, vol. 1, pp. 226–234 (1995)
4. Choset, H., Burdick, J.: Sensor based planning, part ii: Incremental construction of the generalized voronoi graph. In: *Proc. IEEE Conf. on Robotics and Automation*, pp. 1643–1648 (1995)
5. Davison, A.: Real-time simultaneous localisation and mapping with a single camera. In: *Proc. IEEE Int. Conf. on Computer Vision* (2003)
6. Dellaert, F., Burgard, W., Fox, D., Thrun, S.: Using the condensation algorithm for robust, vision-based mobile robot localization. In: *IEEE Computer Society Conf. on Computer Vision and Pattern Recognition* (1999)
7. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society series B* **39**, 1–38 (1977)
8. Dudek, G., Jenkin, M.: *Computational Principles of Mobile Robotics*. ISBN: 0521568765. Cambridge University Press (2000)
9. Dudek, G., Jenkin, M., Milios, E., Wilkes, D.: A taxonomy for multi-agent robotics. *Autonomous Robots* **3**, 375–397 (1996)
10. Fenwick, J.W., Newman, P.M., Leonard, J.J.: Cooperative concurrent mapping and localization. In: *Proc. IEEE Int. Conf. on Robotics and Automation*, vol. 2, pp. 1810–1817 (2002)
11. Fox, D., Burgard, W., Kruppa, H., Thrun, S.: A probabilistic approach to collaborative multi-robot localization. *Autonomous Robots* **8**(3), 325–344 (2000)
12. Fox, D., Burgard, W., Thrun, S.: Active markov localization for mobile robots. *Robotics and Autonomous Systems (RAS)* **25**, 195–207 (1998)
13. Fox, D., Ko, J., Konolige, K., Limketkai, B., Schulz, D., Stewart, B.: Distributed multi-robot exploration and mapping. *Proceedings of the IEEE* **94**(7), 1325–1339 (2006)
14. Hagen, S., Krose, B.: Towards global consistent pose estimation from images. In: *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 466–471 (2002)
15. Howard, A., Mataric, M.J., Sukhatme, G.S.: Localization for mobile robot teams using maximum likelihood estimation. In: *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 434–459 (2002)
16. Kuipers, B., Byun, Y.T.: A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations. *Robotics and Autonomous Systems* **8**, 46–63 (1991)
17. Kurazume, R., Hirose, S.: An experimental study of a cooperative positioning system. *Autonomous Robots* **8**(1), 43–52 (2000)
18. Leonard, J.J., Durrant-Whyte, H.F.: Mobile robot localization by tracking geometric beacons. *IEEE Transactions on Robotics and Automation* **7**(3), 376–382 (1991)
19. Leonard, J.J., Feder, H.J.S.: A computationally efficient method for large-scale concurrent mapping and localization. In: *Robotics Research: The 9th Int. Symposium*, pp. 169–176 (2000)
20. Lu, F., Milios, E.: Optimal global pose estimation for consistent sensor data registration. In: *Int. Conf. in Robotics and Automation*, vol. 1, pp. 93–100. IEEE (1995)
21. Lumelsky, V., Mukhopadhyay, S., Sun, K.: Sensor-based terrain acquisition: The ‘sightseer’ strategy. In: *Proc. IEEE Conf. on Decision and Control Including The Symposium on Adaptive Processes*, vol. 2, pp. 1157–1161 (1989)

22. Mataric, M.: Environmental learning using a distributed representation. *IEEE Int. Conf. on Robotics and Automation* **1**, 402–406 (1990)
23. Montemerlo, M., Thrun, S., Koller, D., Wegbreit, B.: FastSLAM: A factored solution to the simultaneous localization and mapping problem. In: *Proc. AAAI National Conf. on Artificial Intelligence* (2002)
24. Murphy, R.R., Ausmus, M., Bugajska, M., Ellis, T., Johnson, T., Kelley, N., Kiefer, J., Pollock, L.: Marsupial-like mobile robot societies. In: *Agents*, pp. 364–365 (1999)
25. Nayar, S.K., Murase, H., Nene, S.A.: Learning, positioning, and tracking visual appearance. In: *Proc. IEEE Conf. of Robotics and Automation*, pp. 3237–3244 (1994)
26. Pourraz, F., Crowley, J.L.: Continuity properties of the appearance manifold for mobile robot position estimation. In: *Proc. IEEE Computer Society Conf. on Pattern Recognition Workshop on Perception for Mobile Agents* (1999)
27. Rekleitis, I.M.: Cooperative localization and multi-robot exploration. Ph.D. thesis, School of Computer Science, McGill University (2003)
28. Rekleitis, I.M., Dudek, G., Milios, E.: Multi-robot collaboration for robust exploration. In: *Proc. Int. Conf. in Robotics and Automation*, pp. 3164–3169 (2000)
29. Rekleitis, I.M., Dudek, G., Milios, E.: Multi-robot collaboration for robust exploration. *Annals of Mathematics and Artificial Intelligence* **31**(1–4), 7–40 (2001)
30. Rekleitis, I.M., Dudek, G., Milios, E.: Probabilistic cooperative localization and mapping in practice. In: *Proc. IEEE Int. Conf. on Robotics and Automation*, pp. 1907–1912 (2003)
31. Rekleitis, I.M., Dudek, G., Milios, E.E.: On multi agent exploration. In: *Proc. Vision Interface*, pp. 455–461 (1998)
32. Roumeliotis, S.I., Bekey, G.A.: Bayesian estimation and kalman filtering: A unified framework for mobile robot localization. In: *Proc. IEEE Int. Conf. on Robotics and Automation*, pp. 2985–2992 (2000)
33. Roumeliotis, S.I., Bekey, G.A.: Collective localization: A distributed kalman filter approach to localization of groups of mobile robots. In: *Proc. IEEE Int. Conf. on Robotics and Automation*, pp. 2958–2965 (2000)
34. Roumeliotis, S.I., Bekey, G.A.: Distributed multirobot localization. *IEEE Transactions on Robotics and Automation* **18**(5), 781–795 (2002)
35. Se, S., Lowe, D.G., Little, J.: Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks. *Int. Journal of Robotics Research* **21**(8), 735–758 (2002)
36. Sim, R., Dudek, G.: Learning generative models of scene features. *Int. Journal of Computer Vision* **60**(1), 45–61 (2004)
37. Thrun, S., Fox, D., Burghard, W.: A probabilistic approach to concurrent mapping and localization for mobile robots. *Autonomous Robots* **5**, 253–271 (1998)
38. Tinbergen, N.: *The animal in its world; explorations of an ethologist, 1932-1972*. Harvard University Press (1972)
39. Ulrich, I., Nourbakhsh, I.: Appearance-based place recognition for topological localization. In: *Proc. IEEE Int. Conf. on Robotics and Automation*, vol. 2, pp. 1023–1029 (2000)
40. Wurm, K., Dornhege, C., Eyerich, P., Stachniss, C., Nebel, B., Burgard, W.: Coordinated exploration with marsupial teams of robots using temporal symbolic planning. In: *Proc. IEEE/RSJ Int.l Conf. on Intelligent Robots and Systems* (2010)