

# Examining Exploratory Trajectories for Minimizing Map Uncertainty

Robert Sim and Gregory Dudek  
{simra,dudek}@cim.mcgill.ca  
Centre for Intelligent Machines, McGill University  
3480 University St., Montreal, Canada H3A 2A7

## Abstract

We examine the problem of minimizing uncertainty in the automated construction of a visual map of an unknown environment. Our work is motivated by the idea that a robot’s exploration policy can impact the accuracy of the resulting map, and we seek to determine a policy that optimizes a trade-off between accuracy and efficiency. We are further motivated by the specific requirements of our map representation, which learns a set of implicit models of visual features. Such a representation precludes the instantiation of explicitly parameterized landmarks, such as those employed in standard concurrent mapping and localization frameworks. This paper examines a parameterized family of spiral trajectories and determines parameterizations that yield reliable maps. We present experimental results demonstrating the map construction framework and discuss the implications for future work.

## 1 Introduction

This paper considers the problem of determining a trajectory for automatically constructing a visual map of an unknown environment. In particular, we examine a parameterized family of spiral-shaped trajectories and examine the properties of the generated map, including accuracy and coverage, dependent on the trajectory parameterization. Our work is motivated by the assumption that map uncertainty can be minimized without resort to expensive update methods by collecting observations in a principled manner. We will demonstrate that, using a conventional Kalman Filter parameterized only in the robot’s pose, an accurate map can be constructed.

An important aspect of our work is the visual map representation [18]. Visual maps encode visual properties of the environment in the image domain. The encodings are implicit in nature, and hence are not easily represented in a traditional concurrent mapping and localization framework (CML), particularly those approaches where spatial domain features are inferred. A visual map is constructed by tracking salient image features over a set of training images and then computing generative models of the features as functions of the robot’s pose. The implicit representation of the feature models poses a difficult challenge for accurate map construction— depending on the interpolating scheme employed to construct the generative model, there is a tension between the tendency to linearize away the non-linear aspects of feature behavior and the tendency to model the training data with too much specificity, resulting in instability. Regardless of which end of the spectrum the model favors, it is often the case that an implicit model will “believe” too strongly in the accuracy of the training data, and while there are a variety of methods available for smoothing and removing outliers in the training set, our goal is to ensure a high standard of accuracy at the outset.

This paper builds on our recent work in which a set of hand-crafted exploratory policies were examined for their accuracy, coverage and efficiency [19]. The primary results from that work indicated that the most accurate exploratory policy was also the most inefficient. The policy in question involved exploring at increasing distances along a series of rays emanating from a home position (Figure 1a)). After each ray is explored, the robot returns to the home position, re-localizing against the current map as it travels. By contrast, the most efficient policy was a series of concentric circles (Figure 1b)). While not the least

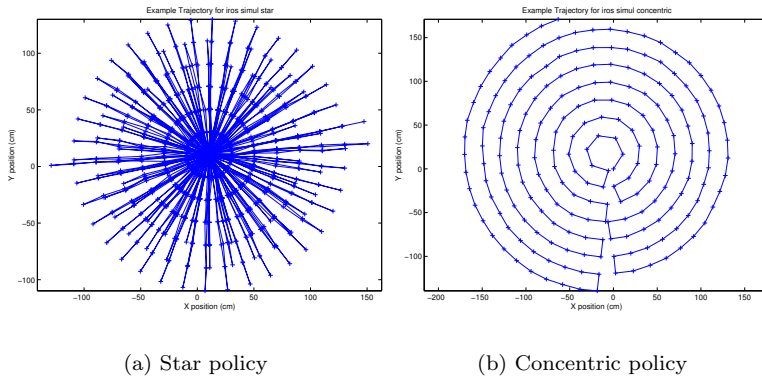


Figure 1: Example Star and Concentric exploratory policies

accurate of the policies examined, this policy demonstrated a clear tendency to propagate and amplify errors over time, resulting in a map that was accurate near the home position but increasingly inaccurate as the circles grew. These two policies can be thought of as lying at opposite extremes of a family of parametric curves— at one end, the robot returns to the home position at a maximal frequency, and at the other the robot never returns. The goal of this paper is to consider one such family of curves and find a reliable medium between the needs for accuracy and efficiency.

The remainder of this paper will consider our problem in greater depth, first examining related work, followed by a description of the visual map framework and our exploration framework. We will then present our experimental approach and results, followed by a discussion of the results and directions for future work.

## 2 Related Work

Our work is an instance of the problem of concurrent mapping and localization (CML), also known as simultaneous localization and mapping (SLAM). This problem has received considerable attention in the robotics community [20, 7, 21, 23, 2], primarily in the context of computing range-based maps with spatially localized features. The state of the art in CML can be broadly subdivided into one of two approaches (and various hybrids). One family of methods collects measurements and incrementally builds the map while the robot moves (i.e. in an on-line fashion). Usually the map is represented as a set of landmarks derived from a range sensor, and a Kalman filter is employed to minimize the total uncertainty of the robot pose and the individual landmark positions [8, 6]. These techniques differ from earlier Kalman filters employed for localization (c.f. [20, 7]) in that the landmark positions, as well as the robot pose, are being estimated and updated over time. While there exist approximation techniques for reducing the computational expense (cf [10]), each update in the standard on-line approach is quadratic in the number of landmarks.

The second family of methods for CML involves first collecting measurements and then post-processing them in a batch. The standard post-processing method is to employ Expectation Maximization (EM), again to minimize the total uncertainty of robot poses and landmark positions [21]. One goal of our work is to develop an on-line exploration method which maximizes the accuracy of the map without resort to expensive map updating. While outside the scope of this paper, this result can in turn be employed as a reliable prior for subsequent EM-style post-processing.

Maps based on visual information have also been examined. Nayar, *et al* pioneered the application of principal components analysis (PCA) to construct an appearance-based map that enables a homing behavior for robotic navigation [12]. Pourraz and Crowley considered the stability of PCA-based methods for navigation [13], and Jugessur and Dudek looked at voting-based methods to make appearance-based methods robust to changes in the scene or illumination [5]. Vision has also been employed for constructing geometric maps, which can then be used in a more traditional CML context. Se *et al* extract stereo-based

landmarks using a scale-invariant filter [16], and Davison and Kita considered the problem of actively servoing a stereo head for landmark acquisition as a robot traverses uneven terrain [3]. Finally, Dellaert *et al* take advantage of environmental invariants, such as a planar ceiling, to construct a mosaic-like map by registering an ensemble of images [4].

Of particular relevance to this paper is the problem of planning a trajectory for minimizing uncertainty while maximizing the utility of the observed data. MacKay considered the problem of optimally selecting sample points in a Bayesian context for the purposes of inferring an interpolating function [9]. Whaite and Ferrie employed this approach as motivation for their 'curious machine', a range-finder object recognition system that selected new viewing angles in order to optimize information gain [22], and Arbel and Ferrie further applied this approach to appearance-based object models, selecting the viewing angle that maximized the ability to discriminate between objects [1]. In the context of mobile robotics, Moorehead *et al*, considered the problem of maximizing information gain from multiple sources in order to control an exploratory robot [11]. Finally, Roy *et al* applied this principle to the problem of robotic path planning, instantiating the problem as a partially observable Markov decision process [15]. An important result from that work was the observation that a full solution to the problem is NP-hard.

Our work is based on the visual map framework described by Sim and Dudek [18], and employs the exploration framework described in our previous examination of exploration trajectories [19]. In the following sections we present the details of these approaches.

### 3 Visual Maps

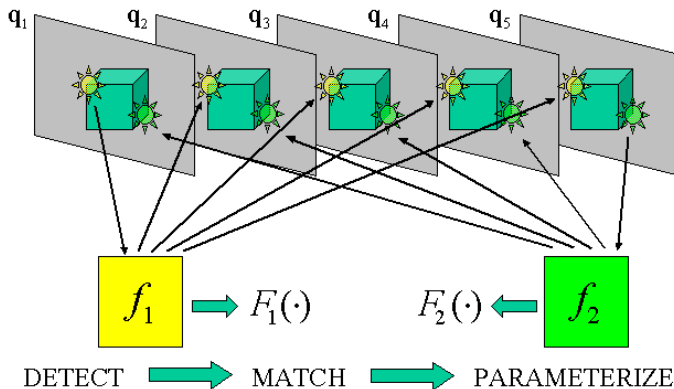


Figure 2: Landmark learning framework: Salient features are detected in the input images and tracked across the ensemble. The resulting feature sets are subsequently parameterized as functions  $F_i(\cdot)$  of the robot pose.

Our visual map representation employs the landmark learning framework described in prior work [18]. We review it here in brief and refer the reader to the cited work for further details.

The object of visual mapping is to learn a set of image-domain features of a scene, and describe them using a generative model so that they can be used to predict maximum-likelihood observations from arbitrary camera poses. The features are initially selected using a model of visual saliency, and subsequently tracked over the pose space. The resulting models are then cross-validated in order to select only those features that demonstrate stability.

The framework operates as follows: Assume for the moment that the robot has collected an set of observations of a scene with ground-truth position information associated with each image. A corner detector is applied to a selection of the images to select an initial set of candidate features [17]. The selected candidates are then tracked across the ensemble of images by maximizing the correlation of the

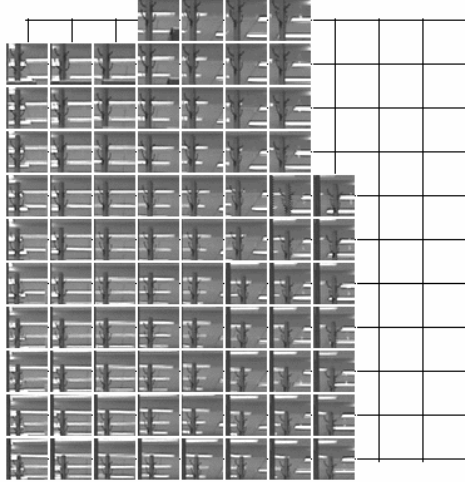


Figure 3: A set of observations of an extracted scene feature. The grid represents an overhead view of the pose space of the camera, and feature observations are placed at the pose corresponding to where they were observed.

local intensity image of the feature. Figure 3 depicts the result of tracking one feature across an image ensemble, wherein the local feature intensity image is depicted at the pose from which it was observed.

The resulting sets of tracked features are subsequently employed to construct an interpolation framework for generating novel feature observations from arbitrary views. The observations themselves are represented as the position of the feature in the image,  $\mathbf{z} = [x \ y]^T$ , and the interpolator is constructed using simple bilinear interpolation between neighboring observations in the Delaunay triangulation of the observation poses. This approach differs somewhat from the feature parameterization employed in [18], which computes a radial basis network interpolator of a wider variety of feature properties. In practice, an arbitrary interpolation scheme can be employed and in this work we employ a triangulation-based approach for reasons of efficiency. However, this approach does make the interpolation more susceptible to outlier observations. In order to evaluate the features and guard against outliers, the resulting models are validated using leave-one-out cross-validation.

## 4 Exploration Framework

We have briefly described how a visual map can be constructed from an ensemble of images of the environment acquired from known poses. In our prior work, these poses have typically been measured by hand or using an observing robot [14]. One motivation for the current work is to automate this data collection task using a single robot.

We have adapted the Extended Kalman Filter (EKF) localization framework described in the seminal papers by Smith *et al* and Leonard and Durrant-Whyte as the basis for our exploration framework [20, 7]. While these papers assumed a geometric representation of the environment, the visual map representation instead employs landmark observations in the image domain. It should be noted that, unlike EKF implementations deployed for CML which encode both robot pose and landmark position parameters, the only parameters maintained in our implementation are those of the robot pose. Given that the EKF has been studied extensively, we repeat here only those aspects of our implementation that are particular to our work. Note that this description also appears in our previous work on exploration trajectories [19].

At each time step  $k$ , the robot executes an action  $\mathbf{u}(k)$ , and takes a subsequent observation  $\mathbf{z}$ . The plant model is updated from  $\mathbf{u}$  according to the standard EKF formulation, and a set of matches to known features  $\mathbf{z}_i$  are extracted from the observed image.

For each successfully matched feature, a predicted observation  $\hat{\mathbf{z}}_i$  is generated using the visual map

and the current pose estimate, and the innovation  $v_i(k + 1)$  is computed

$$v_i(k + 1) = \mathbf{z}_i(k + 1) - \hat{\mathbf{z}}_i(k + 1) \quad (1)$$

The innovation covariance requires estimation of the Jacobian of the predicted observation given the map and the plant estimate. We approximate this Jacobian as the gradient of the nearest face of the model triangulation and define it as  $\nabla \mathbf{h}_i$ . Defined as such, the innovation covariance follows the standard observation model:

$$\mathbf{S}_i(k + 1) = \nabla \mathbf{h}_i \mathbf{P}(k + 1 | k) \nabla \mathbf{h}_i^T + \mathbf{R}_i(k + 1) \quad (2)$$

where  $\mathbf{P}$  is the pose covariance following the action  $\mathbf{u}$ , and  $\mathbf{R}$  is the cross-validation covariance associated with the learned feature model. It is important to note that  $\mathbf{R}$  serves several purposes— it is simultaneously an overall indicator of the quality of the interpolation model, as well as the reliability of the matching phase that led to the observations that define the model; finally it also accommodates the stochastic nature of the sensor.

## 4.1 Outlier Detection

Feature correspondence takes place once an observation is obtained. However, there may be outlier matches that must be filtered out. As such, we employ the gating procedure described in [7], with the additional constraint that the gating parameter  $g$  is computed adaptively. Specifically, we accept feature observations that meet the constraint

$$v_i(k + 1) \mathbf{S}_i^{-1}(k + 1) v_i^T(k + 1) \leq g^2 \quad (3)$$

where

$$g^2 = \max(g_{base}, \bar{g} + 2\sigma_g) \quad (4)$$

and  $g_{base}$  is a user defined threshold, and  $\bar{g}$  and  $\sigma_g$  are the average and standard deviation of the set of gating values computed for each feature observation (that is, the left-hand side of Equation 3). This selection of  $g$  allows the filter to correct itself when several observations indicate strong divergence from the predicted observations— indicating a high probability that the filter has diverged and affording the opportunity to correct the error.

## 4.2 Map Update

Given the set of gated observations, the EKF is updated according to the standard formulation, whereby the set of filtered innovation measurements is compounded into a single observation vector and a least-squares solution is computed for the Kalman gain. Combined with the plant model, a pose estimate and associated covariance are obtained. Once an updated pose estimate is available, the successfully matched features are inserted into the visual map, using the estimated pose as their observation pose. It should be noted that we also insert those observations that were invalidated by the gating procedure. We take this approach because it serves to increase the cross-validation covariance associated with the mis-matched feature, thereby reducing its influence for future localization. At the end of the exploration procedure, only those features that serve to match reliably *and* localize reliably can be selected and retained.

## 5 Parameterized Trajectories

In our previous work, we examined a set of hand-crafted trajectories based on an intuitive sense of how the robot might be able to minimize uncertainty while exploring [19]. In this paper, we will formalize our approach and examine an analytic family of trajectories, parameterized over a single parameter, that aims to capture the variety of properties that are important for accurate and efficient exploration. The

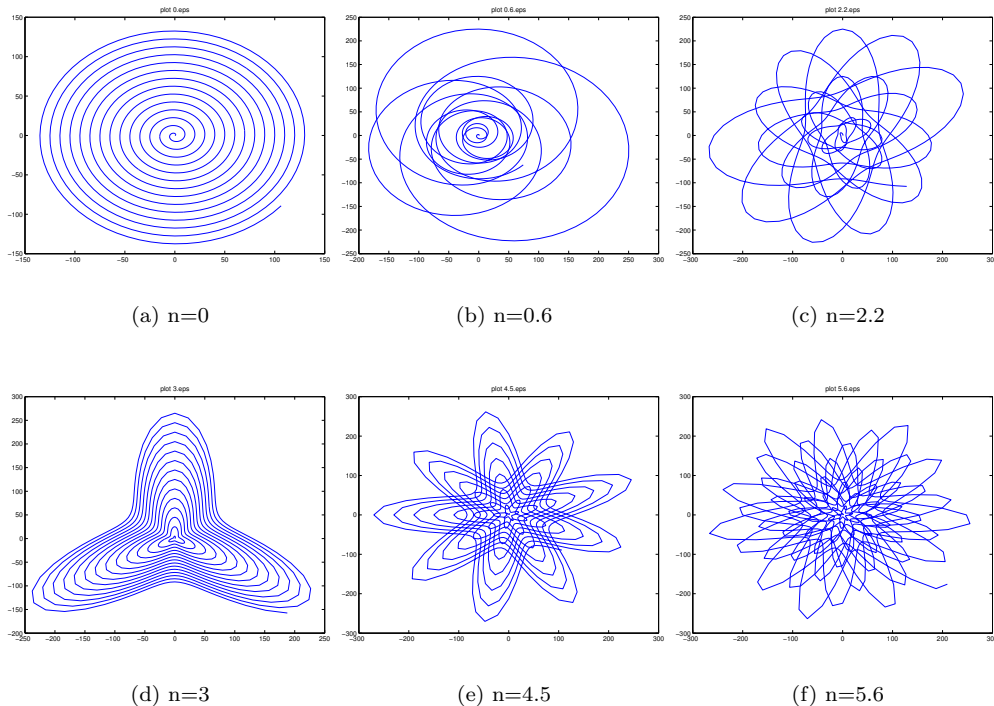


Figure 4: Sample trajectories for a variety of values of  $n$ .

specific parametric curve that we examine is expressed as the distance  $r$  from the origin as a function of time:

$$r_n(t) = \frac{kt}{2 + \sin nt} \quad (5)$$

where  $k$  is a dilating constant that is fixed for our experiments and  $n$  parameterizes the curve to control the frequency with which the robot moves toward the origin. Some examples of the curve for a variety of values of  $n$  are shown in Figure 4. Note that in the extreme cases, the curve never moves toward the origin ( $n = 0$ ), or will do so with very high frequency ( $n \rightarrow \infty$ ). Also of interest are integral values of  $n$ , where the curve never self-intersects, and has  $n$  distinct lobes. Finally, note that from an efficiency standpoint, the new space covered as a function of  $\theta$  decreases roughly monotonically as  $n$  increases, since for larger  $n$  the robot spends an increasing amount of time in previously explored territory.

## 6 Experimentation

### 6.1 Setup

We chose to run our experiments in a simulated environment in order to obtain accurate ground truth. The environment is composed of a single 1200cm by 600cm rectangular room, with images from a real laboratory environment texture-mapped on to the walls. The robot has a ring of sixteen evenly spaced sonar sensors which are employed solely for detecting collisions. The robot’s odometry model is set to add normally distributed zero-mean, 1% standard deviation error to translations and normally distributed zero-mean, 2% standard deviation error to rotations. Each observation is collected by placing a simulated camera at the ground truth pose of the robot, and snapping two images, one along the global  $x$  axis and one along the  $y$  axis. It is assumed that in a real-world setting the camera has the ability to align itself using a procedure which is external to the robot drive mechanism, possibly using a compass and pan-tilt unit or an independent turret, such as that which is available on a Nomad 200 robot. A single

observation is defined as the composite image obtained by tiling the two images side by side. Figure 5 illustrates a typical image returned by the camera in one direction in the simulated environment.

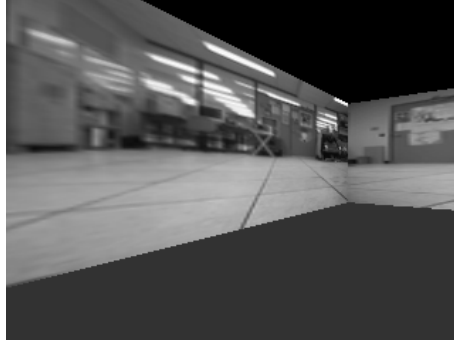


Figure 5: Simulated camera view.

The experiments were conducted as follows: for values of  $n \in [0.0, 8.0]$  at increments of 0.1, the robot was placed at the center of the room, and the trajectory  $r_n(t)$  was executed over five degree increments in  $t$  for 1000 time steps (whereby one time step involved a rotation followed by a translation). The constant  $k$  in Equation 5 was set to 20cm. At each pose, an observation was obtained and the Kalman Filter was updated. The visual map was updated whenever the filter indicated that the robot was more than 6.7cm from the nearest observation in the visual map. The ground-truth pose, the filter pose and the control inputs were recorded for each pose along the trajectory.

## 6.2 Results

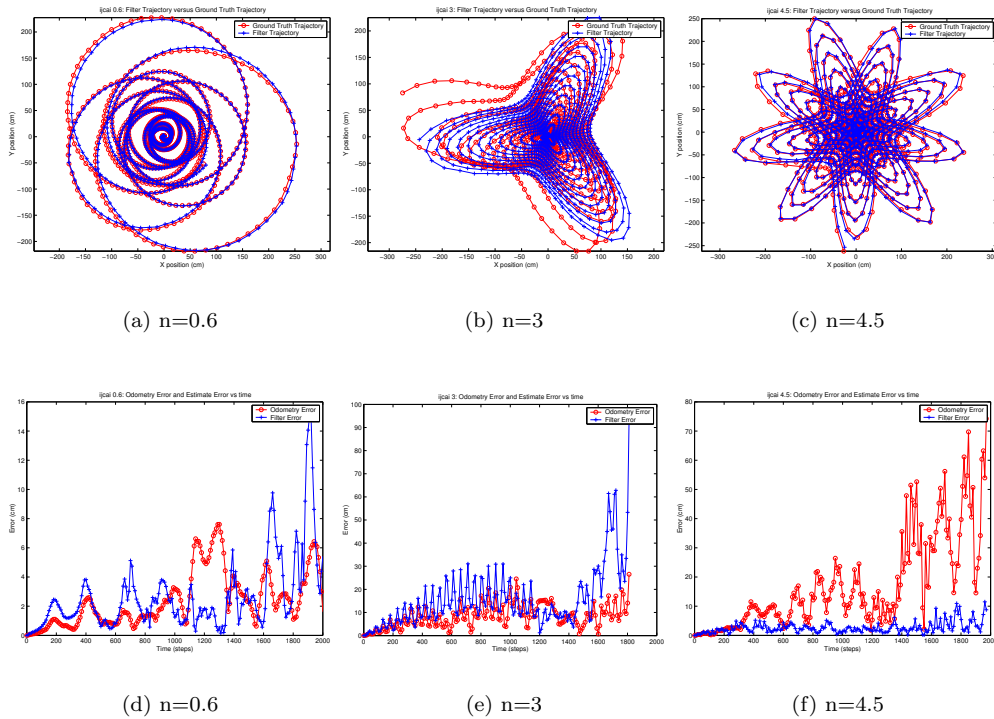
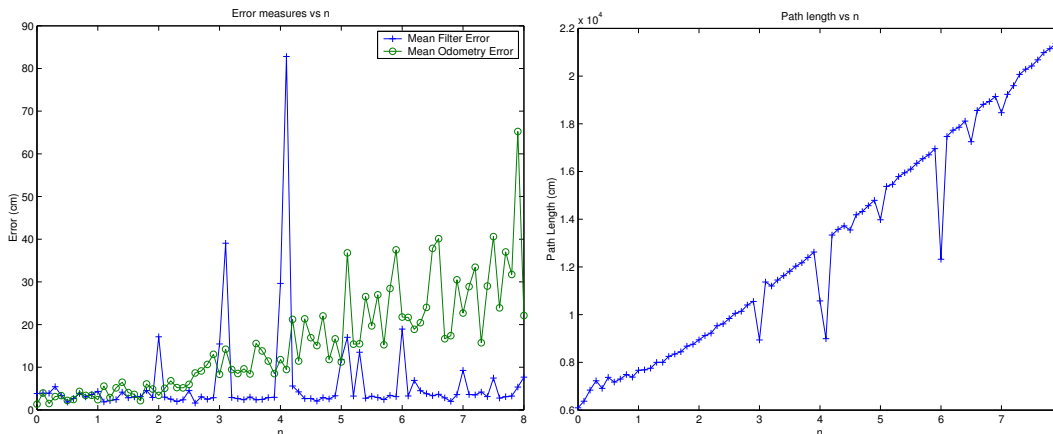


Figure 6: (a,b,c): Filter vs ground truth trajectories for various  $n$ . (d,e,f): Filter error and odometry error versus time for various  $n$ .

Figure 6a),b),c) depicts a selection of the ground-truth trajectory plotted against the filter trajectory for different values of  $n$ . The disparity between the two trajectories is an indicator of the accuracy of the visual map, since the poses of the images inserted into the visual map correspond to the filter poses. Given that small rotation errors near the beginning of the trajectory can lead to large errors at the edges of the map, even if the map itself is conformal, we adjusted the orientation of the filter trajectory around the starting pose to find the best fit against the ground-truth.



(a) Mean Filter Error and Mean Odometry Error vs  $n$ .

(b) Trajectory length vs  $n$ .

Figure 7: Mapping accuracy and efficiency as a function of  $n$ .

Figure 6d),e),f) depicts the filter error versus ground truth, as well as the odometry error versus ground truth over time for the corresponding values of  $n$ . From these figures, one can observe that for some values of  $n$ , odometry out-performs the filter, whereas for other values the filter tracks ground truth more accurately.

For each value of  $n$  it is possible to compute the mean filter and odometric error over the entire trajectory. Figure 7a) plots the mean error values for odometry and the filter as a function of  $n$ . It is interesting to note that as  $n$  increases the odometry error tends to increase, due to the increasing total magnitude of rotations performed by the robot, but the filter error remains roughly constant. This suggests that mapping accuracy is roughly independent of the choice of  $n$ . Note, however, the prominent spikes in the filter error corresponding to neighborhoods of integral values of  $n$ . These values correspond to trajectories that never self-intersect, or demonstrate a high degree of parallelism with nearby observations. As such, it appears that errors will propagate significantly or the filter may even diverge when insufficient constraints are available between neighboring paths (an instance of the aperture problem). The lone exception to this trend is the value for  $n = 0$ . In this particular case, however, the small amount of rotation at each time step leads to a well-constrained plant model in the Kalman Filter.

Finally, Figure 7b) depicts the length of each trajectory as a function of  $n$ . The trajectory length is an approximate measure of the inefficiency of the trajectory for exploration, since the radius of the convex hull of the explored space is bounded from above by  $kt_{max}$ , where  $t_{max}$  is the maximal time value, a constant across our experiments. Periodic minima in the trajectory length correspond to points where the exploration was terminated prematurely because the robot was unable to safely continue. In all of these cases, the filter estimate had diverged significantly from ground truth. As expected, increasing values of  $n$  lead to increased inefficiency.



## 7 Discussion and Future Work

We have presented an analysis of a family of parametric curves and its suitability for generating exploration trajectories for solving the concurrent mapping and localization problem, as it pertains to the construction of a visual map of an unknown environment. This study was conducted under the premise that mapping uncertainty can be managed by selecting an appropriate exploratory trajectory. The results demonstrate that the parametric family under consideration is in general a suitable choice for exploration in that for most parameterizations the error in the generated map is small relative to odometric error. However, it was interesting to note a subset of parameterizations that systematically led to divergence. A somewhat surprising result was that mapping error was relatively small for the simple spiral trajectory corresponding to  $n = 0$ . While this can be explained in part by the small amount of accumulated odometric error, it runs contrary to our prior findings. Nonetheless, our hypothesis is that if the exploration were to continue over a longer time interval, the filter will eventually diverge. Furthermore, the potential for the presence of obstacles in a real environment would impose circumnavigation requirements which could introduce the kinds of odometric errors that might lead to divergence. As such, our conclusions are that it is worth the additional effort of “re-homing” the robot from time to time, corresponding to employing a larger value of  $n$ .

An important aspect of the exploratory trajectories considered in this work is that they are computed independently of the state of the robot’s map or the uncertainty in the filter. An obvious direction for future investigation is the question of determining a locally optimal trajectory given the current map and filter state. Furthermore, while it is beyond the scope of this paper, the family of curves studied pose the problem that the robot needs to determine a starting position at the outset, not to mention that the rotational symmetry of the curves make them less suitable for irregularly shaped environments. One potential solution to this problem is to partition the environment and build a separate visual map for each partition. Our ongoing work reflects this goal of dealing with larger environments.

## References

- [1] Tal Arbel and Frank P. Ferrie. Viewpoint selection by navigation through entropy maps. In *Seventh IEEE International Conference on Computer Vision*, Kerkyra, Greece, 1999. IEEE Press.
- [2] Hans Blaasvaer, Paolo Pirjanian, and Henrik I. Christensen. Amor: An autonomous mobile robot navigation system. In *IEEE, Int. Conference on Systems, Man, and Cybernetics*, pages 2266–2271, 1994.
- [3] A. J. Davison and N. Kita. Sequential localisation and map-building for real-time computer vision and robotics. *Robotics and Autonomous Systems*, 36(4):171–183, 2001.
- [4] 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*. IEEE Press, June 1999.
- [5] G. Dudek and D. Jugessur. Robust place recognition using local appearance based methods. In *International Conference on Robotics and Automation*, San Francisco, April 2000. IEEE Press.
- [6] Jose Guivant, Eduardo Nebot, and Hugh Durrant-Whyte. Simultaneous localization and map building using natural features in outdoor environments. In *Sixth International Conference on Intelligent Autonomous Systems*, Italy, 2000.
- [7] J. Leonard and H. F. Durrant-Whyte. Simultaneous map building and localization for an autonomous mobile robot. In *Proc. IEEE Int. Workshop on Intelligent Robots and Systems*, pages 1442–1447, Osaka, Japan, November 1991.
- [8] John J. Leonard and Hans Jacob S. Feder. A computationally efficient method for large-scale concurrent mapping and localization. In John Hollerbach and Dan Koditschek, editors, *Robotics Research: The Ninth International Symposium*, London, 2000. Springer-Verlag.

- [9] D. MacKay. Information-based objective functions for active data selection. *Neural Computation*, 4(4):590–604, 1992.
- [10] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM: A factored solution to the simultaneous localization and mapping problem. In *Proceedings of the AAAI National Conference on Artificial Intelligence*, Edmonton, Canada, 2002. AAAI.
- [11] Stewart Moorehead, Reid Simmons, and William Red L. Whittaker. Autonomous exploration using multiple sources of information. In *IEEE International Conference on Robotics and Automation*, May 2001.
- [12] S.K. Nayar, H. Murase, and S.A. Nene. Learning, positioning, and tracking visual appearance. In *Proc. IEEE Conf on Robotics and Automation*, pages 3237–3246, San Diego, CA, May 1994.
- [13] F. Pourraz and J. L. Crowley. Continuity properties of the appearance manifold for mobile robot position estimation. In *Proceedings of the 2nd IEEE Workshop on Perception for Mobile Agents*, Ft. Collins, CO, June 1999. IEEE Press.
- [14] Ioannis Rekleitis, Robert Sim, Gregory Dudek, and Evangelos Miliotis. Collaborative exploration for the construction of visual maps. In *2001 IEEE/RSJ Conference on Intelligent Robots and Systems (IROS)*, Hawaii, October 2001.
- [15] Nicholas Roy, Wolfram Burgard, Dieter Fox, and Sebastian Thrun. Coastal navigation – robot motion with uncertainty. In *AAAI Fall Symposium: Planning with POMDPs*, 1998.
- [16] S. Se, D. Lowe, and J. Little. Vision-based mobile robot localization and mapping using scale-invariant features. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 2051–2058, Seoul, Korea, May 2001.
- [17] J. Shi and C. Tomasi. Good features to track. In *Proc. IEEE International Conf. Computer Vision and Pattern Recognition (CVPR)*. IEEE Press, 1994.
- [18] R. Sim and G. Dudek. Learning generative models of scene features. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Lihue, HI, December 2001. IEEE Press.
- [19] Robert Sim and Gregory Dudek. Effective exploration strategies for the construction of visual maps. Submission to 2003 IEEE/RSJ Conference on Intelligent Robots and Systems (IROS), March 2003.
- [20] R. Smith, M. Self, and P. Cheeseman. Estimating uncertain spatial relationships in robotics. In I.J. Cox and G. T. Wilfong, editors, *Autonomous Robot Vehicles*, pages 167–193. Springer-Verlag, 1990.
- [21] S. Thrun, D. Fox, and W. Burgard. A probabilistic approach to concurrent mapping and localization for mobile robots. *Machine Learning*, 31:29–53, 1998. also appeared in *Autonomous Robots* 5, 253–271.
- [22] P. Whaite and F. P. Ferrie. Autonomous exploration: Driven by uncertainty. In *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pages 339–346, Los Alamitos, CA, USA, June 1994. IEEE Computer Society Press.
- [23] B. Yamauchi, A. Schultz, and W. Adams. Mobile robot exploration and map building with continuous localization. In *Proc. IEEE Int. Conf. on Robotics and Automation*, pages 3715–2720, Leuven, Belgium, May 1998. IEEE Press.