Title

Adaptive Sampling for Estimating a Scalar Field using a Robotic Boat and a Sensor Network

Authors
Zhang, Bin
Sukhatme, Gaurav

Publication Date
2007-04-10

DOI
10.1109/ROBOT.2007.364041

Peer reviewed

2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007

Adaptive Sampling for Estimating a Scalar Field using a Robotic Boat and a Sensor Network

Bin Zhang and Gaurav S. Sukhatme

Abstract— **This paper introduces an adaptive sampling algorithm for a mobile sensor network to estimate a scalar field. The sensor network consists of static nodes and one mobile robot. The static nodes are able to take sensor readings continuously in place, while the mobile robot is able to move and sample at multiple locations. The measurements from the robot and the static nodes are used to reconstruct an underlying scalar field. The algorithm presented in this paper accepts the measurements made by the static nodes as inputs and computes a path for the mobile robot which minimizes the integrated mean square error of the reconstructed field subject to the constraint that the robot has limited energy. We assume that the field does not change when robot is taking samples. In addition to simulations, we have validated the algorithm on a robotic boat and a system of static buoys operating in a lake over several km of traversed distance while reconstructing the temperature field of the lake surface.**

I. INTRODUCTION

Sensor networks provide new tools for observing and monitoring the environment. In aquatic environments, accurately measuring quantities such as temperature, chlorophyll, salinity, and concentration of various nutrients is useful to scientists seeking a better understanding of aquatic ecosystems, as well as government officials charged with ensuring public safety via appropriate hazard warning and remediation measures.

Broadly speaking, these quantities of interest are scalar fields. Each is characterized by a single scalar quantity which varies spatiotemporally. A characteristic of such scalar fields is that the sensor readings are only valid locally. That is, the correlation between the sensors at different locations decreases rapidly with increasing distance between the sensor nodes. To estimate the scalar field where no sensor nodes are deployed, we need to interpolate the data. Intuitively, the more the readings near the location where a field estimate is desired, the less the reconstruction error. In other words, the spatial distribution of the measurements (the *samples*) affects the estimation error.

In many cases, it may not be feasible to move the static sensor nodes after deployment. In such cases, one or more mobile robots could be used to augment the static sensor network, hence forming a sensor-actuator network or a robotic sensor network. Static nodes and mobile robots both have their advantages and constraints. Static nodes generally consume less energy than mobile robots and their batteries last longer. However, static nodes cannot move and cannot change the spatial distribution of the sensor readings while mobile robots are able to move and take sensor readings at different locations.

The problem of adaptive sampling: An immediate question to ask is how to coordinate the mobile robots and the static nodes such that the error associated with the estimation on the scalar field is minimized subject to the constraint that the the energy available to the mobile robot(s) is bounded. Specifically, if each static node makes a measurement in its vicinity, and the total energy available to the mobile robot is known, what path should the mobile robot take to minimize the mean square integrated error associated with the reconstruction of the entire field? Here we assume that the energy consumed by communications and sensing is negligible compared to the energy consumed in moving the mobile robot. We also assume that the mobile robot can communicate with all the static nodes and acquire sensor readings from them. Finally, we focus on reconstructing phenomena which do not change temporally(or change very slowly compared to the time it takes the mobile robot to complete a tour of the environment).

The domain: In this paper, we develop a general solution to the above problem and test it on a particular set up designed to monitor an aquatic environment. The experimental set up is a systems of anchored buoys (the static nodes), and a robotic boat (the mobile robot) capable of measuring temperature and chlorophyll concentrations. This testbed is part of the NAMOS (Networked Aquatic Microbial Observing System) project (`http://robotics.usc.edu/˜namos`), which is used in studies of microbial communities in freshwater and marine environments [1], [2].

Contributions of this paper: In general, the optimality of the path for the mobile robot depends on the approach used for estimating the field. Model-based estimation (and hence optimal sampling design based on linear or non-linear models) is well studied [3]. In the environmental monitoring context a prior model is normally unknown and it might even be the goal of the project to learn a model from the data collected by the sensor network. Therefore, non-parametric estimation is appropriate. In this paper, we propose an adaptive sampling algorithm based on local linear regression [4], [5] which is guaranteed to be optimal in the sense of minimizing the integrated mean square error (IMSE) of the field reconstruction. The energy consumption model depends on the dynamics of the robotic boat. The adaptive sampling

This work is supported in part by the National Science Fundation (NSF) under grants CNS-0325875, IIS-0133947, CNS-0540420, CNS-0520305 and grants CCR-0120778, ANI-00331481 (via subcontract).

B. Zhang `binz@usc.edu` and G. S. Sukhatme `gaurav@usc.edu` are with the Robotic Embedded Systems Laboratory, Computer Science Department, University of Southern California, 941 W. 37th Place, Los Angeles, CA 90089, USA

algorithm does not depend on the energy consumption model but the optimal paths generated by the algorithm do. We show tests of our algorithm on the robotic boat (henceforth simply "the boat") executing optimal trajectories (exceeding an aggregate of 3 km in length over the course of the trials reported here) operating with data collected from the buoys.

This paper is organized as follows. We discuss related work in the following section. In section III, the adaptive sampling algorithm based on local linear regression is discussed without considering the constraints from the dynamics/kinematics of the boat. We introduce those constraints in section IV and propose a breadth first search algorithm for path planning. The experimental results are discussed in section V. Concluding remarks are in section VI

II. RELATED WORK

Adaptive sampling has been studied in sensor networks. Willett [6] proposed an algorithm based on wavelets for static sensor networks to extend the battery life of sensor nodes while providing enough estimation accuracy by trying to keep as many nodes in sleep mode as possible. Rahimi [7] proposed an algorithm for an actuated sensor network, which refines the field estimation by taking more readings in locations where estimation variance is high. This approach is similar to the adaptive sampling algorithm to be introduced next, but did not consider a dynamic or kinematic model of the mobile robot. Krause [8] proposed an near-optimal algorithm to solve a problem similar to ours. However, he assumed that many sensors are deployed at the beginning, the correlations between the value of the scalar field at different locations are learned from the data collected from the sensors and then the sensors are redeployed with some sensors removed.

Optimal experimental design is an area that is closely related to adaptive sampling; it has been well studied in the literature of statistics. The goal of optimal experimental design is to minimized the estimation error with given number of samples/readings to take. If the model of the phenomenon is known to be linear or can be transformed to linear, there exists an algorithm to find the optimal solution [3], [9]. If the model is known to be non-linear, a sequential design can be used [3]. If the model is unknown and no prior knowledge is assumed, the reasonable design is maximin or minimax design [9]. There are also studies on optimal design based on local linear regression [10], [11] and Kriging [12], [9]. All these studies on sampling design assume the samples to be taken manually and the cost related to the design is strictly a function of the number of samples instead of the path connecting all the sample locations.

A penalty associated with the path is considered in various incarnations of the robot exploration/mapping problems, which are well studied in the robotics literature [13], [14], [15], [16]. Roy [13] proposed an Entropy based approach for path planning so that better localization could be achieved. Sim [14] proposed an A-optimality criterion for SLAM and a Breadth First Search algorithm to find the approximate global optimal path. Mei [15] proposed a orientation based target selection strategy, where the energy cost for turning is incorporated in the optimal criterion. While there are similarities between the adaptive sampling design problem studied here and mapping/exploration problems it is important to note that there are some differences. In robot mapping/exploration the criterion of optimality is defined as the decrease in the uncertainty of the map, which normally assumes that the estimation bias is zero. In adaptive sampling for a scalar field, interpolation is necessary to estimate the value where no direct sensing information is available and hence bias is inevitable. Therefore, a new criterion is needed to define the optimality of the path.

III. ADAPTIVE SAMPLING BASED ON LINEAR LOCAL REGRESSION

A. Linear Local Regression

Non-parametric regression has been well studied in statistics and many methods have been proposed. Because it is easy to understand, analyze and implement on a computer, the kernel estimator is one of the most popular non-parametric estimators. Kernel estimators include Nadaraya-Waston, Gasser-Muller and local linear regression estimators. In this paper, local linear regression is used to estimate the phenomenon. The asymptotic properties of 1 dimensional local linear Regression were first studied by Fan [5] and later extended by Ruppert to the multi-dimensional case [4]. The main take away message from these two studies is that the Integrated Mean Square Error (IMSE) of the estimator is related to the second derivatives of the phenomenon investigated.

If $m(\mathbf{x})$ is the function to be estimated (reconstructed), we assume the following model

$$Y_i = m(\mathbf{X_i}) + v^{1/2}(\mathbf{X_i})\epsilon_i, \tag{1}$$

where $i = 1, ..., n$, d is the number of dimensions, $\mathbf{X_i}$ are $\Re^d$-valued predictor variables, Y_i are scalar response variables, $v(\mathbf{X_i}) = Var(Y|\mathbf{X} = \mathbf{x})$ is finite and the ϵ_i are mutually independent and identically distributed random variables with zero mean and unit variance and are independent of $\mathbf{X_i}$. Then the local linear regression estimator is the linear estimator that minimizes

$$\sum_{i=1}^{n}\{Y_i - \alpha - \beta^T(\mathbf{X}_i - \mathbf{x})K_H(\mathbf{X}_i - \mathbf{x})\}^2, \tag{2}$$

where H is a $d \times d$ symmetric positive definite matrix, K is a d-variate kernel with $\int K(\mathbf{u})d\mathbf{u} = 1$ and $K_H(\mathbf{u}) = |H|^{-1/2}K(H^{-1/2}\mathbf{u})$. $H^{1/2}$ is called the bandwidth matrix. Solving the previous optimization problem, we have the following LLR estimator:

$$\hat{m}(\mathbf{x}, H) = e_1 \cdot (X_x^T W_x X_x)^{-1} X_x^T W_x Y, \tag{3}$$

where $\mathbf{X_x} = \begin{bmatrix} 1 & (\mathbf{X}_1 - \mathbf{x})^T \\ \cdots & \cdots \\ 1 & (\mathbf{X}_n - \mathbf{x})^T \end{bmatrix}$, $Y = [Y_1, \cdots, Y_n]^T$, $W_x = diag\{K_H(\mathbf{X}_1 - \mathbf{x}), \cdots, K_H(\mathbf{X}_n - \mathbf{x})\}$ and $e_1 = [1, 0, \cdots, 0]$.

If the kernel satisfies $\mu_2(K)\mathbf{I} = \int \mathbf{u}\mathbf{u}^T K(\mathbf{u})d\mathbf{u}$ and $\int u_1^{l_1} \cdots u_d^{l_d} K(\mathbf{u})d\mathbf{u} = 0$ for all non-negative integers $l_1, \cdots, l_d$ such that their sum is odd, it has been proved that the estimation error associated with the local linear regression is given by the following equation [4]:

$$\begin{aligned}\text{MSE}\{\hat{m}(\mathbf{x}; H)\} &= \frac{R(K)v(\mathbf{x})}{n|H|^{1/2} f(\mathbf{x})} \\ &+ \frac{1}{4}\mu_2(K)^2 tr^2\{H\text{H}_m(\mathbf{x})\} \\ &+ \mathbf{o}_p\{n^{-1}|H|^{-1/2} + tr^2(H)\}. \quad (4)\end{aligned}$$

where $R(k) = \int K(u)^2 du$, $f(\mathbf{x})$ is the density function with $\int f(\mathbf{x})d\mathbf{x} = 1$, and $\text{H}_m(\mathbf{x})$ is the Hessian matrix of $m(\mathbf{x})$.

B. Optimal Design for Linear Local Regression

In equation 4, when $n|H|^{1/2}$ is big enough and H is small enough, the infinitesimal $\mathbf{o}_p\{n^{-1}|H|^{-1/2} + tr^2(H)\}$ is negligible and the IMSE can be approximated as following

$$\int \left\{ \frac{R(K)v(x)}{n|H|^{1/2} f(\mathbf{x})} + \frac{1}{4}\mu_2(K)^2 tr^2\{H\text{H}_m(\mathbf{x})\} \right\} d\mathbf{x} \quad (5)$$

If we assume the Hessian matrix of $m(\mathbf{x})$ is known, we can determine the optimal bandwidth and optimal density function by minimizing the IMSE with the constraint that $\int f(\mathbf{x})d\mathbf{x} = 1$. By applying the Lagrange-Euler differential equation, we have the optimal bandwidth and density function as follows:

$$h^* = \left(\frac{dR(K)v(\mathbf{x})}{nf(\mathbf{x})\mu_2(K)^2 tr^2\{\text{H}_m(\mathbf{x})\}} \right)^{\frac{1}{d+4}}, \quad (6)$$

$$f^*(\mathbf{x}) \propto (\mu_2(K) tr\{\text{H}_m(\mathbf{x})\})^{\frac{2d}{d+8}} (R(K)v(\mathbf{x}))^{\frac{4}{d+8}}, \quad (7)$$

where we assume the bandwidth matrix is defined as $H = h^2\mathbf{I}$. $f^*(\mathbf{x})$ is called the optimal design. When the cost of moving from one sample location to another is small compared to the cost of taking sensor readings, we can use $f^*(\mathbf{x})$ to generate the sample locations. Assume that there are a small number of initial sensor readings available, the Hessian Matrix can be estimated by using local polynomial regression. Then new samples can be drawn according to the optimal density function computed from the Hessian Matrix. Figure 1 compares the performance of the optimal design and a random design. The comparison is based on the water surface temperature data taken in the field using a raster scan of the environment. From the figure, it is obvious that the optimal designs outperform the random design.

IV. PATH PLANNING FOR THE MOBILE ROBOT

A. Objective Function for Path Planning

When samples are to be taken by a mobile robot, optimal design alone is not feasible because now the constraint is not the number of samples to be taken but the distance that the mobile robot can travel. This limits how many samples can be taken and where those samples can be taken. For instance, in the worst case, the sample locations generated by optimal design are all far away so that energy available to the mobile node is not enough even to reach one of those locations.

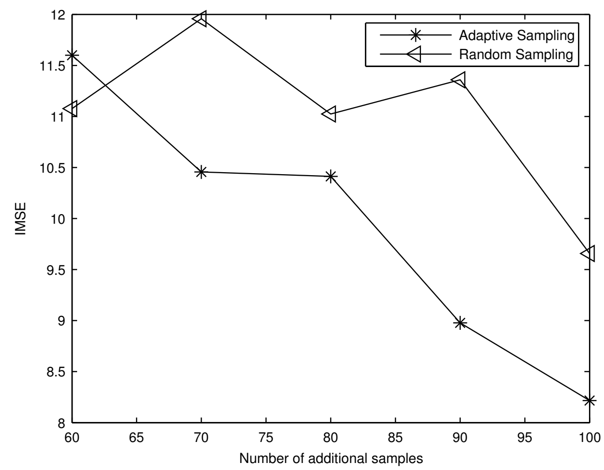


Fig. 1. A comparison between Adaptive Sampling and Random Sampling. No energy consumption model is considered here.

Assume that initially there are n_0 readings from static sensor nodes $(\mathbf{X_1}, Y_1), \ldots, (\mathbf{X_{n_0}}, Y_{n_0})$, the path of the mobile robot passes through the points $\mathbf{X}_{n_0+1}, \ldots, \mathbf{X}_{n_0+n}$, then the optimal path should minimize IMSE, which can be estimated as follows

$$\text{IMSE}(\mathbf{X}_1, \ldots, \mathbf{X}_{n0+n}) \propto \int \left(\frac{tr^d\{\text{H}_m(\mathbf{x})\}v^2(\mathbf{x})}{n^2 \hat{f}^2(\mathbf{x})} \right)^{\frac{2}{d+4}} d\mathbf{x}, \quad (8)$$

where $\hat{f}(\mathbf{x}) = n^{-1}\sum_{i=1}^{n} K_H(\mathbf{X_i} - \mathbf{x})$ is the estimation of density function. Similar to the information gain defined in robot exploration literature, we define the gain for each point as

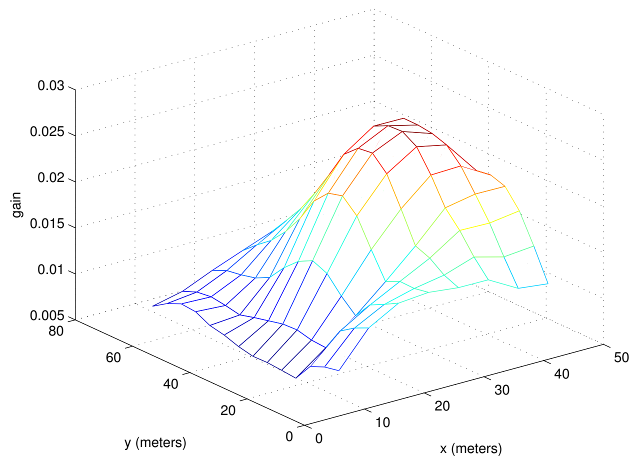
$$G(\mathbf{x}) = \text{IMSE}(\mathbf{X}_1, \ldots, \mathbf{X}_{n_0}) - \text{IMSE}(\mathbf{X}_1, \ldots, \mathbf{X}_{n_0}, \mathbf{x}), \quad (9)$$

and the gain of a whole path p as

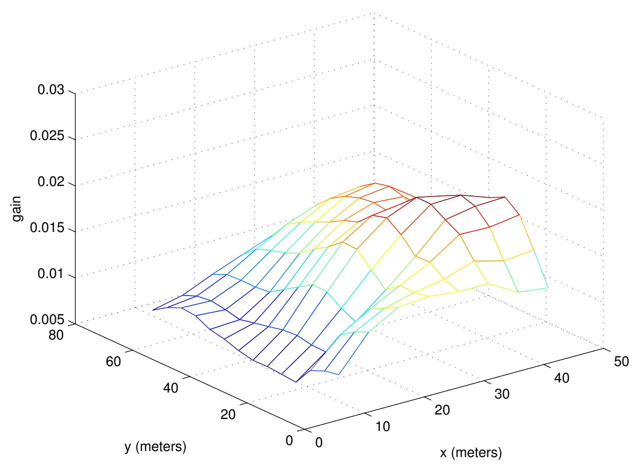
$$\begin{aligned}G(p) &= \text{IMSE}(\mathbf{X}_1, \ldots, \mathbf{X}_{n_0}) - \quad (10) \\ &\quad \text{IMSE}(\mathbf{X}_1, \ldots, \mathbf{X}_{n_0}, \mathbf{X}_{n_0+1}, \ldots, \mathbf{X}_{n_0+n}).\end{aligned}$$

Note that the gain of the whole path is not the sum of the gain of all points on the path because the gain at each location is not independent with each other. On the one hand, one new sampling point not only increases the sampling density at/near that point. On the other hand, MSE is not linear in the density function. The higher the initial density function at the location $\mathbf{x}$, the less the gain that is achieved by taking one more sample at $\mathbf{x}$. Figure 2 illustrates the effect of this dependency. Figure 2(a) shows the gain for each location after the initial sensor readings are taken and Figure 2(b) shows the gain for each location after one more sensor reading is taken at location (35,35). From the figure, we can see that not only is the gain at (35,35) decreased, the gains of the locations close to (35,35) also decrease. As a result, at some stage of the plan if two competing paths A and B reach exactly the same state (same location, same direction and same energy level), and A has higher gain than

B, one cannot prune B because A might already decrease the gains of other unvisited locations so much that B has the potential to catch up. This suggests that the state of the mobile robot depends on the whole path, not just current gain and position.



(a)



(b)

Fig. 2. (a) The gain of each location before taking one more reading at (35,35); (b) The gain of each location after taking one more reading at (35,35). Once a location has been visited it value (and the values of points in its vicinity) drops.

However, if the whole path is taken as the state, the size of the state space is exponential in the available energy and hence is computationally intractable. By examining the effect of dependency closely, we find that the effect is local and the range it affects depends on the bandwidth. Therefore, if we discretize the sensing field sparsely enough, we can approximate the gain of the path by using the sum of the gains of all the points on the path. In this case, the state can be represented by the position of the node, current energy available to the robot, and the gain achieved. Now the search space is linear in both the size of the environment and the energy level and a breadth first search can be applied to find the optimal path.

It is possible that when the mobile robot reaches the location with the maximum gain it stays in the same location thereafter since it can achieve maximum gain there. Because of the non-linear properties of the MSE, the gain achieved by taking a second reading at the same location is much less than the first one. So, we explicitly disallow the mobile robot to go back to the locations it already visited. We represent the state of the mobile robot as $S = (x, y, \theta, e)$, where x, y are the coordinates of the mobile robot, θ is its direction and e is the energy available to it. Note that θ and e do not have a direct effect on the estimation error, but they affect how many samples would be taken given that the initial energy available is fixed. Suppose $\{x_0,y_0,\theta_0,e_0\}$ is the initial state of the mobile robot, R_0, X_0, Y_0 are the initial sensor readings and corresponding coordinates, X_1, Y_1, Θ_1 are the sequence of stated the mobile robot should follow, i.e., the path, the breadth search algorithm is shown as follows:

Algorithm 1: Breadth First Search Algorithm for Path Planning

input : x_0, y_0, θ_0, e_0, X_0, Y_0, R_0
output: X_1, Y_1, Θ_1
Initialize all G(x, y, θ, e) = 0;
Q = MakeQueue();
enqueue(Q, x_0, y_0, θ_0, e_0);
while *Q is not empty* **do**
  x, y, θ, e = dequeue(Q);
  for *each vertices x', y', θ', e' adjacent to x, y, θ, e* **do**
    gain = ComputeGain(x', y', X_0, Y_0, R_0);
    if *G(x, y, θ, e) + gain $>$ G(x', y', θ', e') $\wedge$ $\{x'$, y', θ', $e'\}$ is not one of the ancestors of $\{x$, y, θ, $e\}$* **then**
      G(x', y', θ', e') = G(x, y, θ, e) + gain;
      ϕ(x', y', θ', e') = x, y, θ, e;
      enqueue (Q, x', y', θ', e') with priority G(x', y', θ', e');
    end
  end
end
$\{x, y, \theta, e\}$ = $argmax_{\{x,y,\theta,e\}} G(x, y, \theta, e)$;
while *x, y, θ, e is not $\{x_0, y_0, \theta_0, e_0\}$* **do**
  push x into X_1;
  push y into Y_1;
  push θ into Θ_1;
  $\{x$, y, θ, $e\}$ = ϕ(x, y, θ, e);
end
return X_1, Y_1, Θ_1;

B. Energy Consumption Model

In the algorithm described in the previous subsection, an energy consumption model is necessary to determine if two states are adjacent to each other. The energy model we use is based on the boat, which is part of the NAMOS

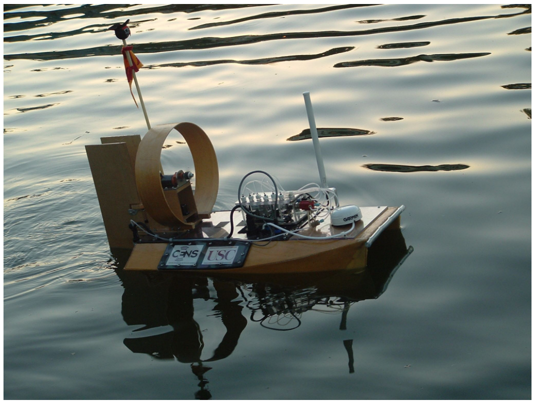


Fig. 3. The robotic boat in the NAMOS project.

system. NAMOS, Networked Aquatic Microbial Observing System [1], is a sensor-actuator network for freshwater and marine monitoring. The system consists of 10 static sensor nodes and one boat as shown in Figure 3, which is the mobile node we use in all tests reported here. Unlike wheeled mobile robots, the boat tends to keep moving when the propeller stops because of the low friction of the water. Additionally, the boat makes turns by using a rudder and hence it cannot turn in place. If the boat is moving at a certain speed, there exists a minimum turning circle and the radius depends on the speed of the boat.

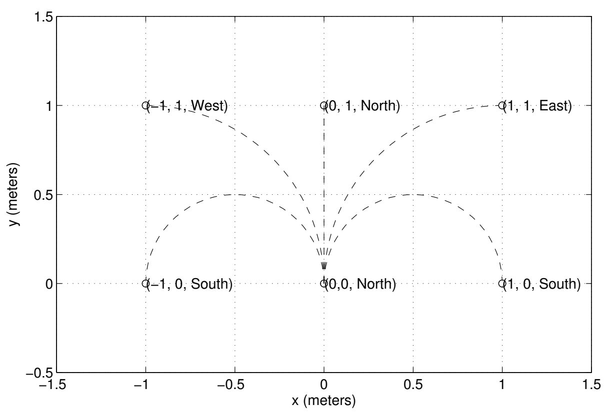


Fig. 4. The trajectories considered for the energy consumption model. See Table I

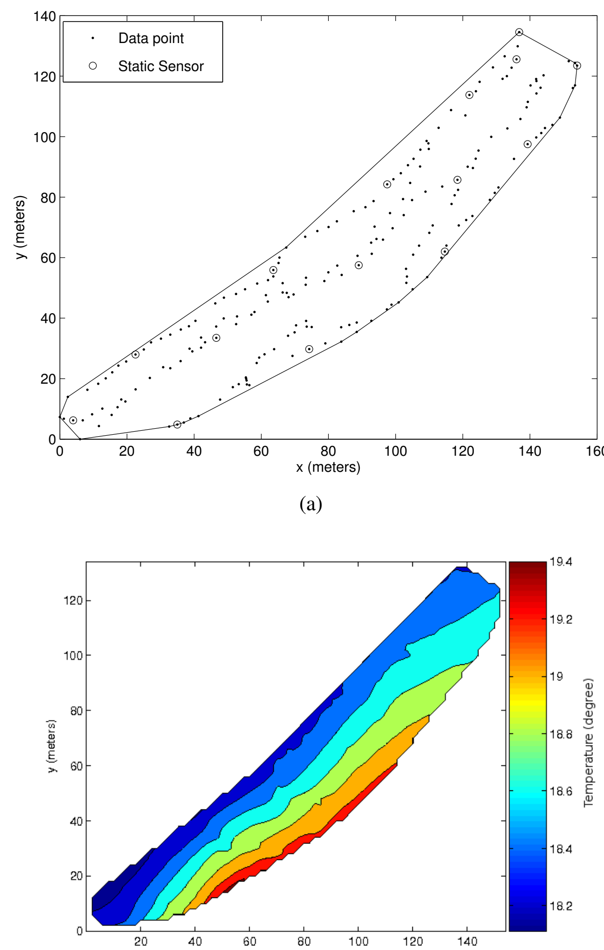
With the above limitations in mind, we propose the energy consumption model for the boat as depicted in Figure 4. The position of the robot is represented as (x, y, direction). At each location (x, y), the robot can be in 4 directions: North, East, South and West. From each state, there are 5 states the boat can transit to. For examples, if current state of the boat is (0, 0, North), it is able to transit to state (1, 0, South), (1, 1, East), (0, 1, North), (-1, 1, West), and (-1, 0, South). Note that the boat cannot transit directly to the states behind it, such as (-1, -1, South), but it is able to reach those state through other state, such as (-1, 0, South). We assume that the energy caused by state transition is proportional to the distance the robot traveled. The ideal curve connecting two states should be the one with minimum length but the minimum radius of curvature is not less than the radius of the minimum turning circle. In this paper, we use circular arc to approximate the optimal curve. The energy consumed when the robot transits from (0, 0, North) to other state is listed as Table I, where we assume the energy cost from (0, 0, North) to (0, 1, North) is 1 unit.

TABLE I

THE ENERGY CONSUMPTION TABLE

Destination State	Energy consumed
(1, 0, South)	1.6
(1, 1, East)	1.6
(0, 1, North)	1
(-1, 1, West)	1.6
(-1, 0, South)	1.6

V. EXPERIMENTS



(a)

(b)

Fig. 5. (a) The raster scan data set; (b) The interpolated temperature field by using the whole raster scan data set.

Our algorithm has been tested with the raster scan data taken by the robotic boat in Lake Fulmor, CA, during a period of 40 minutes in the afternoon of May 9th, 2006. Figure 5(a) shows the points in the entire raster scan data set. The solid curve is the boundary of the sensing field and the black dots are the data points. Temperature readings are used in the tests and Figure 5(b) shows the temperature field interpolated with the whole data set. We mine this data set (the ground truth) as follows to test the algorithm via a data-denial simulation. We assume that the boat is able to follow the optimal path generated by the adaptive sampling algorithm accurately. When one sampling location $\mathbf{x}$ is generated, we search in the raster scan data set for the point closest to $\mathbf{x}$ and take the corresponding temperature reading as the sensor readings at location $\mathbf{x}$. Once the process of sampling is done, the temperature field is estimated and the result is compared with the raster scan data instead of current temperature on the surface of the lake. As a result, even if the temperature field change during the time when the robotic boat is doing raster scan, the data set still can be used as taken from a static field. It is also assumed that there exist 15 static nodes to provide the initial data set and the locations of those static nodes are shown as small circles in Figure 5(a). Because we are going to use some of the temperature readings in the raster scan data set as the readings of the static sensors, the locations of the static nodes are picked so that they coincide with some of the points in the raster scan data set.

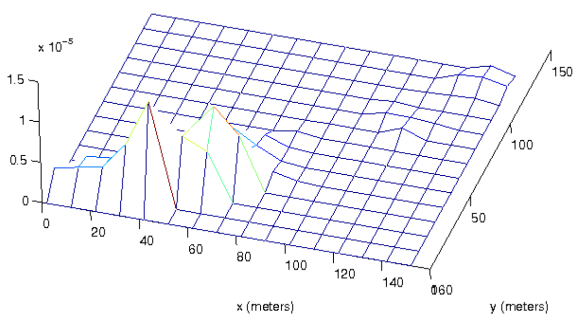


Fig. 6. The square of the trace of the estimated hessian matrix

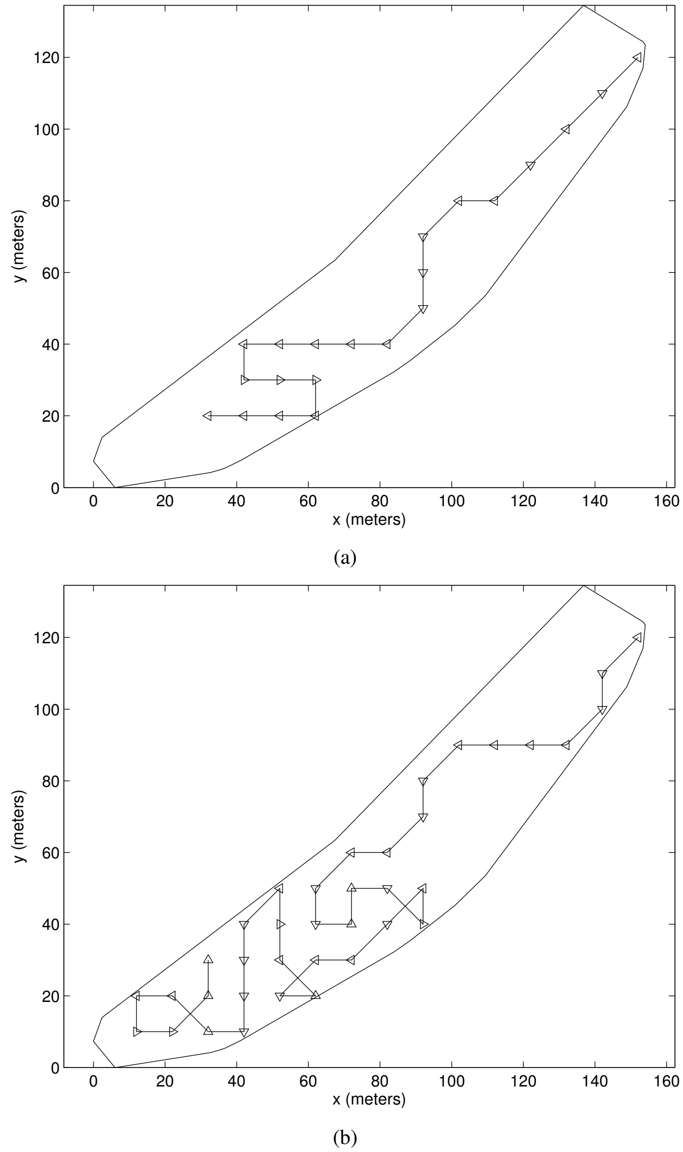


Fig. 7. Paths generated by adaptive sampling. (a) Initial energy is 25; (b) Initial energy is 50.

We assume that the boat starts from the top right corner of the lake. The initial direction of the boat is west, i.e., facing toward the center of the lake. Figure 6 shows the trace of the estimated Hessian matrix. The higher the trace, the higher the gain can be achieved by taking a reading there. As shown in Figure 6, the trace of the estimated Hessian matrix is higher in the bottom left part of the lake than in the top right part. As a result, the boat tends to take more samples in the bottom left part. the Figure 7(a) and Figure 7(b) show two typical paths generated by the algorithm. The triangles indicate the directions of the boat. The Figure 7(a) shows the path with initial energy 25 units and Figure 7(b) shows the path with the initial energy 50 units. When the initial energy is high, the path planner tends to allow the boat to wander a little on its way to the place with maximum gain so that it does not have to move back later. When the initial energy is low, the path planner requires the boat to go straight to those locations with maximum gain.

Our algorithm is compared to a random walk with same initial state. For each method, we generate a set of new data points and then estimate the temperature with the new data set together with the initial readings from the static sensors, which are still part of the raster scan data. Then the estimation is compared to the whole raster scan data set assuming that the raster scan is the ground truth. The IMSE is approximated by summing the square error at each data point of the raster scan. For each initial energy value, the random walk runs for 50 trials and the IMSE reported is the average over these 50 trials. Figure 8 shows the comparison of the IMSE between the two methods with the initial energy being varied from 5 to 60 units. From the figure, we can see that with the adaptive sampling algorithm, the IMSE is about 20% less than a random walk. Since $IMSE \propto n^{-\frac{2}{3}}$ for the best case in 2D problem, a 20% improvement is significant.

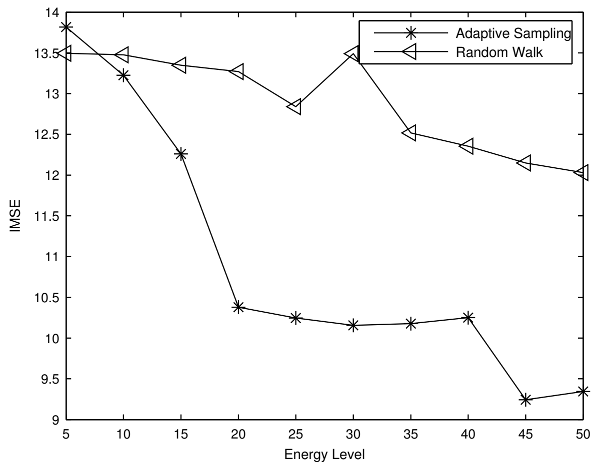


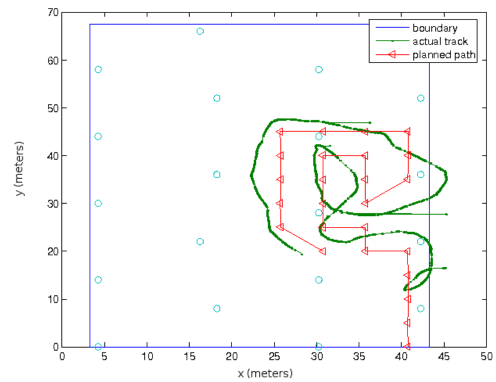
Fig. 8. Comparison between adaptive sampling and random walk

Our algorithm has also been tested on the boat in a lake in Echo Park in Los Angeles and the the target scalar field is the temperature field. Once again, we use the raster scan data as the ground truth. The boat was first used to quickly perform a raster scan in a 40m x 70m area, which is part of the lake. We assume that the temperature field did not change significantly during the period when the experiments are carried out. Figure 9 shows the results of two experiments. The solid lines in Figures 9(a) and 9(c) show the boundary of the test area and the circles are the locations where initial temperature readings are available.

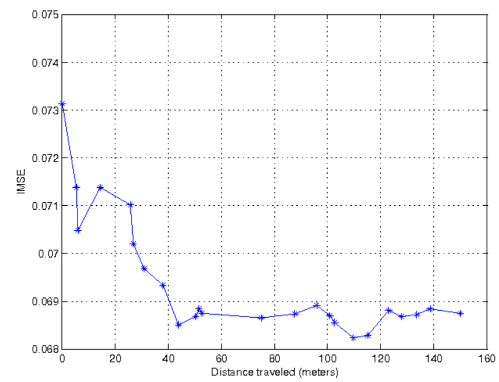
The boat starts from the right bottom corner facing north in the first experiment and top right corner facing west in the second experiment. The triangles in Figure 9(a) and Figure 9(c) are the planned locations where the boat should go and take temperature readings and the dots show the GPS record of the boat, which show the actual track. Note that the actual track did not follow the planned path strictly. This is due to the GPS resolution and the effect of the wind. Currently, there is no battery monitor on the boat and hence we cannot measure the energy consumed by the boat directly. However, during the experiments, the speed of the propeller on the boat is approximately constant. That is, the thrust generated by the propeller is approximately constant and hence we can approximate the energy consumption by measuring the distance the boat traveled. With the dense GPS readings along the track, it is easy to compute the distance. Figure 9(b) and Figure 9(d) show the IMSE of the estimation vs. the distance the boat traveled in both experiments respectively.

VI. CONCLUSIONS AND FUTURE WORK

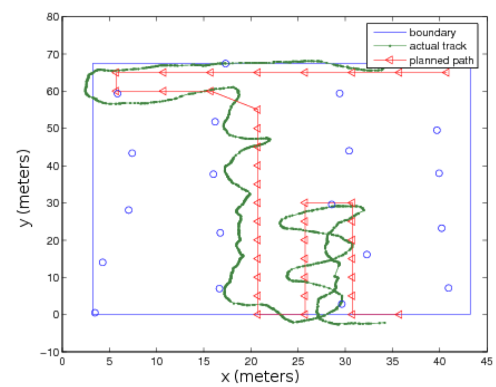
We have proposed an adaptive sampling algorithm for a mobile sensor network consisting of a set of static nodes and a mobile robot tasked to reconstruct a scalar field. Our algorithm is based on local linear regression. Sensor readings from static nodes (a set of buoys) are sent to the mobile robot (a boat) and used to estimate the Hessian Matrix of the scalar field (the surface temperature of a lake), which is directly



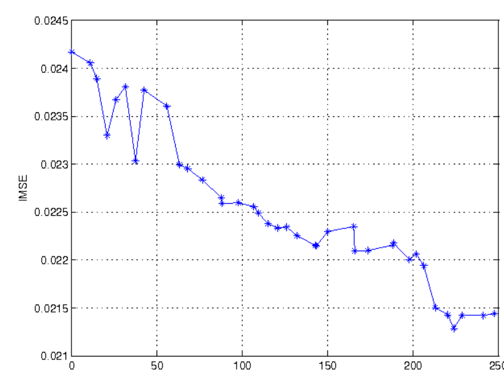
(a)



(b)



(c)



(d)

Fig. 9. (a) and (c) show the planned path and the GPS trace of the actual track the boat took. (b) and (d) show estimation error v.s. distance the boat traveled (the energy consumed by the boat is approximately proportional to the distance traveled).

related to the estimation error. Based on this information, a path planner generates a path for the boat such that the resulting integrated mean square error (IMSE) of the field reconstruction is minimized subject to the constraint that the boat has a finite amount of energy which it can expend on the traverse. Data from extensive (several km) traverses in the field as well as simulations, validate the performance of our algorithm.

We are currently working on how to determine the appropriate resolution to discretize the sensed field. One interesting observation from the simulations and experiments is that when the initial available energy is increased, the estimation errors decrease rapidly and level off instead of decreasing to zero. Theoretically, when the energy available to the mobile node increases, more sensor readings can be taken and hence the estimation errors should keep decreasing. By examining the path generated by the adaptive sampling algorithm, we found that when the initial energy is enough for the mobile node to go through all the 'important' locations, increasing the initial energy does not have much effect on the estimation error. We plan to investigate advanced path planning strategies and alternative sampling design strategies in future work.

VII. ACKNOWLEDGMENTS

The authors thank David Caron, Amit Dhariwal, Mark Hansen, Arvind Pereira, Carl Oberg, Stefanie Moorithi and Beth Stauffer for their comments and help with the experiments.

REFERENCES

[1] A. Dhariwal, B. Zhang, C. Oberg, B. Stauffer, A. Requicha, D. Caron, and G. S. Sukhatme, "Networked aquatic microbial observing system," in *the Proceedings of the IEEE International Conference of Robotics and Automation (ICRA)*, May 2006.

[2] G. S. Sukhatme, A. Dahriwal, B. Zhang, C. Oberg, B. Stauffer, and D. Caron, "The design and development of a wireless robotic networked aquatic microbial observing system," *Environmental Engineering Science*, 2007.

[3] S. Silvey, *Optimal Design*. Chapman and Hall, 1980.

[4] D. Ruppert and M. P. Wand, "Multivariate locally weighted least squares regression," *The Annals of Statistics*, vol. 22, no. 3, pp. 1346–1370, 1994.

[5] J. Fan, "Local linear regression smoothers and their minimax efficiences," *The Annals of Statistics*, vol. 21, no. 1, pp. 196–216, 1993.

[6] R. Willett, A. Martin, and R. Nowak, "Backcasting: Adaptive sampling for sensor networks," in *Proceedings of the third international symposium on Information processing in sensor networks*, 2004, pp. 124–133.

[7] M. H. Rahimi, "Bioscope: Actuated sensor network for biological science," Ph.D. dissertation, Computer Science Department, Viterbi School of Engineering, USC, 2005.

[8] A. Krause, C. Guestrin, A. Gupta, and J. Kleinberg, "Near-optimal sensor placements: Maximizing information while minimizing communiction cost," in *Proccedings of the Fifth International Conference on Information Processing in Sensor Networks (IPSN'06)*, Apr 2006, pp. 2–10.

[9] A. Atkinson and A. Donev, *Optimum Experimental Designs*. Oxford Science Publications, 1992.

[10] M.-Y. Cheng, P. Hall, and D. M. Titterington, "Optimal design for curve estimation by local linear smoothing," *Bernoulli*, vol. 4, pp. 3–14, 1998.

[11] V. V. Fedorov, G. Monteiedra, and C. J. Nachtsheim, "Design of experiments for locally weighted regression," *Journal of Statistical Planning and Inference*, no. 81, pp. 363–382, 1999.

[12] W. G. Muller, *Collecting Spatial Data*. Physica-Verlag, 2001.

[13] N. Roy, W. Burgard, D. Fox, and S. Thrun, "Coastal navigation – robot motion with uncertainty," in *In AAAI Fall Symposium: Planning with POMDPs*, vol. 1, 1998, pp. 35–40.

[14] R. Sim and N. Roy, "Global a-optimal robot exploration in SLAM," in *the Proceedings of the IEEE/RSJ International Conference of Robotics and Automation (ICRA)*, 2005, pp. 661–666.

[15] Y. Mei, Y.-H. Lu, C. S. G. Lee, and Y. C. Hu, "Energy efficient mobile robot exploration," in *the Proceedings of the IEEE International Conference of Robotics and Automation (ICRA)*, May 2006, pp. 505–551.

[16] B. Yamauchi, "A frontier-based approach for autonomous exploration," in *the Proceedings of the IEEE International Conference of Robotics and Automation (ICRA)*, Jul 1997, pp. 146–151.