Maximum Mutual Information Principle for Dynamic Sensor Query Problems

Emre Ertin[1], John W. Fisher[2], and Lee C. Potter[1]

[1] The Ohio State University, Columbus, OH
{ertin.1,potter}@osu.edu,
[2] MIT Artificial Intelligence Laboratory, Cambridge, MA
fisher@ai.mit.edu

Abstract. In this paper we study a dynamic sensor selection method for Bayesian filtering problems. In particular we consider the distributed Bayesian Filtering strategy given in [1] and show that the principle of mutual information maximization follows naturally from the expected uncertainty minimization criterion in a Bayesian filtering framework. This equivalence results in a computationally feasible approach to state estimation in sensor networks. We illustrate the application of the proposed dynamic sensor selection method to both discrete and linear Gaussian models for distributed tracking as well as to stationary target localization using acoustic arrays.

1 Introduction

There has been renewed interest in the notion of deploying large numbers of networked sensors for applications ranging from environmental monitoring to surveillance to "intelligent" rooms(c.f. [2]). Envisioned are smart sensor nodes with on-board sensing, computation, storage and communication capability. Such sensor networks simultaneously present unprecedented opportunities and unique challenges in collaborative signal processing. A particular challenge in the wireless sensor network setting is the need for distributed estimation algorithms which balance the limited energy resources at a node with costs of communication and sensing.

If one considers the distributed tracking problem, for example, it is not hard to imagine that one need not incorporate *every* sensor measurement in order to compute a reliable, if not optimal, estimate of the state of an object (or more properly the posterior distribution thereof). This is particularly true in the case where sensors have a limited field of regard with limited overlap between sensors. Distributed processing strategies that use a subset of sensor measurements directly mitigate the volume of inter-node communication thereby conserving power. The challenge is to decide in an intelligent manner which sensor measurements to use.

In the context of just such a scenario, Zhao *et al.* [1] recently suggested a novel approach, the Information-Driven Sensor Querying (IDSQ) algorithm, as a means of selecting the "best" sensor measurement for updating the posterior

F. Zhao and L. Guibas (Eds.): IPSN 2003, LNCS 2634, pp. 405–416, 2003.
© Springer-Verlag Berlin Heidelberg 2003

belief of an object's state. In that work, a utility measure of a node measurement was proposed based on an estimate of the *expected* posterior state distribution conditioned on the, as yet unobserved, measurement at that node. This led to a direct method of selecting which node to query.

In this paper we further investigate aspects of one of the information utility functions suggested in [1,3], specifically state uncertainty as quantified by conditional entropy. We begin by formulating the problem in a Bayesian estimation framework (as is commonly done) and decomposing state estimation into prediction (prior belief) and update (posterior belief) steps. We first show that, not surprisingly, functions which attempt to select the next sensor measurement based on *expected* posterior belief do nothing more than exploit information already contained in the prior belief as both are the same prior to taking a measurement. Consequently, utility functions based on *expected* posterior beliefs are more properly cast as utility functions on the prior belief (i.e. the belief over the current set of measurements). Next we consider the expected posterior uncertainty as quantified by conditional entropy (conditioned on previous measurements and a single new measurement) indexed by sensors. We show that this utility function simplifies to selecting the sensor measurement which has maximum mutual information with the object state at the next time step. The primary consequence of this analysis is that the utility function can be computed in a lower-dimensional space and, importantly, in a computationally feasible manner.

We present three experimental examples. The first example uses a simple discrete model to illustrate the maximum mutual information principle. The second example discusses the application of the maximum mutual information based sensor selection method to linear Gaussian Models. The third example is a simulation study of a stationary target localization problem using an acoustic array.

2 Bayesian Filtering with Dynamic Sensor Selection

We adopt a probabilistic state space model for the tracking problem. The state of the target at time step (t) is denoted by $x^{(t)}$. In this paper we will assume that the state space for the tracking problem can be approximated with a finite state space $\{x_i\}_{i=1}^{N}$. The sensor network consists of M sensors. Each sensor can be queried to provide a noisy measurements $z_j^{(t)}$ of the state of the target. The state transition and observation model is given as:

$$\begin{aligned} x^{(t+1)} &= F(x^{(t)}, v^{(t)}) \\ &\Rightarrow q(x^{(t+1)}|x^{(t)}) \end{aligned} \tag{1}$$

$$\begin{aligned} z_j^{(t)} &= H_j(x^{(t)}, w^{(t)}) \\ &\Rightarrow f_j(z_j^{(t)}|x_j) \end{aligned} \tag{2}$$

where F and H_j are arbitrary functions of the state and unknown disturbance variables $v^{(t)}$ and $w^{(t)}$. The state space model suggests a conditional probability distribution $q(x^{(t+1)}|x^{(t)})$ for the target state at time $(t+1)$ and a conditional probability density $f_j(z_j^{(t)}|x_j)$ for the j'th sensors measurement.

The Bayesian filtering solution recursively calculates degree of belief in a state $x^{(t+1)}$, given the sensor measurements. The prediction step computes the prior belief in state $x^{(t+1)}$ before a measurement is taken at $(t+1)$:

$$p(x^{(t+1)}|\overline{z^{(t)}}) = \sum_i q(x^{(t+1)}|x_i^{(t)})p(x_i^{(t)}|z^{(t)}) \ . \tag{3}$$

The update step computes the posterior belief in state $x^{(t+1)}$ after the measurement at $(t+1)$:

$$p(x^{(t+1)}|\overline{z^{(t+1)}}) = \frac{f_j(z^{(t+1)}|x^{(t+1)})p(x^{(t+1)}|\overline{z^{(t)}})}{g_j(z^{(t+1)}|\overline{z^{(t)}})} \ , \tag{4}$$

where $\overline{z^{(t)}}$ denotes the measurements $\{z^{(1)}, z^{(2)}, \ldots, z^{(t)}\}$ up to time (t). The normalization constant $g_j(z^{(t+1)}|\overline{z^{(t)}})$ can be computed using:

$$g_j(z^{(t+1)}|\overline{z^{(t)}}) = \sum_i f_j(z^{(t+1)}|x_i)p(x_i|\overline{z^{(t+1)}}) \ . \tag{5}$$

Zhao *et al.* [1] describes a strategy for tracking problems to implement Bayesian Filtering in a distributed setting. At each time step one sensor node labeled as the leader makes a measurement and computes the belief $p(x^{(t+1)}|\overline{z^{(t)}})$. Then it select a sensor node to lead the tracking effort and passes the current belief to the chosen leader node. The next sensor to lead the tracking algorithm can be chosen to maximize a utility function of the form:

$$\begin{aligned} U(\overline{z^{(t)}} \cup z_j^{(t+1)}) &= -H[p(x^{(t+1)}|\overline{z^{(t)}} \cup z_j^{(t+1)})] \\ &= \sum_i p(x_i^{(t+1)}|\overline{z^{(t)}} \cup z_j^{(t+1)}) \log p(x_i^{(t+1)}|\overline{z^{(t)}} \cup z_j^{(t+1)}) \ , \end{aligned}$$

where $U(\overline{z^{(t)}} \cup z_j^t)$ is the utility received from decreased uncertainty in the state of the target, which is measured as the entropy of the conditional probability density of $x^{(t+1)}$ given the sensor measurements up to time $(t+1)$. This utility function can be augmented with the communication cost of relaying the current belief from the current leader to the next. For example, the communication cost component of the utility can encompass the bandwidth utilization, transmission and reception power costs.

In this paper we focus on estimable measures of information utility, but a suitable communication cost can easily be integrated with our approach. Typically the measurement for the next sensor is unknown at $(t+1)$. Further, the expectation of the posterior belief $p(x^{(t+1)}|\overline{z^{(t)}} \cup z_j^t)$ is equal to the predicted belief $p(x^{(t+1)}|\overline{z^{(t)}}))$.

$$E[p(x^{(t+1)}|\overline{z^{(t)}} \cup z_j^{(t+1)})|\overline{z^{(t)}}]$$

$$
\begin{aligned}
&= E\left[\frac{f_j(z_j^{(t+1)}|x^{(t+1)})p(x^{(t+1)}|\overline{z^{(t)}})}{g_j(z_j^{(t+1)}|\overline{z^{(t)}})}\Big|\overline{z^{(t)}}\right] \\
&= p(x^{(t+1)}|\overline{z^{(t)}})E\left[\frac{f_j(z_j^{(t+1)}|x^{(t+1)})}{g_j(z_j^{(t+1)}|\overline{z^{(t)}})}\Big|\overline{z^{(t)}}\right] \\
&= p(x^{(t+1)}|\overline{z^{(t)}})\int_{z\in Z_j}\frac{f_j(z|x^{(t+1)})}{g_j(z|\overline{z^{(t)}})}g_j(z|\overline{z^{(t)}})dz \\
&= p(x^{(t+1)}|\overline{z^{(t)}})\int_{z\in Z_j} f_j(z|x^{(t+1)})dz \\
&= p(x^{(t+1)}|\overline{z^{(t)}})\ .
\end{aligned}
$$

Zhao *et al.* [1] compute a proxy $\hat{p}((x^{(t+1)}|\overline{z^{(t)}}\cup z_j^{(t+1)})$ to the expected posterior belief by averaging $f_j(z_j^{(t+1)}|x^{(t+1)})$ over estimated measurement values using predicted belief. Although, this approximation to the expected posterior belief will not be equal to the predicted belief, the above result indicates that any utility measure based on expected posterior belief will be of limited use for sensor selection. Instead we employ an expected posterior uncertainty measure for sensor selection. In particular, we consider the expected posterior entropy, one of the information measures suggested in [3]:

$$
\hat{j} = \arg\max_{j\in V} E\left[-H(p(x^{(t+1)}|\overline{z^{(t)}}\cup z_j^t))|\overline{z^{(t)}}\right]\ .
$$

In other words the sensor which will result in the smallest *expected* posterior uncertainty of the target state will be chosen to be the leader node of the tracking algorithm. In general, a direct computation of the expected posterior entropy is computationally infeasible. It requires computing the posterior belief for each possible measurement value and then averaging the entropy of the computed posterior belief over all possible measurement values. Even if the measurement space is discretized it requires computationally expensive calculations in the high dimensional state space. In the following, we show that maximizing the mutual information between the sensor output and target state is equivalent to minimizing expected posterior uncertainty. This observation yields a computationally feasible sensor selection method based on a maximum mutual information principle. The expected entropy of the posterior density can be evaluated using 3 and 4.

$$
\begin{aligned}
&E\left[-H(p(x^{(t+1)}|\overline{z^{(t)}}\cup z_j^t))|\overline{z^{(t)}}\right] \\
&= \int_{z\in Z_j}\left(\sum_i p(x_i^{(t+1)}|\overline{z^{(t)}}\cup\{z\})\log p(x_i^{(t+1)}|\overline{z^{(t)}}\cup\{z\})\right)q_j(z|\overline{z^{(t)}})dz \\
&= \int_{z\in Z_j}\sum_i f_j(z|x_i^{(t+1)})p(x_i^{(t+1)}|\overline{z^{(t)}})\log\frac{f_j(z|x_i^{(t+1)})p(x_i^{(t+1)}|\overline{z^{(t)}})}{g_j(z|\overline{z^{(t)}})}dz
\end{aligned}
$$

$$= \sum_i \left(\int_{z \in Z_j} f_j(z|x_i^{(t+1)}) \log f_j(z|x_i^{(t+1)}) dz \right) p(x_i^{(t+1)}|\overline{z^{(t)}})$$
$$- \int_{z \in Z_j} g_j(z|\overline{z^{(t)}}) \log g_j(z|\overline{z^{(t)}}) dz$$
$$+ \sum_i p(x_i^{(t+1)}|\overline{z^{(t)}}) \log p(x_i^{(t+1)}|\overline{z^{(t)}})$$
$$= -H(Z_j^{(t+1)}|X^{(t+1)}) + H(Z_j^{(t+1)}) - H(X^{(t+1)})$$
$$= I(Z_j^{(t+1)}; X^{(t+1)}) - H(X^{(t+1)})$$

We note that the second term does not depend on the sensor measurement at $(t+1)$. Hence, in a Bayesian Filtering framework minimizing the expected uncertainty in the posterior belief is equivalent to maximizing the mutual information between the state $X^{(t+1)}$ and measurement vector $Z_j^{(t+1)}$.

3 Applications

3.1 Example 1: Discrete Observations

We consider a simple two state, two sensor problem to illustrate the concepts presented in Section 2. There are two possible states for the target $x^{(t)} \in \{-1, 1\}$. The state transition model is given by:

$$x^{(t+1)} = F(x^{(t)}, v^{(t)}) = x^{(t)} v^{(t)} \ . \qquad (6)$$

where $v^{(t)}$ is a binary random variable which takes values $\{-1, 1\}$ with probability q and $1 - q$ respectively. The observation model for the two sensors is given by.

$$z_1^{(t)} = H_1(x^{(t)}, w^{(t)}) = \mathrm{sgn}(x^{(t)} + w_1^{(t)}) \qquad (7)$$
$$z_2^{(t)} = H_2(x^{(t)}, w^{(t)}) = \mathrm{sgn}(x^{(t)} - w_2^{(t)}) \qquad (8)$$

The state space model suggests the following conditional probability distributions for state $x^{(t+1)}$ and sensor measurement $z_j(t)$.

$$p(x^{(t+1)} = 1|x^{(t)}) = 1 - p(x^{(t+1)} = -1|x^{(t)}) = \begin{cases} 1-q & \text{if } x^{(t)} = 1 \\ q & \text{if } x^{(t)} = -1 \end{cases}$$

$$f_1(z_1^{(t)} = 1|x^{(t)}) = 1 - f_1(z_1^{(t)} = -1|x^{(t)}) = \begin{cases} 1 & \text{if } x^{(t)} = 1 \\ r & \text{if } x^{(t)} = -1 \end{cases}$$
$$f_2(z_2^{(t)} = -1|x^{(t)}) = 1 - f_2(z_2^{(t)} = 1|x^{(t)}) = \begin{cases} r & \text{if } x^{(t)} = 1 \\ 1 & \text{if } x^{(t)} = -1 \end{cases}$$

i.e., sensor 1 makes an erroneous measurement with probability r if the state is -1, and sensor 2 makes an erroneous measurement with probability r if the state is 1. For this simple model we can parametrize the prior and posterior belief using a scalar variable:

$$p^{(t+1)|\overline{(t)}} \stackrel{\text{def}}{=} p(x^{(t+1)} = 1|z^{\overline{(t)}}) \tag{9}$$

$$p^{(t+1)|\overline{(t+1)}} \stackrel{\text{def}}{=} p(x^{(t+1)} = 1|z^{\overline{(t+1)}}) \tag{10}$$

We can verify that the expected posterior belief $E[p^{(t+1)|\overline{(t+1)}}|z^{\overline{(t)}}]$ is equal to the prior belief $p^{(t+1)|\overline{(t)}}$ irrespective of the sensor choice at time $(t+1)$. If sensor 1 is queried at time $(t+1)$,

$$\begin{aligned}
&E[p^{(t+1)|\overline{(t+1)}}|z^{\overline{(t)}}] \\
&= p(x^{(t+1)} = 1|z^{\overline{(t)}} \cup \{z_1^{(t+1)} = 1\})p(z_1^{(t+1)} = 1|z^{\overline{(t)}}) \\
&\quad + p(x^{(t+1)} = 1|z^{\overline{(t)}} \cup \{z_1^{(t+1)} = -1\})p(z_1^{(t+1)} = -1|z^{\overline{(t)}}) \\
&= \frac{p^{(t+1)|\overline{(t)}} \cdot 1}{p^{(t+1)|\overline{(t)}} \cdot 1 + (1 - p^{(t+1)|\overline{(t)}}) \cdot r}(p^{(t+1)|\overline{(t)}} \cdot 1 + (1 - p^{(t+1)|\overline{(t)}}) \cdot r) \\
&\quad + \frac{p^{(t+1)|\overline{(t)}} \cdot 0}{p^{(t+1)|\overline{(t)}} \cdot 0 + (1 - p^{(t+1)|\overline{(t)}}) \cdot (1-r)}(p^{(t+1)|\overline{(t)}} \cdot 0 + (1 - p^{(t+1)|\overline{(t)}}) \cdot (1-r)) \\
&= p^{(t+1)|\overline{(t)}} .
\end{aligned}$$

Similarly if sensor 2 is queried at time $(t+1)$,

$$\begin{aligned}
&E[p^{(t+1)|\overline{(t+1)}}|z^{\overline{(t)}}] \\
&= p(x^{(t+1)} = 1|z^{\overline{(t)}} \cup \{z_2^{(t+1)} = 1\})p(z_2^{(t+1)} = 1|z^{\overline{(t)}}) \\
&\quad + p(x^{(t+1)} = 1|z^{\overline{(t)}} \cup \{z_2^{(t+1)} = -1\})p(z_2^{(t+1)} = -1|z^{\overline{(t)}}) \\
&= \frac{p^{(t+1)|\overline{(t)}} \cdot (1-r)}{p^{(t+1)|\overline{(t)}} \cdot (1-r)1 + (1 - p^{(t+1)|\overline{(t)}}) \cdot 0}(p^{(t+1)|\overline{(t)}} \cdot (1-r) + (1 - p^{(t+1)|\overline{(t)}}) \cdot 0) \\
&\quad + \frac{p^{(t+1)|\overline{(t)}} \cdot r}{p^{(t+1)|\overline{(t)}} \cdot r + (1 - p^{(t+1)|\overline{(t)}}) \cdot 1}(p^{(t+1)|\overline{(t)}} \cdot r + (1 - p^{(t+1)|\overline{(t)}}) \cdot 1) \\
&= p^{(t+1)|\overline{(t)}} .
\end{aligned}$$

The mutual information between state at time $(t+1)$ and sensor j's output is given by:

$$\begin{aligned}
I(Z_1^{(t+1)}; X^{(t+1)}) &= H(Z_1^{(t+1)}) - H(Z_1^{(t+1)}|X^{(t+1)}) \\
&= \mathcal{H}((1 - p^{(t+1)|\overline{(t)}})(1-r)) - (1 - p^{(t+1)|\overline{(t)}})\mathcal{H}((1-r)) \\
I(Z_2^{(t+1)}; X^{(t+1)}) &= H(Z_2^{(t+1)}) - H(Z_2^{(t+1)}|X^{(t+1)}) \\
&= \mathcal{H}(p^{(t+1)|\overline{(t)}}(1-r)) - p^{(t+1)|\overline{(t)}}\mathcal{H}((1-r)) ,
\end{aligned}$$

where the function $\mathcal{H}$ is defined as $\mathcal{H}(x) = -x\log(x) - (1-x)\log(1-x)$. It is easy to verify that $I(Z_1^{(t+1)}; X^{(t+1)}) > I(Z_2^{(t+1)}; X^{(t+1)})$ for $p^{(t+1)|\overline{(t)}} > 0.5$. For this example minimizing the expected entropy of posterior belief is equivalent to choosing the sensor that is ideal for the most likely state.

3.2 Example 2: Linear Gaussian Model

In this section we consider the sensor selection for the Bayesian filtering problem with linear Gaussian models. We assume the following linear state space model:

$$x^{(t+1)} = Fx^{(t)} + v^{(t)} \tag{11}$$

$$z_j^{(t)} = H_j x^{(t)} + w^{(t)} \tag{12}$$

We assume the disturbances $v^{(t)}, w^{(t)}$ are zero mean Gaussian processes with covariances Σ_v and Σ_w respectively. For a linear Gaussian model and Gaussian prior belief $p(x^{(t)}|\overline{z^{(t)}})$, it can be proved that both $p(x^{(t+1)}|\overline{z^{(t)}})$ and $p(x^{(t+1)}|\overline{z^{(t+1)}})$ are also Gaussian [4]. The mean and covariance for $p(x^{(t+1)}|\overline{z^{(t)}})$ and $p(x^{(t+1)}|\overline{z^{(t+1)}})$ can be computed using the mean and covariance of $p(x^{(t)}|\overline{z^{(t)}})$ and the measurement $z^{(t+1)}$ through Kalman filter recursions.

The observation model in 12 suggests a normal conditional distribution for $z_j^{(t+1)}$:

$$f_j(z_j^{(t+1)}|x^{(t+1)}) = \mathcal{N}(z_j^{(t+1)}; H_j x^{(t+1)}, \Sigma_w) \ , \tag{13}$$

where $\mathcal{N}(y; \mu, \Sigma)$ denotes the Gaussian distribution with mean μ and Σ:

$$\mathcal{N}(y; \mu, \Sigma) \stackrel{\text{def}}{=} ((2\pi)^n |\Sigma|)^{-0.5} \exp\left(-(y-\mu)^T \Sigma^{-1} (y-\mu)\right) .$$

Given the predicted belief $p(x^{(t+1)}|\overline{z^{(t)}}) = \mathcal{N}(x^{(t+1)}; \mu^{(t+1)|\overline{(t+1)}}, \Sigma^{(t+1)|\overline{(t+1)}})$ we can derive the distribution for j'th sensors measurement at time $(t+1)$ as:

$$\begin{aligned}
g_j(z_j^{(t+1)}|\overline{z^{(t)}}) &= \int f_j(z_j^{(t+1)}|x^{(t+1)}) p(x^{(t+1)}|\overline{z^{(t)}}) dx^{(t+1)} \\
&= \int \mathcal{N}(z_j^{(t+1)}; H_j x^{(t+1)}, \Sigma_w) \mathcal{N}(x^{(t+1)}; \mu^{(t+1)|\overline{(t+1)}}, \Sigma^{(t+1)|\overline{(t+1)}}) dx^{(t+1)} \\
&= \mathcal{N}(z_j^{(t+1)}; H_j \mu^{(t+1)|\overline{(t+1)}}, \Sigma_w + H_j \Sigma^{(t+1)|\overline{(t+1)}} H_j^T) \ .
\end{aligned} \tag{14}$$

The mutual information between the sensor measurement and target state at time $(t+1)$ can be calculated using 13,14

$$\begin{aligned}
& I(Z_j^{(t+1)}; X^{(t+1)}) \\
&= H(Z_1^{(t+1)}) - H(Z_1^{(t+1)}|X^{(t+1)}) \\
&= H[\mathcal{N}(z_j^{(t+1)}; H_j \mu^{(t+1)|\overline{(t+1)}}, \Sigma_w + H_j \Sigma^{(t+1)|\overline{(t+1)}} H_j^T)] \\
&\quad - \int H[\mathcal{N}(z_j^{(t+1)}; H_j x^{(t+1)}, \Sigma_w)] p(x^{(t+1)}|\overline{z^{(t)}}) dx^{(t+1)} \\
&= c \log |\Sigma_w + H_j \Sigma^{(t+1)|\overline{(t+1)}} H_j^T| - \int c \log |\Sigma_w| p(x^{(t+1)}|\overline{z^{(t)}}) dx^{(t+1)} \\
&= c \log \frac{|\Sigma_w + H_j \Sigma^{(t+1)|\overline{(t+1)}} H_j^T|}{|\Sigma_w|}
\end{aligned}$$

To summarize the sensor selection rule for minimizing the expected posterior entropy is given as:

$$\hat{j} = \arg\max_{j \in V} \frac{|\Sigma_w + H_j \Sigma^{(t+1)|\overline{(t+1)}} H_j^T|}{|\Sigma_w|} \ .$$

Since the posterior density is Gaussian, this sensor selection rule minimizes the covariance determinant for the posterior belief. We should note that since the covariance (or equivalently the entropy) of the updated belief $p(x^{(t+1)}|z^{\overline{(t+1)}})$ does not depend on the measurement value $z_j^{(t+1)}$, sensor selection for the linear Gaussian model is straightforward.

3.3 Example 3: Acoustic Array

In this section we consider the distributed localization of a single target using an acoustic sensor network. We assume a single target is present in a square 1 km × 1km region, which is divided into 50m × 50m cells. We assume the target is stationary:

$$x^{(t+1)} = F(x^{(t)}, v^{(t)}) = x^{(t)}$$

There are five microphones (range sensors) randomly placed in the region. Each sensor makes a time of arrival measurement (TOA) from an acoustic emission of the target. The sensor measurement model is given as:

$$z_j = \frac{\|x - y_j\|}{c} + n_j \ .$$

where x denotes the target location and y_j denotes the location of the j'th sensor. The speed of sound is given by c and the disturbances n_j's are Gaussian random variables with variance σ_j. The error variance of the maximum likelihood TOA detector is inversely proportional to the signal to noise ratio, which in general depends on the distance of the target to the sensor location [5,6]. In part A below, we assume constant noise variance for all the sensors and in part B, we consider the general case where the noise variance increases with increasing distance to the target. We assume each sensor can be interrogated only once. We also assume that the sensor locations and target emission time are known. A self localization method for microphone arrays is given in [7].

Part A:

In this case we assume the noise variance $\sigma_j = \sigma_0 = 50$msec is constant for all five sensors. For this case the mutual information between the state $X^{(t+1)}$ and measurement vector Z_j is given by

$$I(Z_j^{(t+1)}; X^{(t+1)}) = H(Z_j^{(t+1)}) - H((Z_j^{(t+1)}|X^{(t+1)}) \ ,$$

where:

$$H(Z_j^{(t+1)}) = -\int \sum_{x_i} \mathcal{N}(z, \frac{\|x_i - y_j\|}{c}, \sigma_0^2) p(x_i|z^{\overline{(t)}})$$

$$\times \log \left(\sum_{x_i} \mathcal{N}(z, \frac{\|x_i - y_j\|}{c}, \sigma_0^2) p(x_i|\overline{z^{(t)}}) \right) dz$$

$$H(Z_j^{(t+1)}|X) = \frac{1}{2} \log \left(2\pi e \sigma_0^2 \right) \ .$$

We note that for constant noise variance, maximizing mutual information $I(Z_j^{(t+1)}; X^{(t+1)})$ is equivalent to maximizing the entropy of the Gaussian mixture $H(Z_j^{(t+1)})$. The entropy of the Gaussian mixture can be calculated using numerical integration. Alternatively, we can obtain an approximation to $H(Z_j^{(t+1)})$ by fitting a single Gaussian to the mixture distribution.

$$H(Z_j^{(t+1)}) \approx \frac{1}{2} \log \left(2\pi e \sigma_{Z_j}^2 \right) \ ,$$

where

$$\sigma_{Z_j}^2 = \sum_{x_i} p(x_i|\overline{z^{(t)}}) \left((\frac{\|x_i - y_j\|}{c})^2 + \sigma_0^2 \right) - \left(\sum_{x_i} p(x_i|\overline{z^{(t)}}) \frac{\|x_i - y_j\|}{c} \right)^2 \ .$$

In our simulations we observed virtually no difference in sensor selection performance between actual $H(Z_j^{(t+1)})$ and its approximation.

We used 500 monte carlo simulations, for three methods of sensor selection: Random sensor selection, Maximum Mutual Information based sensor selection and Mahalanobis distance based sensor selection discussed in [3]. The results are given in Figure 1. We consider root mean square error as a measure of target localization performance. For this experiment Maximum Mutual Information based sensor selection results in the best localization performance, followed by Mahalanobis distance based method.

Part B:

In this case we assume the noise variance is dependent on the target distance

$$\sigma_j = \sigma(r) = (\frac{r}{r_0})^{\alpha/2} \sigma_0$$

In general the value of alpha depends on temperature and wind conditions and can be anisotropic. For this experiment we used $\alpha = 2$, $r_0 = 0.5$km, and $\sigma_0 =$ 30msec. For the distance dependent noise model, the mutual information between the state $X^{(t+1)}$ and measurement vector Z_j is given by

$$I(Z_j^{(t+1)}; X^{(t+1)}) = H(Z_j^{(t+1)}) - H((Z_j^{(t+1)}|X^{(t+1)}) \ ,$$

where:

$$H(Z_j^{(t+1)}) = -\int \sum_{x_i} \mathcal{N}(z, \frac{\|x_i - y_j\|}{c}, \sigma(\|x_i - y_j\|)^2) p(x_i|\overline{z^{(t)}})$$

$$\times \log \left(\sum_{x_i} \mathcal{N}(z, \frac{\|x_i - y_j\|}{c}, \sigma(\|x_i - y_j\|)^2) p(x_i|\overline{z^{(t)}}) \right) dz$$

$$H(Z_j|X) = \sum_{x_i} \frac{1}{2} \log \left(2\pi e \sigma(\|x_i - y_j\|)^2 \right) p(x_i|\overline{z^{(t)}}) \ .$$

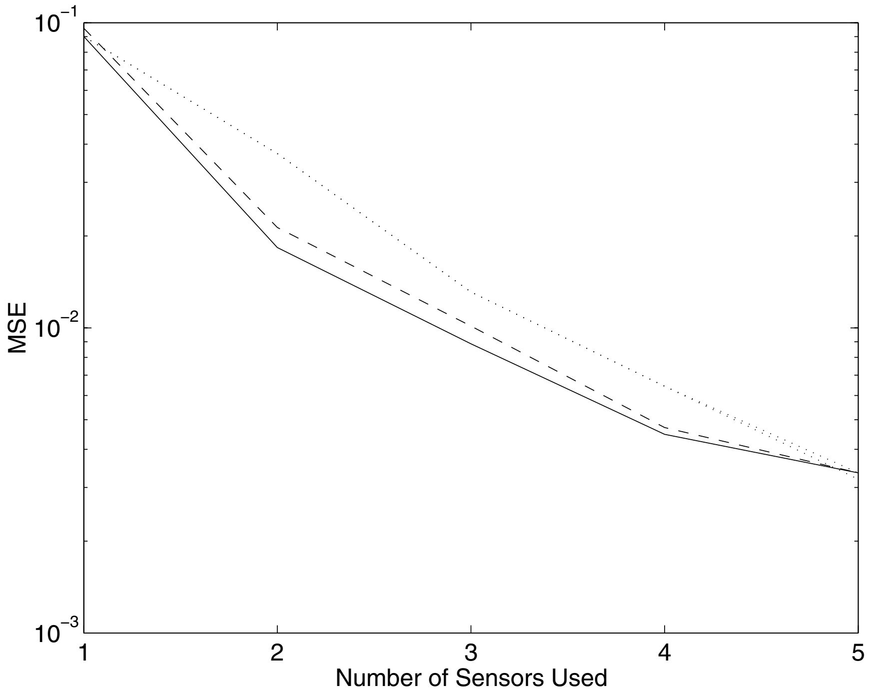

Fig. 1. Performance of the Sensor Selection Methods for constant noise variance. (Solid: Mutual Information, Dashed: Mahalanobis Distance Based Method, Dotted:Random)

Again if the distribution of Z_j can be approximated with a Gaussian we can approximate $H(Z_j^{(t+1)}) \approx \frac{1}{2}\log\left(2\pi e\sigma^2_{Z_j}\right)$, where

$$\sigma^2_{Z_j} = \sum_{x_i} p(x_i|\overline{z^{(t)}})((\frac{\|x_i - y_j\|}{c})^2 + \sigma(\|x_i - y_j\|)^2) - \left(\sum_{x_i} p(x_i|\overline{z^{(t)}})\frac{\|x_i - y_j\|}{c}\right)^2 .$$

We used 500 monte carlo simulations for the range dependent noise case. The results are given in Figure 2. We consider root mean square error as a measure of target localization performance. For this experiment Maximum Mutual Information and Mahalanobis distance based methods are very close in performance. The advantage of dynamic sensor selection over random sensor selection is again evident from the simulation results.

4 Conclusions

Motivated by the work of Zhao *et al.* [1] we have presented an extension to the problem of distributed tracking in sensor networks. Specifically, we considered

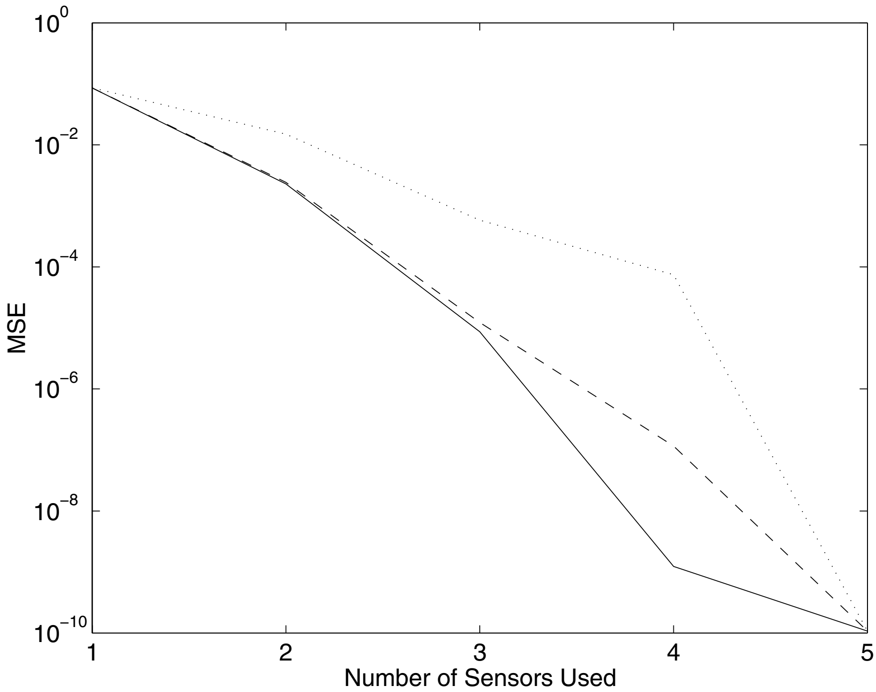

Fig. 2. Performance of the Sensor Selection Methods for range dependent noise variance. (Solid: Mutual Information, Dashed: Mahalanobis Distance Based Method, Dotted:Random)

the expected posterior uncertainty, quantified by conditional entropy as the utility function for choosing the next measurement node in a distributed Bayesian sequential estimation framework. The demonstrated equivalence of expected conditional entropy (over measurements) to the mutual information between future state and the node measurements led to a computationally feasible method for employing the suggested utility function.

Additionally we presented three example problems for which the method could be used along with empirical results. The results indicate that maximum mutual information principle presents a computationally attractive method for dynamic sensor selection problems.

Some interesting questions arise in the context of sensor networks which motivate future research. For example, how does additional attribution of object state (e.g., class) complicate the analysis? How might one incorporate these ideas into heterogeneous networks where measurement models are less well understood? It is unlikely that such modifications will lead to such tractable measurement models; however, it is also the case that estimation of statistical dependence (i.e., mutual information) remains tractable in lower dimensional spaces.

References

1. F. Zhao, J. Shin, and J. Reich, "Information-driven dynamic sensor collaboration," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 61–72, 2002.
2. S. Kumar, D. Shepherd, and F. Zhao, "Collaborative signal and information processing in micro-sensor networks," *IEEE Signal Processing Magazine*, vol. 19, no. 1, 2002.
3. M.Chu, H. Haussecker, and F. Zhao, "Scalable information-driven sensor querying and routing for ad hoc heteregeneneous sensor networks," *Int. J. High-Performance Compu. Applicat.*, 2002. to appear.
4. Y. C. Ho and R. C. K. Lee, "A bayesian approach to problems in stochastic estimation and control," *IEEE Trans. Automat. Contr.*, vol. 9, pp. 333–339, 1964.
5. C. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. on ASSP*, vol. 4, pp. 320–326, 1976.
6. D. Krishnamurty, "Self calibration techniques for acoustic sensor arrays," Master's thesis, The Ohio State University, 2002.
7. R. L. Moses, D. Krishnamurthy, and R. M. Patterson, "An auto-calibration method for unattended ground sensors," in *icassp*, vol. 3, (Orlando, FL), 2002.