Journal of VLSI Signal Processing 2007
© 2007 Springer Science + Business Media, LLC. Manufactured in The United States.
DOI: 10.1007/s11265-007-0079-0

Information Regularized Sensor Fusion: Application to Localization With Distributed Motion Sensors

UMUT OZERTEM AND DENIZ ERDOGMUS
CSEE Department, Oregon Health and Science University, Portland, OR, USA

Received: 2 May 2006; Revised: 20 October 2006; Accepted: 2 April 2007

Abstract. We propose the information regularization principle for fusing information from sets of identical sensors observing a target phenomenon. The principle basically proposes an importance-weighting scheme for each sensor measurement based on the mutual information based pairwise statistical similarity matrix between sensors. The principle is applied to maximum likelihood estimation and particle filter based state estimation. A demonstration of the proposed regularization scheme in centralized data fusion of dense motion detector networks for target tracking is provided. Simulations confirm that the introduction of information regularization significantly improves localization accuracy of both maximum likelihood and particle filter approaches compared to their baseline implementations. Outlier detection and sensor failure detection capabilities, as well as possible extensions of the principle to decentralized sensor fusion with communication constraints are briefly discussed.

Keywords: information regularization principle, mutual information, target localization and tracking, binary motion detector network

1. Introduction

Recent advances in wireless networking and the increasing availability of small, low-power, inexpensive sensors with wireless communication capabilities have driven the advent of many interesting applications of distributed sensing and information fusion. Sensor network research includes various detection, tracking, and localization applications [1–4], ranging from geophysical studies [5–7], to health monitoring systems [8–10], and military applications [11]. The most common environmental studies include fire, and flood detection, and biocomplexity mapping, where mostly pressure, humidity, and temperature sensors are deployed. Since the sensors are typically left unattended for extended durations in environmental applications, they are preferably equipped with effective power recharging capabilities, such as solar cells. Monitoring of human cognitive and physical activity data is the most common health application of sensor networks. In some studies, for example, the aim is to detect some patterns in human behavior are early indicators of neurodegenerative diseases.

In many application domains, distributed binary motion detector networks merit attention due to computationally cheap sensing and processing characteristics. Bayesian decision schemes for binary sensor networks have been studied by Tsitsiklis [2], Chair and Varshney [12], and other researchers. However, if the phenomena that we want to sense have some temporal–spatial dependencies, these original approaches have to be modified to account for sensor location dependent hypotheses. Another similar issue is the dependency to the individual detection characteristics of sensors, such as false alarm and miss probabilities or detection range, which changes the sensor behavior throughout the sensor network field. To address spatial dependencies in target localization, Rodriguez, Tong, and colleagues proposed improvements to maximum likelihood decision fusion by considering source localization [13, 14].

Another significant problem in target tracking is data association for the scenarios that include multiple targets. This problem can be regarded as the way of determining the correspondences between the sensor measurements and multiple targets. The most widely used data association algorithm is multiple hypothesis tracking (MHT) algorithm [15]. Chong and colleagues propose a distributed version of this algorithm [16]. However, the main drawback of this approach is that the computational complexity of the algorithm increases exponentially with the number of sensors, which makes it impractical to use in dense sensor network fields. Recently, a graphical model optimization based approach is proposed; Chen and colleagues overcome this problem, which makes MHT bases approaches applicable in dense sensor networks [17].

There is also a recent interest in incorporating machine learning techniques to extract the characteristics of the environment in order to design adaptive and environment-aware systems, called cognitive devices [18]. The idea behind this approach is to exploit statistical redundancies in the environment behavior to modify the operating mode, and increase performance and efficiency. In the context of target tracking, if the target has a statistically stationary or slowly varying nonstationary behavior, sensor networks can exploit the statistical similarities in their measurements to improve their localization performance. By incorporating information about past sensing activity into the current estimation process, one can improve the performance beyond the standard naïve Bayesian estimation approaches that assume identical independent sensors that are unaware of their environment and their history, which is the current benchmark. Clearly, in the ideal case, the sensor network would have accurate knowledge of the prior distribution or statistical model of the phenomenon being observed, and using this prior knowledge, would expand its estimation scheme in a maximum a posteriori (MAP) context. Such information could also be used to achieve effective measurement dimensionality reduction. However, obtaining such detailed prior information is expected to pose a couple of practical difficulties. The distributed nature of the phenomenon prevents the network from acquiring sufficient data to accurately estimate the prior with small amounts of training data. Moreover, the global nature of the prior prevents decentralization of estimation procedure. This contradicts with the current trend in distributed sensor network research, which is to achieve decentralized data/decision fusion with nearly optimal performance and minimal bandwidth and energy requirements [19]. In localization using only instantaneous sensor measurements, optimal typically refers to maximum likelihood, and can only be achieved through centralized means, especially because all sensor measurements are treated as equally important. On the other hand, if prior knowledge of the target characteristics can be summarized utilizing past sensor activity, the sensors can be ranked/weighted individually according to the information that they provide related to the target phenomenon. Alternatively, pairwise relevance of sensors can be incorporated to minimize communication requirements between the sensors that provide redundant or irrelevant information. Both of these approaches will lead to solutions that reduce the communication requirements of the sensor networks through selective sensor information transmissions. Consequently, the proposed information regularization principle lends itself to decentralized decision making conveniently. Moreover, the pairwise statistical similarity between sensor measurements can be used to estimate the information diversity in the network, thus achieve self-organization of the network through parameter modifications (such as sensor location) to maximize diversity.

In this paper, we demonstrate an application of the information regularization principle to target localization and tracking using a dense motion sensor network. Specifically, we implement modifications to two benchmark Bayesian approaches, namely maximum likelihood (ML) and particle filter (PF) estimation, by incorporating importance-weighting to sensor measurements using mutual information based statistical similarity of sensor activation history. Both the information regularized maximum likelihood (IRML) and the information regularized particle filter (IRPF) are shown to improve localization performance significantly relative to their original counterparts at minimal additional computational cost of evaluating the information regularization weights (which essentially corresponds to a training phase). The application presented here is also illustrative of the fact that the information regularization principle can be utilized for various other sensor network applications such as faulty sensor identification.

2. Problem Statement

We consider the problem of target localization using binary motion sensor measurements and the *firing* (indicating detection) characteristics of the sensors. The proposed principle is employed in both ML estimation of target localization based on the instantaneous sensor firing patterns, which does not exploit the temporal dynamics of target trajectories, and in PF estimation that utilizes complete statistical knowledge about target dynamics [20, 21]. In both ML and IRML, optimal instantaneous estimates are determined by initializing the iterative maximization algorithm to the estimate at the previous time step, which incorporates some rudimentary temporal information.

Identical motion sensors that are sensitive to the target in a radially symmetric fashion are assumed (for simplicity) with a monotonically decreasing probability of detection for increasing sensor to target distance. Given the sensor position s, and the target position x, the firing probability of the sensor should satisfy

$$\begin{aligned}&1)\, p(f=1|x,s)=p(f=1|\|x-s\|)\\&2)\, \alpha\le p(f=1|\|x-s\|)\le 1-\beta\\&3)\, \|x-s\|\le\|x'-s'\|\\&\quad\Leftrightarrow p(f=1|\|x'-s'\|)\le p(f=1|\|x-s\|)\\&4)\, \lim_{\|x-s\|\to 0} p(f=1|\|x-s\|)=1-\beta\\&5)\, \lim_{\|x-s\|\to\infty} p(f=1|\|x-s\|)=1-\alpha\end{aligned} \tag{1}$$

where f is the sensor output, α is the probability of false alarm, and β is the probability of misdetection. Although it is not necessarily the best model for real sensors, a single-sided Gaussian profile is the most typical model choice in the literature for illustrative purposes. Here, we also employ the one sided Gaussian model, which is

$$p(f=1|\|x-s\|)=\alpha+(1-\alpha-\beta)e^{-\|x-s\|^2/h^2} \tag{2}$$

where h defines the half detection probability range. Hence, $p(f=1|h)=1/2$. Consequently, $p(f=0|\|x-s\|)=1-p(f=1|\|x-s\|)$.

Given the sensor locations $\{s_1,...,s_n\}$and the sensor outputs $\{f_1,...,f_n\}$ at any time instant, the ML estimate for the target location is the solution of the following optimization problem.

$$\begin{aligned}\hat{x}^{\mathrm{ML}}=&\arg\max_x \textstyle\sum_{f_i=1}\log p(f_i=1|\|x-s_i\|)\\&+\textstyle\sum_{f_i=0}\log\left(1-p(f_i=1|\|x-s_i\|)\right)\end{aligned} \tag{3}$$

At this point, note that the ML solution does not consider any prior information about the probability distribution of the target location. In a static estimation scenario, the MAP estimate, on the other hand, would utilize such prior distribution information $p(x)$ to augment Eq. (3) into [22]:

$$\begin{aligned}\hat{x}^{\mathrm{MAP}}=&\arg\max_x \textstyle\sum_{f_i=1}\log p(f_i=1|\|x-s_i\|)\\&+\textstyle\sum_{f_i=0}\log\left(1-p(f_i=1|\|x-s_i\|)\right)\\&+\log\left(p(x)/p(f_1,...,f_n)\right)\end{aligned} \tag{4}$$

In the case of dynamic target tracking, recursive Bayesian estimation utilizing knowledge about the target dynamics would provide a more appropriate and convenient implementation of the MAP principle. For instance, assuming a simple circular Gaussian random walk model for the target position and, we would have the state transition probability distribution as $p(x_t|x_{t-1})=G\left(x_t;x_{t-1}+\mu_v,\sigma_v^2 I\right)$.[1] Also assuming independent sensor measurements, the state-conditional measurement distribution $p(f_t|x_t)=\prod_{s=1}^{n} q_{ts}^{f_{ts}}(1-q_{ts})^{1-f_{ts}}$, where $q_{ts}=p(f_s=1|\|x_t-s_s)$. Given this model, one can then employ standard Kalman filter extensions including the sigma-point or particle filters to perform the following distribution of position given past measurements [20]:

$$p(x_t|f_{0:t})\propto p(f_t|x_t)\int p(x_t|x_{t-1})p(x_{t-1}|f_{0:t-1})dx_{t-1} \tag{5}$$

Note that $p(f_t|x_t)$ is the likelihood of the instantaneous sensor measurements given the target position, which yields the log-likelihood in Eq. (3). The integral term in Eq. (5) is the prior information equivalent to that in Eq. (4). In particular, for the basic particle filter, one updates a number of particles and their weights as follows:

$$x_t^p = x_{t-1}^p + v_{t-1} \quad u_t^p \propto u_{t-1}^p p\left(f_t \middle| x_t^p\right) \tag{6}$$

where v_{t-1} is drawn from $G\left(.; \mu_v, \sigma_v^2 I\right)$. The estimate can then be obtained using a weighted sample average over the particles: $\hat{x}_t^{PF} = \sum_p u_t^p x_t^p$ (although one could easily argue against the use of this method for cases where the state distribution has multiple modes).

3. Information Regularization Using Pairwise Sensor Activity Mutual Information Graph

Consider a distributed motion sensor network that is utilized for target detection/localization over a predefined bounded region. Suppose that the target trajectories follow a stationary statistical generative model (e.g., Markov process). Under these assumptions one would expect a given target to follow some random trajectory T, and the sensors to provide outputs in a consistent fashion with the random trajectory T. If the trajectory comes from a stationary probability distribution, the sensor network outputs are expected to have statistical similarities for each target trajectory. Here, note that the sensor outputs are discrete-values point-processes and the term correlation is used in its broader meaning including nonlinear and discrete-valued dependencies between sensor outputs.

As the target follows the random trajectory T, suppose $f_i(T)$ be the corresponding random binary firing sequence of sensor i. Here, we investigate the mutual information between the pairs of sensor outputs, denoted by $I(f_i(T), f_j(T))$, and we expect the mutual information to be high for the sensor pairs that fire and remain silent simultaneously. On the other hand, for the pairs that detect the target asynchronously, the mutual information is expected to be small. Similarly, the sensor pairs that contain one or two sensors that never detect the target rather than occasional false alarm firings will show almost zero mutual information. According to the sensor model given in Eq. (1), which defines a monotonically decreasing function of sensor-to-target distance, the pairwise mutual information graph (where each sensor is a node and the edge strength connecting two nodes is given as I_{ij}) is expected to connect the neighboring sensors at the regions with $p(x_t)/p(f_1,\ldots,f_n)$, while isolating sensors where this ratio is small.

The mutual information between the sensor output sequences $f_i(T)$ and $f_j(T)$ is defined as the expected value of the ratio of the joint and marginal distributions of these sequences: $p(f_i(T), f_j(T))$, $p(f_i(T))$, and $p(f_j(T))$.

$$I\left(f_i(T), f_j(T)\right) = E\left[\log \frac{p\left(f_i(T), f_j(T)\right)}{p(f_i(T))p\left(f_j(T)\right)}\right] \tag{7}$$

In practice, the estimation of this mutual information requires the estimation of the joint probability densities, which increases exponentially with the length of the sequences. Assuming (reasonably) that each sensor exhibits independent firing behavior at each time step T_k of the trajectory, the joint distribution of each sensor's firing sequence over T becomes separable into the product of firing probabilities at each time step: $p(f_i(T)) = p(f_i(T_1))\ldots p(f_i(T_N))$. Similarly, assuming each pair of sensors are independent from their pairwise behavior at other time steps, we obtain the separability of the joint distribution of the time sequences into the product of the individual firing behaviors at each time instant: $p\left(f_i(T), f_j(T)\right) = p\left(f_i(T_1), f_j(T_1)\right)\ldots p\left(f_i(T_N), f_j(T_N)\right)$. This simplifies the statistical difficulty posed by the high dimensionality of Eq. (5), and reduces the expression to the simple bivariate mutual information between two binary random variables.

$$\begin{aligned} I\left(f_i(T), f_j(T)\right) &\approx E\left[\log \prod_{k=1}^{N} \frac{p\left(f_i(T_k), f_j(T_k)\right)}{p(f_i(T_k))p\left(f_j(T_k)\right)}\right] \\ &= \sum_{k=1}^{N} E\left[\log \frac{p\left(f_i(T_k), f_j(T_k)\right)}{p(f_i(T_k))p\left(f_j(T_k)\right)}\right] \\ &= \sum_{k=1}^{N} I\left(f_i(T_k), f_j(T_k)\right) \end{aligned} \tag{8}$$

At this point, one can utilize the sensor output sequences while a set of *training* targets to estimate the pairwise mutual information between pairs of sensors i and j. In a more realistic scenario, one would initialize the mutual information between sensors uniformly for all pairs (thus, no regulariza-

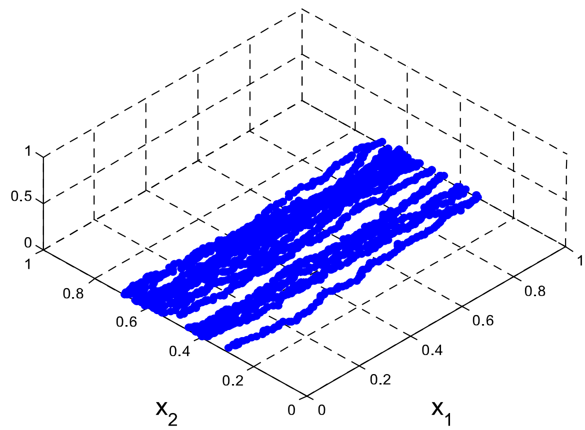

Figure 1. Sample target trajectories using the random walk model given in Eq. (14) for 20 sample Monte Carlo (MC) simulations.

tion is imposed and IRML starts as ML, IRPF starts as PF), and recursively updating these quantities according to the sensor sequences for the current target. Since the aim of this paper is to present the information regularization concept, here we assume that we are given a set of training targets to be utilized for the estimation of the mutual information for each sensor pair. Recursive updates for the information regularization with a *suitable* forgetting factor, on the other hand, can also be used for extending the modeling capabilities of information regularization to target processes from a nonstationary distribution. These will be investigated in a future publication. Given concatenated firing sequences $f_i(T)$ and $f_j(T)$ for many target trajectories (represented by T), the joint and marginal firing probabilities $p(f_i,f_j)$, $p(f_i)$, and $p(f_j)$, where $f_i,f_j \in \{0,1\}$, are estimated by relative occurrence frequencies. That is, if the concatenated firing sequences $f_i(T)$ and $f_j(T)$ have N sampling time instants, the frequency of four possible co-firing patterns, $\{(0,0),(0,1),(1,0),(1,1)\}$ for sensor i and j are counted and normalized by N to estimate the joint and marginal probability distributions in Eq. (8).

A sensor that is at a central location for a given target dynamics is expected to have large mutual information with its neighboring sensors. The matrix constructed using these pairwise mutual information values $I_{ij} = I\big(f_i(T), f_j(T)\big)$ can be regarded as a similarity matrix (generalized version of the adjacency matrix in graph theory; used in spectral machine learning methods). This adjacency matrix represents a graph that summarizes the statistical similarity of the sensor activity in the network, which in the case of target tracking is also related to the spatial organization of the sensors. We define the information regularization weights as the degree of each node (sum of edge strengths that are connected to a particular node/sensor) normalized by the volume of the graph (sum of all edges in the graph):

$$w_i = \frac{\sum_j I\big(f_i(T), f_j(T)\big)}{\sum_j \sum_l I\big(f_l(T), f_j(T)\big)} \tag{9}$$

An illustration of the IR weights in Eq. (9) for a random walk target behavior is provided in Figs. 1 and 2. Note that this selection corresponds to assuming a fully connected graph where each sensor is a node and the pairwise mutual information values are edge weights. Alternatively, sparsely connected geographical neighborhood graphs could be incorporated to impose communication only between spatially nearby nodes. Thresholding the pairwise mutual information values and maintaining only the strong connections would also yield another sparse connectivity graph based on statistical similarity and could encourage both geographical selectivity and improved outlier rejection. Imposing zero entries in the similarity matrix using one of these methods corresponds to enforcing no-similarity between certain sensors, thus helps reduce centralization, communication requirements, and overall computation load.

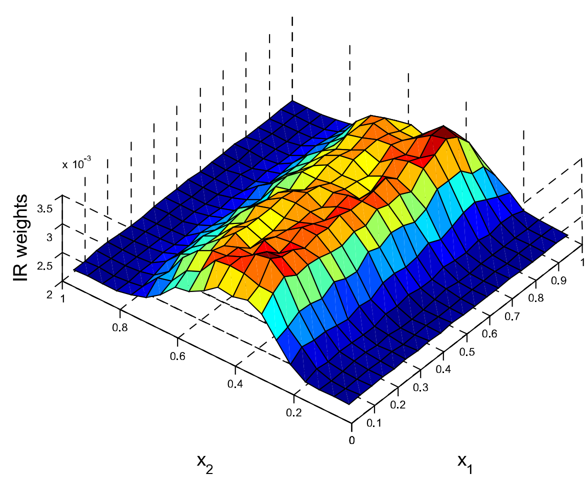

Figure 2. Information regularization weights over a 20×20 uniform sensor grid using 100 MC realizations of Eq. (14) as training data.

Using the information regularization weights proposed in Eq. (8), we define the information regularized maximum likelihood (IRML) criterion for target localization. IRML criterion is obtained by modifying ML criterion in Eq. (3) using the information regularization weights assigned to each sensor in Eq. (9). The regularized likelihood function becomes a weighted geometric mean:

$$p(f_t|x_t;w) = \prod_{s=1}^{n} q_{ts}^{w_s f_{ts}} (1 - q_{ts})^{w_s(1-f_{ts})} \qquad (10)$$

At this point it is important to emphasize that IRML uses only the independent sensor firing assumption that is also utilized by the ML approach. Moreover, the independent sensor firing assumption here does not imply anything about the mutual information between the sensor output sequences given in Eqs. (5) and (6). The latter simply defines a statistical similarity measure between these time series, and does not imply statistical dependency of instantaneous sensor firings.

Considering the logarithm of Eq. (10), the objective criterion of IRML takes a weighted log-likelihood form:

$$\begin{aligned}\hat{x}^{\mathrm{IRML}} = \arg\max_x \textstyle\sum_{f_i=1} w_i \log p(f_i = 1 | \|x - s_i\|) \\ + \textstyle\sum_{f_i=0} w_i \log(1 - p(f_i = 1 | \|x - s_i\|))\end{aligned} \qquad (11)$$

The IRML estimate can be obtained by maximizing the criterion in Eq. (11) using a fixed-point iterative algorithm similar to the mean-shift clustering algorithm [23]. For the firing model of Eq. (3) with $\alpha=\beta$, taking the gradient of Eq. (11) with respect to x, equating to zero, and rearranging terms yields the following simple iterative optimization rule (detailed derivation in the Appendix):

$$\hat{x}_t = \frac{\left[\sum_{f_i=1} w_i \frac{(q_i(\hat{x}_t)-\alpha)s_i}{q_i(\hat{x}_t)} + \sum_{f_i=0} w_i \frac{(q_i(\hat{x}_t)-\alpha)s_i}{(1-q_i(\hat{x}_t))}\right]}{\left[\sum_{f_i=1} w_i \frac{(q_i(\hat{x}_t)-\alpha)}{q_i(\hat{x}_t)} + \sum_{f_i=0} w_i \frac{(q_i(\hat{x}_t)-\alpha)}{(1-q_i(\hat{x}_t))}\right]} \qquad (12)$$

where $q_i(x) = p(f_i = 1 | \|x - s_i\|) = \alpha + (1 - 2\alpha) e^{-\|x-s_i\|^2/h^2}$.

In order to reduce iterations-to-convergence, avoid irrelevant local optima, and incorporate temporal information regarding the dynamics of the target trajectory, the iterations in Eq. (12) are initialized to the IRML-estimate of the target location obtained at the previous time step using Eq. (11).

The application of the proposed regularization to PF is straightforward. Simply, the regularized likelihood function in Eq. (10) is used in the weight updates of the particles:

$$x_t^p = x_{t-1}^p + v_{t-1} \quad u_t^p \propto u_{t-1}^p p(f_t | x_t^p; w) \qquad (13)$$

4. Applications of the IR Principle

Before proceeding to the experimental results regarding the IRML and IRPF algorithms, we will briefly discuss how information regularization can be utilized to address common problems in sensor fusion. The redundancy between sensor outputs (in the sense of insignificant performance gain relative to added communication and computational resource requirements) can be eliminated in a variety of ways depending on the application. To reduce the communication requirements, one can use the pairwise mutual information graph as a similarity matrix of sensor pairs, which facilitates the clustering of *similar* sensors together for communication purposes, yielding a hierarchical decision fusion scheme, or enabling selective sensor communication based on past information relevance. Sensor clustering can be utilized to define subnetworks to minimize unnecessary network-wide communications. Furthermore, one can replace the group of similar sensors with fewer sensors depending on the application and the coverage area of this group of sensors. Similarity based clustering—for instance spectral clustering—techniques (see [23] for some pointers to relevant literature) based on the mutual information similarity measure in Eq. (7) would be essential for these modifications.

Another prospect would be to reorganize the sensor locations to avoid similarly behaving sensors, which will efficiently maximize the information diversity of the overall sensor network, and increase the performance under the given statistics of the target phenomenon. Sensor network reorganization could also be achieved in a self-organizing fashion by employing sensors that have the ability to modify their location (and orientation if directional sensors are used).

Sensor failure detection is another promising application for the information regularization principle. Temporal evolution of the pairwise mutual information matrix can be exploited to identify sensors that exhibit unexpected sudden changes in their co-sensing behavior with neighboring sensors, presumably due to failure.

5. Experimental Results

In this section we present experimental results for a sensor network with firing probabilities as given in Eq. (2). We compare results obtained with proposed IRML and IRPF with benchmarks ML and PF. For simplicity, we use sensors uniformly deployed on a grid in all of the experiments. Note that, this is not necessarily an optimal choice. As mentioned in the previous section, sensor locations can in fact be selected according to the pairwise mutual information matrix to increase the information diversity among the sensors for a particular target trajectory model.

Mean localization errors (average Euclidean distance between true and estimated position of target) averaged over the trajectories generated in 100 Monte Carlo runs are provided. In these experiments, the target trajectories were random walks generated according to

$$x_t = x_{t-1} + \mu + v_{t-1} \tag{14}$$

where $\mu=[0.01,0]^T$ and v is a circular bivariate Gaussian random variable with covariance $(0.005^2 I)$. The probability of miss and false alarms are assumed to be equal: $\alpha=\beta$, and a 1×1 unit-distance-square area of interest covered by a uniform 10×10 sensor grid is considered.

Figure 3 shows the mean localization error of all four methods as a function of false alarm rate α (maintaining $\alpha=\beta$). Figure 4 shows the mean localization error of all four methods as a function of equivalent density of sensors (simulated by manipulating the half-detection range h of the sensors). Proposed IRML and IRPF estimators significantly outperform their benchmark counterparts ML and PF estimators. This is due to the fact that the proposed regularization scheme achieves an effective dimensionality reduction in the measurement vector by projecting the log likelihood function on a more informative subspace determined by the IR weight vector, which is illustrated for a 20×20 uniform sensor grid in Fig. 2 that demonstrates how sensors in greater synchrony with others receive higher importance weights. In addition, as expected, the PF approaches outperform the ML approaches due to the additional advantage provided by the target dynamic model information.

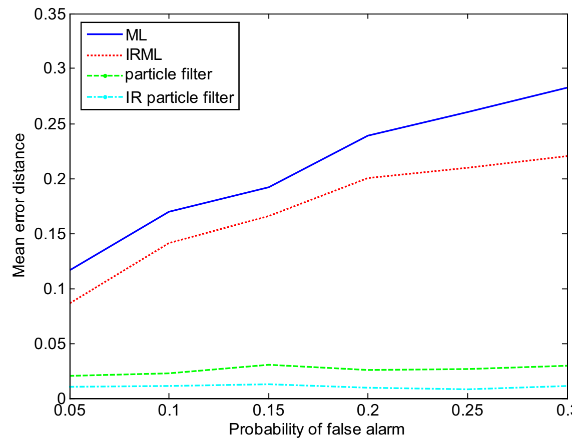

Figure 3. Mean of the distance between the actual target position and the estimated location in 100 Monte Carlo simulations for a 10×10 uniform sensor grid as a function of false alarm (detection) probability: *ML* (*solid*), *IRML* (*dashed*).

6. Discussions

Effective and efficient utilization cheap, low-power, and reliable sensing devices in dense sensor networks for various problems such as detection, localization, and tracking provides an accuracy-complexity tradeoff. In ideal situations with complete or very accurate

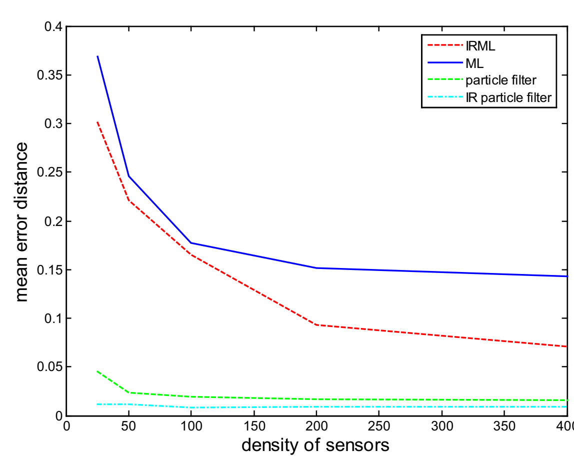

Figure 4. Mean error distance for sensor networks of different spatial density: *ML* (*solid*), *IRML* (*dashed*).

prior knowledge regarding the measured phenomena, sensor fusion using Bayesian techniques usually provides accuracy-optimal but when such prior information is not available maximum likelihood provides a simple yet relatively inaccurate solution. In this paper, we presented the application of an information regularization principle to maximum likelihood and particle filter target localization using motion sensor networks. Improved accuracy was obtained compared to both benchmarks with insignificant additional computational training burden.

The information regularization principle is based on the assumption that the observed distributed phenomenon exhibits spatial or temporal correlations that can be captured by sensor activity correlations, as well as being (quasi-) stationary. The two assumptions lead to the expectation that the sensor network will comprise of sensors that behave in a *statistically similar* manner. This similarity can be best measured nonparametrically by the pairwise mutual information between the activities of two sensors. Due to the correlations exhibited by the target phenomenon, the sensor activities exhibit strong correlations between sensors measuring *statistically nearby* aspects of the observed variable. Consequently, such sensor pairs develop a strong mutual information connection indicating measurement redundancy.

In this paper, we have focused on the application of information regularization on incorporating prior sensing experience into the Bayesian sensor fusion framework through a modified maximum likelihood target localization criterion that emphasizes sensors that exhibit higher statistical *correlation* in their firing activity. This is motivated by the assumption that these sensors are sensing relevant and consistent information. Information regularization reduces the effect of outlier sensors that fire falsely without the presence of a target in their neighborhoods, thus improve localization accuracy compared to standard maximum likelihood and particle filter estimation significantly.

Contemporary challenges in sensor network research include decentralization and bandwidth constraints. The IR principle can be easily adapted to handle these constraints by assuming a suitable sparse connectivity (adjacency) graph between the sensors, thus enabling sensors to eliminate their communication with and dependency on other sensors that do not provide synchronous relevant information or that are geographically separated.

Acknowledgements

This work is partially supported by the NSF grants ECS-0524835 and ECS-0622239.

1. Appendix

The fixed point iteration in Eq. (12) is derived in this appendix. The gradient of the IRML criterion given in Eq. (11) with respect to the target position x is:

$$\nabla_x^T J^{IRML} = \sum_{f_i=1} w_i \frac{\nabla_x^T q_i(x)}{q_i(x)} + \sum_{f_i=0} w_i \frac{\nabla_x^T (1 - q_i x)}{(1 - q_i x)} \tag{15}$$

where $q_i(x) = p(f_i = 1 | \|x - s_i\|) = \alpha + (1 - 2\alpha) e^{-\|x-s_i\|^2/h^2}$. Note that $\nabla_x^T q_i(x) = (q_i(x) - \alpha)\left[-2(x - s_i)/h^2\right]$. Substituting this in Eq. (15), we obtain:

$$\nabla_x^T J^{IRML} = -2h^{-2} \left[\begin{array}{l} \sum\limits_{f_i=1} w_i \dfrac{(q_i(x) - \alpha)(x - s_i)}{q_i(x)} \\ -\sum\limits_{f_i=0} w_i \dfrac{(q_i(x) - \alpha)(x - s_i)}{(1 - q_i(x))} \end{array} \right] \tag{16}$$

Equating Eq. (16) to zero and rearranging terms to leave x alone on one side of the equation yields a fixed-point update:

$$x = \frac{\left[\sum\limits_{f_i=1} w_i \frac{(q_i(x)-\alpha)s_i}{q_i(x)} + \sum\limits_{f_i=0} w_i \frac{(q_i(x)-\alpha)s_i}{(1-q_i(x))} \right]}{\left[\sum\limits_{f_i=1} w_i \frac{(q_i(x)-\alpha)}{q_i(x)} + \sum\limits_{f_i=0} w_i \frac{(q_i(x)-\alpha)}{(1-q_i(x))} \right]} \tag{17}$$

Note

1. Note that in general target tracking problems, the target motion dynamics are unknown so they must be estimated jointly with the state vector, which would include not only position, but also velocity, and acceleration history.

References

1. P. K. Varshney, *Distributed Detection and Data Fusion*, Springer, New York, 1997.
2. J. N. Tsitsiklis, "Decentralized Detection by a Large Number of Sensors," *Math. Control Signals Syst.*, vol. 1, 1988, pp. 167–182.

3. J. C. Chen, K. Yao, and R. E. Hudson, “Source Location and Beamforming,” *IEEE Signal Process. Mag.*, vol. 19, 2002, pp. 30–39.
4. R. R. Brooks, P. Ramanathan, and A. M. Sayed, “Distributed Target Classification and Tracking in Sensor Networks,” *Proc. IEEE*, vol. 91, no. 8, 2003, pp. 1163–1171.
5. A. Baptista, T. Leen, Y. Zhang, A. Chawla, D. Maier, W. C. Feng, W. C. Feng, J. Walpole, C. Silva, and J. Freire, “Environmental Observation and Forecasting Systems: Vision, Challenges and Successes of a Prototype,” in *Encyclopedia of Physical Science and Technology*, R. A. Meyers (Ed.), Academic, 3rd edn., vol. 5, pp. 565–581.
6. P. Bonnet, J. Gehrke, and P. Seshadri, “Querying the Physical World,” *IEEE Pers. Commun.*, vol. 7, 2000, pp. 10–15.
7. A. Cerpa, J. Elson, M. Hamilton, and J. Zhao, “Habitat Monitoring: Application Driver for Wireless Communications Technology,” *ACM SIGCOMM'2000*, Costa Rica, 2001.
8. H. Jimison, M. Pavel, J. McKanna, and J. Pavel, “Unobtrusive Monitoring of Computer Interactions to Detect Cognitive Status in Elders,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 8, 2004, pp. 248–252.
9. B. G. Celler, T. Hesketh, W. Earnshaw, and E. Ilsar, “An Instrumentation System for the Remote Monitoring of Changes in Functional Health Status of the Elderly, International Conference,” *IEEE-EMBS*, New York, 1994, pp. 908–909.
10. G. Coyle, L. Boydell, and L. Brown, “Home Telecare for the Elderly,” *J. Telemed. Telecare*, vol. 1, 1995, pp. 183–184.
11. F. Ye, H. Luo, J. Cheng, S. Lu, and L. Zhang, “A Two-Tier Data Dissemination Model for Large-Scale Wireless Sensor Networks,” *Proceedings of ACM-ICMCM*, 2002, pp. 148–159.
12. Z. Chair, and P. K. Varshney, “Optimal Data Fusion in Multiple Sensor Detection Systems,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 22, no. 1, 1986, pp. 98–101.
13. A. Artes-Rodriguez, R. Torres-Lopez, and L. Tong, “Distributed Detection by Multiple Tests in Sensor Networks Using Range Information,” *Proceedings of EUSIPCO*, 2004.
14. A. Artes-Rodriguez, “Decentralized Detection in Sensor Networks Using Range Information,” *Proceedings of ICASSP*, vol. 2, 2004, pp. 265–268.
15. D. B. Reid, “An Algorithm for Tracking Multiple Targets,” *IEEE Trans. Automat. Contr.*, vol. 24, 1979, pp. 843–854.
16. C. Y. Chong, S. Mori, and K. C. Chang, “Distributed Multitarget Multisensor Tracking,” *Advanced Applications*, vol. 1, 1990, pp. 247–295.
17. L. Chen, M. J. Wainwright, M. Cetin, and A. S. Willsky, “Data Association Based on Optimization in Graphical Models with Application to Sensor Networks,” *Math. Comput. Model.*, vol. 43, 2006, pp. 1114–1135.
18. S. Haykin, “Cognitive Radar,” *IEEE Signal Process. Mag.*, vol. 23, 2006, pp. 30–40.
19. L. Doherty, B. A. Warneke, B. Boser, and K. S. J. Pister, “Energy and Performance Considerations for Smart Dust,” *International Journal of Parallel and Distributed Sensor Networks*, vol. 4, no. 3, 2001.
20. A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*, Springer, New York, 2001.
21. E. A. Wan, and R. van der Merwe, “Kalman Filtering and Neural Networks,” in *The Unscented Kalman Filter*, Simon Haykin, (Ed.), Wiley, 2001 (chapter 7).
22. R. Duda and P. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.
23. U. Ozertem, and D. Erdogmus, “Spectral Clustering with Mean Shift Vector Quantization,” *Proceedings of MLSP*, 2005.

Umut Ozertem received his B.S. in Electrical & Electronics Engineering in 2003 from the Middle East Technical University, Turkey. After working at Tubitak-BILTEN between August 2003 and July 2004 under the supervision of Aydin Alatan, he joined the Adaptive Systems Lab in the CSEE Department of the Oregon Health & Science University as a Ph.D. student. He is currently working with Deniz Erdogmus, and his research interests include statistical signal processing and machine learning with a focus on information theoretic methods.

Deniz Erdogmus received his B.S. in Electrical & Electronics Engineering (EEE), and B.S. in Mathematics both in 1997, and the M.S. in EEE in 1999 from the Middle East Technical University, Turkey. He received his Ph.D. in Electrical & Computer Engineering from the University of Florida (UF) in 2002. He worked as a research engineer at TUBITAK-SAGE, Turkey from 1997 to 1999 and was a research assistant and a postdoctoral research associate at UF from 1999 to 2004. Currently, he is an assistant professor at the CSEE and BME Departments of the Oregon Health and Science University. His research focuses on information theoretic and nonparametric techniques in statistical signal processing and its applications to biomedical problems. He has served in editorial and scientific boards of a number of journals and conferences. He was the recipient of the IEEE-SPS 2003 Best Young Author Paper Award and 2004 INNS Young Investigator Award.