INFORMATION REGULARIZED MAXIMUM LIKELIHOOD FOR BINARY MOTION SENSORS

Umut Ozertem, Deniz Erdogmus
CSEE Department, Oregon Health and Science University, Portland, Oregon, USA

Abstract. We propose a pairwise mutual information based regularization technique for maximum likelihood sensor fusion in dense distributed sensor networks. The principle is demonstrated in target localization and tracking using a dense binary motion sensor network under a centralized data fusion framework. Simulations demonstrate that the information regularization enables the maximum likelihood localization procedure to provide significantly more accurate target position estimates compared to its unregularized counterpart, which is the current benchmark. The extensions of the information regularization principle to various sensor and data fusion problems such as outlier detection, and sensor failure identification are discussed.
Keywords: sensor networks, decision fusion, information theoretic learning, maximum likelihood

1. INTRODUCTION

Recent advances in cheap, low-power, and reliable integrated sensing and processing devices with wireless communication capabilities have stimulated research on distributed sensing for various detection, tracking, and localization applications [1-4] ranging from geophysical and environmental studies [5] to remote health monitoring systems [6]. Distributed dense networks of simple (binary) sensing devices (motion and contact sensors) have received special attention in application domains, such as monitoring of elderly in their homes for early diagnosis of various neurological disorders, that require cheap distributed sensing and processing power. Optimal Bayesian decision rules for binary distributed detection networks have been studied by Chair & Varshney [7], Tsitsiklis [2], and other researchers. Varshney [1] discusses these various fusion criteria in detail. When sensing distributed phenomena with spatially varying characteristics, these original approaches must be modified to account for the need for sensor-location dependent hypotheses. Similarly, the individual detection characteristics of sensors (such as false alarm and miss probabilities for motion sensors) might vary due to various reasons including manufacturing imperfections and setup variations. Along these required improvement directions, recently Rodriguez, Tong, and colleagues have published a series of papers where they address maximum likelihood decision fusion taking source localization with distributed binary motion networks as the testbed [8,9].

On a parallel course, there is interest in building *cognitive devices* that utilize machine learning techniques to learn the characteristics of the environment in which they operate in order to exploit statistical commonalities to modify their operating mode for the purpose of improving their performances [10]. Similarly, sensor networks that observe statistically stationary or slowly varying nonstationary phenomena can exploit various statistical machine learning concepts to improve their performance above what one would obtain through standard maximum likelihood (ML) type Bayesian estimation approaches. Clearly, in the ideal case, the sensor network would have accurate knowledge of the prior distribution of the phenomenon being observed and augment its decision/estimation using this prior in a maximum a posteriori (MAP) context. Obtaining such detailed prior information is expected to run into a couple of practical difficulties including: (i) the distributed nature of the phenomenon prevents the network from acquiring sufficient data to accurately learn the prior with small amounts of training data, (ii) the global nature of the prior prevents decentralization of estimation rules. Although the main goal of distributed sensing is to achieve decentralized processing and decision fusion, in order to avoid energy-expensive wireless communication between sensing units [11], the optimal detection and localization performance can normally be achieved only through centralized optimization of Bayesian criteria. In the case of uninformed sensors where the history of target and sensor activity is not remembered, such Bayesian approaches require equal consideration of every sensor. If some prior knowledge about target behavior can be encapsulated through past sensor activity, the sensors can be ranked according to their relevance to the behavior of the target phenomenon, thus might reduce communication requirements drastically.

In this paper, we propose the information regularized maximum likelihood (IRML) estimation principle for target localization with binary detection sensor networks. The IRML principle utilizes pairwise mutual informations between sensor measurements to condense the prior information into weights that are utilized to modify the relative importance of each sensor in the ML estimation procedure. In the target localization framework using motion sensors, the IRML framework is shown to improve localization accuracy significantly at minimal additional computational load per sensor to evaluate the pairwise mutual information matrix.

2. PROBLEM STATEMENT

We specifically consider the problem of target localization and tracking with distributed motion detector networks. In this paper, we do not utilize the recursive Bayesian state estimation framework [12,13] and rather focus on instantaneous target localization using the sensor ensemble response. However, the temporal information is still incorporated in tracking by initializing the IRML algorithm to its estimate in the previous time step. Each detection sensor is assumed to be sensitive to target motion in a radially symmetric fashion with monotonically decreasing detection probability with increasing sensor-target distance. More formally, the conditional probability of the sensor *firing* (indicating detection) given target and sensor positions satisfies:

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

Typically, a one-sided Gaussian profile is assumed for this function as a convenient form, but not necessarily the best model for real sensors. Here, we employ the one-sided Gaussian model for sensor detection probability as well:

$$p(f=1 \mid \mathbf{x}_t, \mathbf{x}_s) = \alpha + (1-\alpha-\beta)e^{-\|\mathbf{x}_t-\mathbf{x}_s\|^2/h^2} \quad (2)$$

where h is the half-detection probability range ($\|\mathbf{x}_t-\mathbf{x}_s\|=h$ implies $p(f=1|\mathbf{x}_t,\mathbf{x}_s)=0.5$). Given a sensor network with sensor locations $\{\mathbf{x}_1,\ldots,\mathbf{x}_n\}$ and a firing pattern (snapshot) of the sensor network, the ML estimate for the target location is the solution to the following maximization problem:

$$\begin{aligned}\hat{\mathbf{x}}_t^{ML} &= \arg\max_{\mathbf{x}_t} \sum_{f_i=1} \log p(f_i=1 \mid \mathbf{x}_t, \mathbf{x}_i) \\ &\quad + \sum_{f_i=0} \log(1-p(f_i=1 \mid \mathbf{x}_t, \mathbf{x}_i))\end{aligned} \quad (3)$$

This ML solution does not take into account any prior knowledge about the target location distribution. The MAP estimate, on the other hand, would utilize such prior distribution information $p(\mathbf{x}_t)$ to augment (3) into:

$$\begin{aligned}\hat{\mathbf{x}}_t^{MAP} &= \arg\max_{\mathbf{x}_t} \sum_{f_i=1} \log p(f_i=1 \mid \mathbf{x}_t, \mathbf{x}_i) \\ &\quad + \sum_{f_i=0} \log(1-p(f_i=1 \mid \mathbf{x}_t, \mathbf{x}_i)) \\ &\quad + \log(p(\mathbf{x}_t)/p(f_1,\ldots,f_n))\end{aligned} \quad (4)$$

Neither (3) nor (4) have analytical solutions in general, therefore iterative optimization algorithms must be utilized. Furthermore, obtaining the prior accurately required for the third term in (4) is difficult at best. Assuming an inaccurate prior is known to bias the solutions towards erroneous estimates especially for smaller networks.

3. PAIRWISE MUTUAL INFORMATION REGULARIZED MAXIMUM LIKELIHOOD

Consider a distributed motion sensor network that is deployed to provide target detection and localization coverage over a region R. Suppose that the sensor network is exposed to a (quasi-) stationary target acceleration profile with distribution p(A), where the targets enter the region R from a subset B of the boundary of R according to a distribution $p(B)$. It is expected that as each target passes through R through some random trajectory T, the sensors in the network fire in a certain order consistent with the trajectory T. If T for each target is drawn from a stationary distribution, the sensor network will, on average, provide statistically *correlated* firing patterns. The term correlated is used in its broader meaning including nonlinear and discrete-valued correlations between point-processes. Let $f_i(T)$ be the random sequence of firing (0's and 1's) of sensor i as the target follows trajectory T. It is expected that the mutual information between the firing sequence of sensors i and j, denoted by $I(f_i(T),f_j(T))$ evaluated over many trajectories will be large for sensors that fire and remain silent simultaneously (sensors that detect the target at the same time) and will be small for sensors that detect the target in an asynchronous manner (with time lag). Two sensors that *never* detect the target (allowing occasional false detection firings) will exhibit almost zero mutual information. According to the sensing model (1), the expected pairwise mutual information graph (where I_{ij} is the edge strength) will connect neighboring sensors located at regions with high values of $p(\mathbf{x}_t)/p(f_1,\ldots,f_n)$ while isolating sensors in regions where this ratio is small.

By definition, the mutual information between sequences $f_i(T)$ and $f_j(T)$ is calculated using the joint and marginal distributions of these sequences: $p(f_i(T),f_j(T))$, $p(f_i(T))$, and $p(f_j(T))$ averaged over multiple target trajectories:

$$I(f_i(T), f_j(T)) = E\left[\log \frac{p(f_i(T), f_j(T))}{p(f_i(T))\,p(f_j(T))}\right] \quad (5)$$

In practice, this quantity between long sequences require the estimation of exponentially large discrete joint distributions (increasing in size with the length of the sequences). Assuming that each sensor exhibits independent firing behavior at each time step of 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 the firing behavior of each pair of sensors ae assumed to be independent from their pairwise behavior at previous/other time steps, therefore, we obtain the separability of the overall joint distribution into the product of pairwise joint firing behavior at each time step: $p(f_i(T),f_j(T))=p(f_i(T_1),f_j(T_1))\ldots\ p(f_i(T_N),f_j(T_N))$. These two assumptions simplify the computational and practical difficulties posed by the high dimensionality of (5) and reduces to bivariate mutual information:

$$\begin{aligned}I(f_i(T), f_j(T)) &\approx E\left[\log \prod_{k=1}^{N} \frac{p(f_i(T_k), f_j(T_k))}{p(f_i(T_k))\,p(f_j(T_k))}\right] \\ &= \sum_{k=1}^{N} I(f_i(T_k), f_j(T_k))\end{aligned} \quad (6)$$

In practice, the pairwise mutual information between sensors i and j is estimated using firing sequences obtained from both sensors as multiple *training* targets pass through the detection field (training targets can be previously tracked targets and the mutual information based pairwise statistical sensor similarity can be recursively updated in real time to handle nonstationary target behavior; this will be the topic of a future publication). Given concatenated firing sequences $f_i(\mathbf{T})$ and $f_j(\mathbf{T})$ for many target trajectories, 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.

Since a sensor that is at a central location for a given target dynamics is expected to have large mutual information with its nearby neighbors, a normalized importance weight based on the pairwise mutual informations in (5) can be assigned to each sensor

$$w_i = \frac{\sum_j I(f_i(T), f_j(T))}{\sum_j \sum_l I(f_l(T), f_j(T))} \quad (7)$$

The information regularized maximum likelihood (IRML) criterion for target localization is obtained by modifying (3) using the weights appointed to each sensor as in (7):

$$\begin{aligned}\hat{\mathbf{x}}_t^{IRML} &= \arg\max_{\mathbf{x}_t} \sum_{f_i=1} w_i \log p(f_i=1 \mid \mathbf{x}_t, \mathbf{x}_i) \\ &\quad + \sum_{f_i=0} w_i \log(1-p(f_i=1 \mid \mathbf{x}_t, \mathbf{x}_i))\end{aligned} \quad (8)$$

This modification corresponds to the following assumption regarding the joint data likelihood:

$$\begin{aligned}p(f_1,\ldots,f_n \mid \mathbf{x}_t, \mathbf{x}_i) &= \prod_{i=1}^{n} p(f_i \mid \mathbf{x}_t, \mathbf{x}_i)^{w_i\delta(f_i-1)} \\ &\quad \cdot (1-p(f_i \mid \mathbf{x}_t, \mathbf{x}_i))^{w_i\delta(f_i)}\end{aligned} \quad (9)$$

Consequently, the data likelihood becomes a regularized geometric mean where the weights are determined by statistical sensor fitness as measured by the pairwise mutual information matrix.

The IRML estimator is obtained by maximizing (8) using a fixed point iterative algorithm similar to the mean-shift clustering algorithm [14], which is equivalent to expectation maximization. For the firing model of (3) with $\alpha=\beta$, taking the gradient of (8) with respect to $\mathbf{x}_t$, equating to zero and rearranging terms yields the following simple iterative optimization rule, which is optimized to the previous IRML estimate of the target position at every new estimation time step:

$$C_i = (1-\alpha-\beta)e^{-\|\mathbf{x}_t - \mathbf{x}_i\|^2/h^2}$$
$$\hat{\mathbf{x}}_t \leftarrow \frac{\sum_{f_i=1} w_i \mathbf{x}_i C_i + \sum_{f_i=0} w_i \mathbf{x}_i C_i}{\sum_{f_i=1} w_i h^2 (C_i + \alpha) + \sum_{f_i=0} w_i h^2 (1-\alpha-C_i)} \tag{10}$$

4. EXPERIMENTAL RESULTS

In this section we present experimental results for a network consisting sensors with detection probabilities of the form given in (1). We compare results obtained using the maximum likelihood estimator of (3) and the information regularized maximum likelihood estimator of (8). For illustration simplicity, we utilize uniform sensor grid models in the following experiments, however, this uniformity is not necessary and sensors can be randomly deployed in the detection field according to any distribution without much consequence (provided that some sensors are actually placed at positions where they can sense the target).

We start with the comparison of the performances of ML and the proposed information regularized ML estimators. The mean localization errors are averaged over 100 Monte Carlo runs using the following random walk target trajectory model

$$\mathbf{x}_t(k+1) = \mathbf{x}_t(k) + \boldsymbol{\mu} + \mathbf{v} \tag{11}$$

where $\boldsymbol{\mu}$=$[0.01,0]^T$ and $\mathbf{v}$ is a circular bivariate Gaussian random variable with covariance ($0.005^2\mathbf{I}$). Keeping the miss probability constant at β=0.1, a 10×10 uniform sensor grid is trained on a 1×1 square detection field. The mean localization error of ML and IRML estimators are presented as a function of varying false detection probability of the sensors. The IRML estimator significantly outperforms the ML estimator in target localization. Figure 2 illustrates the distribution of target trajectories for 20 representative Monte Carlo runs. For the same random walk model, Figure 3 illustrates the behavior of information regularization weights over the sensor network (actual physical location) for a 20×20 uniform grid of sensors. Although we cannot offer an analytical proof at this time, visual inspection in numerous Monte Carlo experiments with different random walk models demonstrate that the spatial weight distribution approximates the target prior distribution closely. The sensors that lie in a portion where the target is present with a higher probability, we observe higher pair-wise mutual information; hence, higher weights. The sensors where the pdf is low, the outputs are mainly false alarms and the corresponding weights are low, suppressing misleading sensor activity.

In a second comparison of IRML and ML, we compare the mean localization error versus density of sensors (simulated by keeping the number of sensors fixed and reducing the detection field area, sensor half-detection distances, and the random walk mean and standard deviation proportionally). In this experiment, the false alarm and miss probabilities are both set to 0.1. Figure 4 shows the average localization error over 100 Monte Carlo simulations at each sensor density level. IRML significantly outperforms ML and the performance gap increases for denser networks.

5. DISCUSSIONS

The advent of cheap, low-power, reliable sensors enable the exploitation of dense sensor networks for solving various traditional problems such as detection, localization, and tracking in a distributed manner. Bayesian techniques for sensor fusion in idealized static situations such as no-sensor-failures, no-a-priori-target-information, and no-sensor-self-organization have been studied in detail and are now well understood. Contemporary challenges in sensor network research include incorporating machine learning techniques and statistical reliability measures into data fusion criteria in order to enable sensors to exploit patterns in past-observed activity of the target phenomenon and decentralization and optimization of decision fusion under various power and bandwidth constraints. In this paper, we proposed the information regularization principle based on pairwise sensor activity mutual information estimates in order to tackle the first issue mentioned above.

The information regularization principle is based on the assumption that the observed distributed phenomenon exhibits spatial 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 spatial 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.

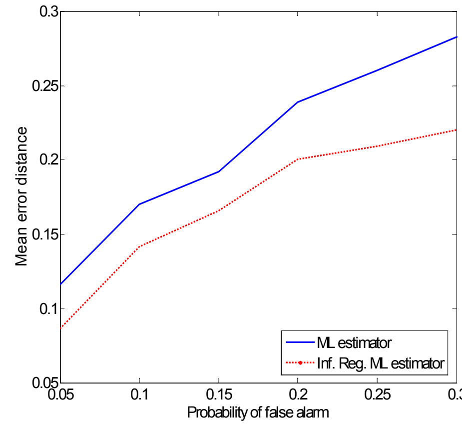

Figure 1. Mean of the distance between the actual target position and the estimated location in 100 Monte Carlo simulations for a 10×10 uniform sensor grid as a function of false alarm (detection) probability: ML (solid), IRML (dashed).

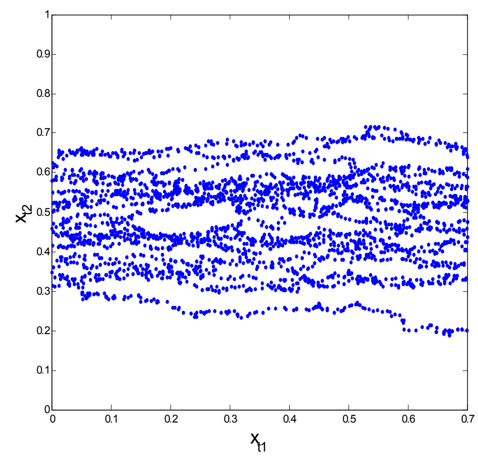

Figure 2. Target trajectories using the random walk model given in (11) for 20 sample Monte Carlo simulations.

The redundancy between sensors can be exploited in a variety of ways depending on the application and the circumstances. For example, one can utilize the pairwise mutual information graph as a similarity matrix between the sensor pairs, thus enabling the abstract clustering of sensors into *similar* sensor groups, where similarity is measured by the amount of common information. Such clustering enables us to reorganize the sensor network more efficiently to maximize the information diversity of the overall network under the given target phenomenon statistics. This network reorganization could be achieved in a self-organizing manner if the sensor units have the ability modify their sensing parameters including their physical location and orientation. The information-cluster structure in the network could also be exploited to identify sub-networks that could form automatically to minimize unnecessary network-wide communications and enable a hierarchical decision fusion scheme, as well as enabling selective sensor communication based on past information relevance. Alternatively, the sensor similarity can be utilized to identify sensor failures. Temporal evolution of the pairwise mutual information matrix can be exploited to identify sensors that change their co-sensing behavior unexpectedly, possibly due to a failure.

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 collective activity, under the assumption that these sensors are most likely sensing relevant and consistent information. The 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 maximum likelihood significantly. Future work will focus on extending the application of this principle to potential problems discussed above, as well as formulating a stronger connection between the pairwise mutual information based weights and Bayesian priors on target phenomena.

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

REFERENCES

[1] P. K. Varshney, *Distributed Detection and Data Fusion*, Springer-Verlag, New York, 1997.

[2] John N. Tsitsiklis, “Decentralized detection by a large number of sensors,” Mathematics of Control, Signals, and Systems, vol 1., pp. 167-182, 1988.

[3] J. C. Chen, K. Yao, and R.E. Hudson, “Source location and beamforming,” IEEE Signal Processing Magazine, vol. 19, pp. 30-39, 2002.

[4] R. R. Brooks, P. Ramanathan, and A. M. Sayed, “Distributed target classification and tracking in sensor networks,” Proceedings of IEEE, vol. 91, no. 8, pp.1163-1171, Aug. 2003.

[5] A. Baptista, T. Leen, Y. Zhang, A. Chawla, D. Maier, W. C. Feng, W. C. Feng, J. Walpole, C. Silva, J. Freire, “Environmental observation and forecasting systems: vision, challenges and successes of a prototype,” Encyclopedia of Physical Science and Technology (R. A. Meyers, Ed.), Academic Press, Third Edition, Vol. 5., pp 565-581.

[6] H. Jimison, M. Pavel, J.McKanna, J. Pavel, “Unobtrusive monitoring of computer interactions to detect cognitive status in elders,” IEEE Transactions on Information Technology in Biomedicine, vol. 8, pp.248-252, 2004.

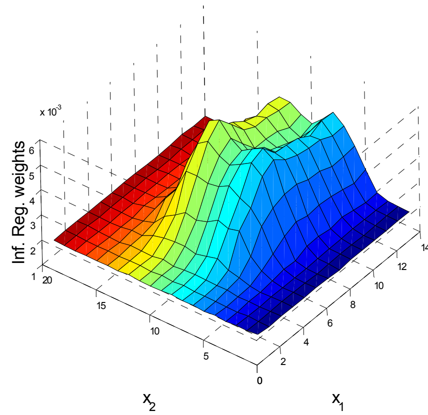

Figure 3. Information regularization weights over the sensor field using 100 realizations of the random walk model in (11) as training data.

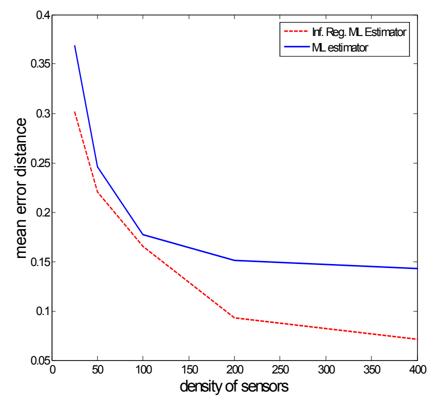

Figure 4. Mean error distance for sensor networks of different spatial density: ML (solid), IRML (dashed).

[7] Z. Chair, and P. K. Varshney, “Optimal data fusion in multiple sensor detection systems,” IEEE Trans. AES, vol. AES-22, no.1, pp. 98-101, Jan 1986.

[8] A. Artes-Rodriguez, R. Torres-Lopez, and L. Tong, “Distributed detection by multiple tests in sensor networks using range information,” Proceedings of EUSIPCO, Sept. 2004.

[9] A. Artes-Rodriguez, “Decentralized detection in sensor networks using range information,” Proceedings of ICASSP, vol. 2, pp. 265-268, 2004.

[10] S. Haykin, “Cognitive radar,” IEEE Signal Processing Magazine, vol.23, pp. 30-40, 2006.

[11] L. Doherty, B. A. Warneke, B. Boser, and K. S. J. Pister, “Energy and performance considerations for Smart Dust,” Int. Jour. of Parallel and Distributed Sensor Networks, vol. 4, no. 3, 2001.

[12] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, New York, 2001.

[13] E. A. Wan, and R. van der Merwe, “Kalman filtering and neural networks”, Chapter 7: The Unscented Kalman Filter, *Wiley Publishing*, Eds. S. Haykin, 2001.

[14] U. Ozertem, D. Erdogmus, “Spectral Clustering with Mean Shift Vector Quantization,” Proceedings of MLSP’05, Mystic, Connecticut, 2005.