INTERIM PROGRESS REPORT

Period covered by report: 22 August 2011 to 31 July 2012

Proposal Title: (MURI) Value-centered Information Theory for Adaptive Learning, Inference, Tracking, and Exploitation

Contract/Grant #: W911NF-11-1-0391

Authors of report: Alfred O. Hero (PI), U of Michigan
Douglas Cochran, Arizona State U
Emre Ertin, Ohio State U
John W. Fisher III, MIT
Jonathon How, MIT
Michael Jordan, UC Berkeley
Randolph L. Moses, Ohio State U
Raj Rao Nadakuditi, U of Michigan
Stefano Soatto, UCLA
Alan Willsky, MIT

ARO proposal number: 59712-CS-MUR

Principal Investigator:
Alfred O. Hero, Professor
Department of Electrical Engineering and Computer Science
The University of Michigan
1301 Beal Ave
Ann Arbor, MI 48109-2122
Tel: 734 763 0564
Email: hero@umich.edu

Submitted to: Dr. Liyi Dai
Program Manager
Army Research Office
Raleigh, NC

Abstract

Our aim is to develop a comprehensive set of principles for task-specific information extraction and information exploitation that can be used to design the next generation of autonomous and adaptive sensing systems. The significance of this research is that it addresses the widespread and longstanding problem of defining, assessing, and exploiting the value of information in active sensing systems. This year we report progress in 21 areas organized around 3 main thrusts: learning and representation of high dimensional data, distributed information fusion, and active information exploitation. In the learning and representation thrust, progress ranges from assessing value of Kronecker representations of high dimensional covariance matrices to learning to rank user preference data, an important task for human-in-the-loop decision systems. In the distributed information fusion thrust, progress is reported in assessing value of information in distributed information gathering and dimensionality reduction systems with application to sensor networks. In the active information exploitation thrust, progress is reported in information geometric trajectory planning, adversarial information collection, active learning in Bayes nets, and multistage adaptive estimation of sparse signals. Our future plans are to continue to develop linkages between these thrust areas, to continue our development of fundamental theory for designing and evaluating distributed active information collection systems, and to account for human interactions in the sensing and processing loop.

1 Overall objective of project

Sensing and actuation systems are inundated with diverse and high volumes of data. Much of this data is uninformative and irrelevant to the end task of the system which can evolve over the mission. The problem of extracting and exploiting the relevant and informative portion of sensor data has been an active area of research for several decades. Despite some progress, notably in information-driven tracking and data fusion, a general solution framework remains elusive, especially for autonomous and distributed sensing systems. The aim of this MURI is to develop a comprehensive set of principles for task-specific information extraction and information exploitation that can be used to design the next generation of autonomous and adaptive sensing systems. These principles will go beyond the standard information theoretic approaches that fail to account for non-classical information structures due to factors such as small sample size, poorly-specified target and clutter models, feedback control actions, hostile or adversarial environments, computation/ communication constraints, distributed sensing resources, and time-critical decision making.

2 Approach

Our research program aims to lay the foundations for a new systems theory that applies to general controlled information gathering and inference systems with mission planning. The research approach comprises three inter-related research themes that collectively address the most critical research challenges. These thrusts are: (1) information-driven structure learning and representation; (2) distributed information fusion for fast paced uncertain environments; and (3) active information exploitation for resource management. We aim to develop an end-to-end framework that will result in better raw sensor data acquisition and processing, improved fusion of multiple sources and modalities, and more effective sensor management and control.

3 Scientific barriers

This research addresses several challenges

1. Reliable value-of-information measures for active multi-modal sensing systems are not available. Existing approaches to learning and representation of information does not account for the sequential nature of data collection. This arises in active sensing systems such as autonomous maneuvering robots with vision/IR/LIDAR capabilities. Quantifying the value of information collected from active sensing systems is essential but there exists no suitable theory to do so. Classical Shannon information theory is inadequate as it was not designed for learning in active sensing systems; rather it was designed for data transmission in communications systems. A new theory for learning the value of information is needed that accounts for real-time feedback and control of the sensor, applies to signals that are non-linearly embedded in high dimension, accounts for models with complex structural components, e.g., hierarchical graphical models of interactions in the scene, has scalable computation even in

large distributed sensor systems, and accounts for the economic or human cost of acquiring data or fielding a new sensor.

2. There is no broadly applicable theory of information fusion for fast paced uncertain environments. The design and operation of sensing systems must accommodate the collection and delivery of a wide range of data at different times, spatial locations, and often with severely limited bandwidth and delay constraints. These systems must not have too many user defined tuning parameters that could overwhelm the human operator. There is no generally applicable theory of multimodal information fusion that accounts for all of these factors. Existing information theoretic measures and associated surrogates, are often only weakly predictive of information fusion performance and usually require careful tuning when used as objective functions that drive the fusion algorithm, Reliable measures are needed for fusion in compromised environments having high background/clutter variability and spotty situational awareness coverage.

3. Most information exploitation algorithms do not accurately predict the ultimate value of a current sensing or navigation action in the presence of uncertain hostile environments. The sensor manager plans ahead and controls the degrees-of-freedom (actions) of the sensor and platform in order to achieve system objectives. These degrees of freedom include: region of focus of attention, choice of modality and mode (for example EO vs LIDAR), transmit waveform selection, and path planning actions (platform maneuvering). The manager must predict the value of information resulting from each of the candidate sensing actions. This prediction must account for the uncertainty of the environment, time-varying visibility constraints (target obscuration), erratic or adversarial target behavior, and sensor resource constraints. To date most plan-ahead sensing and navigation approaches have been based on heuristics, like maximizing Shannon information-gain, and do not account for the value of the information measure as a function of the end task or the uncertainty in the environment.

4 Significance

The significance of this research is that it addresses the longstanding problem of defining, assessing, and exploiting the value of information in active sensing systems. By defining new information measures that account for the future value of data collection we can design better sensing, fusion, and planning algorithms that come with performance guarantees, e.g., tight value-specific bounds and performance approximations. By developing scalable and accurate methods to assess the value of information from empirical data, we can better design active sensor fusion and sensor planning to exploit the information collected thus far. The impact of the research can be summarized by the following four points.

1. The research will result in more accurate prediction of performance using a new class of information measures that account for both quality and value of information.

2. The research will provide a foundational ”systems theory” for active information gathering systems that use these new measures.

3. The research will use this foundational theory to develop highly adaptive and learning-based sensing strategies with significantly enhanced performance having reduced user tuning requirements.

4. The research will apply these sensing strategies to improve sensor signal processing, information fusion, and sensor platform navigation and control

5 Specific accomplishments over the period (8/22/11-7/31/12):

Our efforts are organized around the three research thrusts (1) information-driven structure learning and representation; (2) distributed information fusion; and (3) active information exploitation for resource management. These thrusts are interdependent and most of our efforts fall across the boundaries between them. However, for clarity in presentation we cluster each reported progress and accomplishment around one of these thrusts.

5.1 Information-driven structure learning and representation

As learning and feature representation is one of the basic building blocks of fusion and resource planning, this thrust is crucial for maximizing the value of information collected by a sensing, processing, and decision-making system. Our effort in information-driven learning and representation encompasses three areas: learning and representation of high dimensional data, value of information for vision, and performance quantification for estimation tasks.

5.1.1 Learning and representation of high dimensional data

Contributors:	Hero UM, Jordan UCB, Fisher MIT
Publications:	[17], [19], [24], [3]

Feature representation and learning for high dimensional data is the starting point for studying value of information. Poorly chosen features will deprive the system of the ability to extract useful information for fusion, inference or resource planning. Poor learning rates will make the system unable to adapt to fast changing scenarios. We have made significant progress in both of these domains. This progress is described below.

Progress 1: Sparse Kronecker product representations for high dimensional covariance matrices (Hero UM): Kronecker product representations decompose a high dimensional covariance matrix $\mathbf{\Sigma}$ into an outer product of lower dimensional matrices $\mathbf{A}$ and $\mathbf{B}$, i.e., $\mathbf{\Sigma} = \mathbf{A} \bigotimes \mathbf{B}$. This type of representation has been previously proposed for spatio-temporal network sensor data, e.g., by Werner *et al*in 2008, and results in a significant reduction of the number of unknown covariance parameters. For example, if the number of sensors is p, $p = nm$, and n and m are the dimensions of the Kronecker factors $\mathbf{A}$ and $\mathbf{B}$ the number of unknowns is reduced from $O(m^2n^2)$ to $O(m^2)+O(n^2)$.

As a more concrete example, consider 1000 sensors measuring 1000 time samples of a spatio-temporal random field. The covariance matrix of all the data will be a 1,000,000 by 1,000,000 but under the Kronecker product model can be completely specified by two smaller 1000 by 1000 covariance matrices.

In our recent papers [17] and [19] we have proposed two algorithms for estimating the parameters of Kronecker factors under the matrix normal Gaussian model when the covariance is both sparse and Kronecker structured. The convergence rate of these algorithms was proven and the algorithms were shown to yield consistent estimates. More importantly, as concerns the VoI topic of this MURI grant, in [18] we obtained tight bounds on the rate of convergence of the MSE as a function of the number n of available observations. This establishes the incremental value of observations in terms of reduction of MSE when the covariance has Kronecker product structure. Figure 1 shows a comparison of the convergence region of our algorithm as compared to different algorithms. Our proposed algorithm has the fastest rate of MSE convergence. The work described in [18] has been submitted to a journal.

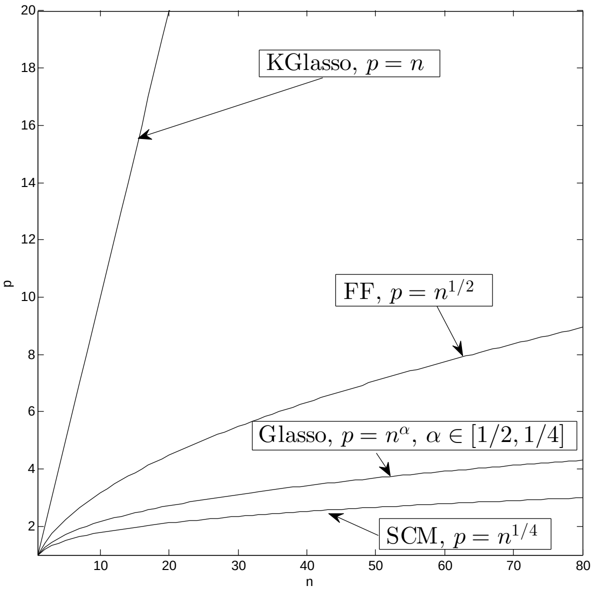

Figure 1: *Illustration of advantages of the proposed sparse Kronecker structured covariance estimation algorithm (KGlasso [19]) as compared to other state of the art algorithms for estimation in spatio-temporal graphical models. Regions of convergence for these algorithms were established in [18] and are upper bounded by each respective curve including: KGlasso (converges in region below upper curve), non-sparse structured Flip Flop estimator (below second highest curve), non-Kronecker structured Glasso (below third highest curve), and the standard non-sparse and non-Kronecker sample covariance matrix estimator (SCM) (below bottom curve). Our proposed KGlasso [19] has the largest region of convergence implying that it can estimate covariance matrices of high dimension (p) with many fewer observations (n) than the other methods.*

Progress 2: Sparse learning in sensor networks (Jordan UCB): Data accumulation and modeling in large sensor networks often run aground on problems of high dimensionality. In such setting it is essential to learn sparse solutions that retain a small subset of the available predictors. To address such sparse regression problems, we have proposed a new family of sparsity-inducing priors that we refer to as *exponential power-generalized inverse Gaussian (EP-GIG)* distributions [24]. Our framework allows us to induce non-convex losses which are effective in obtaining sparse solutions. We have also developed hierarchical Bayesian methods based on these loss functions.

Progress 3: Structure learning for multivariate time series (Fisher MIT): We have extended prior work on learning the structure of interactions between vector-valued time-series. Inference over interactions structures is super-exponential in the number of time-series (i.e., $O(N^N)$). As such, exact inference presents a significant computational challenge. We have extended our prior work on exact polynomial-time algorithms for Bayesian inference over structure - so called, switching temporal interaction models (STIMs) - to include state space models. Prior methods required time-series to be fully observed, whereas the new method allows for noisy observations and (importantly) missing data. The computational complexity of the new formulation remains polynomial, $O(N^k)$, where N is the number of time-series under consideration and k is number of allowed interactions.

Progress 4: Learning to rank from partial user preference data (Jordan UCB): When evaluating proposed actions in uncertainty environments, users often prefer to receive a ranked list of the options, and are most interested in the value associated with the actions near the top of the list. To address this need, we have been studying loss functions and the corresponding risks that are appropriate for supervised ranking [3]. In this problem the task is to rank sets of candidate items returned in response to queries. Although there exist statistical procedures that come with guarantees of consistency in this setting, these procedures require that individuals provide a complete ranking of all items, which is rarely feasible in practice. Instead, individuals routinely provide partial preference information, such as pairwise comparisons of items, and more practical approaches to ranking have aimed at modeling this partial preference data directly. As we show, however, such an approach has serious theoretical shortcomings. Indeed, we demonstrate that many commonly used surrogate losses for pairwise comparison data do not yield consistency; surprisingly, we show inconsistency even in low-noise settings. With these negative results as motivation, we present a new approach to supervised ranking based on aggregation of partial preferences and develop U-statistic-based empirical risk minimization procedures. We present an asymptotic analysis of these new procedures, showing that they yield consistency results that parallel those available for classification.

5.1.2 Value of information for vision

Contributors: Soatto UCLA
Publications: [5], [16]

In the area of value of information for vision the effort is to develop invariant representations of pose and appearance for shape spaces. Such representations can be used to specify subspaces over which information measures can be defined for assessing the value of information for sensor platforms with mobility. For such a representation the relevant observable information is defined on the pose/appearance quotient space of shapes. We are working on quantifying and minimizing the information gap between the quotient shape space and the complete shape space. This gap will specify the information lost due to object pose and appearance uncertainty.

Progress 5: Event detection for live fire monitoring systems (Soatto UCLA): In [16] we have tackled the problem of event detection for processes manifest in time-series that exhibit complex multi-scale temporal variability spanning from the minute to the year. A prototypical problem is *outdoor fire detection*: The data is affected by nuisance variability due to illumination, weather, terrain shape, etc. Simply “learning away” such a variability would require temporal observation over a span of at least one year (to capture seasonal variability), and the procedure would have to be repeated for each sensor (monitoring station). Furthermore, the change due to the onset of a fire is typically trumped by normal-mode variability, and no “positive examples” are available for the majority of monitoring stations. We have therefore set out to study systematic ways to *canonize* nuisance variability that can be factored out from the data via pre-processing in a lossless fashion: *model* nuisance variability that can be easily normalized with side-information (e.g. the dime of the day, or the location of the horizon); and *learn* the residual variability (e.g., cast shadows). The results, published at ECCV [16], represent the basis for a live fire monitoring stations that has benchmarked against human performance, beating it not only for untrained subjects, but also for expert fire monitors.

Progress 6: Salient motion detection from a moving sensor platform (Soatto UCLA): In [5], published at CVPR, we have tackled the problem of "salient" event detection from a moving sensor. If the sensor was static, relatively simple statistical models of the "background" can be used as a test against the hypothesis of an alarm ("foreground"), as customary in surveillance. However, when the camera is moving, the entire domain of the image is deforming, and explicitly modeling the deformation due to the background is highly problematic, because – as shown by Sundaramoorthi *et al*2009, – even a single static scene can induce on the domain of a moving image a deformation that spans the entire group of diffeomorphisms. The situation is even more complex when occlusion phenomena occurs, in which case the deformation cannot even be represented as the graph of (however complex) a function. We have thus defined "salient" motions in the scene that are not compatible with the motion of the camera. Both, however, are unknown, and we were determined to determine saliency *without* explicitly estimating the motion of the camera or the independent motion of objects in 3-D. Using the properties of (infinite-dimensional) orthogonal projector operators, we were able to detect independently moving objects as a violation of the assumption of epipolar domain deformation, but without actually computing the camera motion. We have also discovered empirically, and therefore proved analytically, that absence of knowledge of camera calibration does not affect the classification (saliency), even though of course it would affect the estimate of motion. The resulting system has shown effective at detecting motion of even very small objects seen from a moving platform (e.g., vehicles moving as seen from an aerial platform).

5.1.3 Performance quantification

Contributors: Jordan UCB, Cochran ASU
Publications: [6], [2]

Value of information is closely connected to performance quantification. Performance quantification is divided into 1) offline performance benchmarking and analysis and 2) online empirical performance prediction from measurements. We have made significant advances in both of these areas.

Progress 7: Performance quantification by BLB (Jordan UCB): In the area of performance quantification we have developed a new approach to empirical estimation of estimator uncertainty using a "bag of little bootstraps" (BLB) [6]. Our sampling approach allows one to use compact representations of the resamples to provide major reductions in the computational complexity of the bootstrap while preserving its asymptotic theoretical guarantees. Recently we have completed the development of a large-scale distributed implementation of BLB using the Spark cluster computing framework. This framework allows us to either read data from disk (making its performance comparable to traditional Hadoop MapReduce) or use a cluster of compute nodes to cache it in memory (when such an amount of memory is indeed available) making repeated accesses of the data significantly faster. As a result, in [6] we were able to analyze datasets of size 150GB, which would be computationally intractable for traditional bootstrap. In ongoing research we are developing a system for applying BLB to the estimation of quality of query processing algorithms in very large databases. The goal of this work is to not only return BLB-based estimates when to a wide range of queries, but to also automatically identify statistical situations where bootstrap and hence BLB cannot be applied.

Progress 8: Quantifying VoI for networked resources (Cochran ASU): In the area of quantifying the value of information sharing among networked resources, we have drawn on an example from multiple-channel signal detection. In this setting, detection performance in classical terms (e.g., probability of correct detection obtainable given a particular probability of false alarm) is precisely known as a function of signal-to-noise rations on the channels. But these performance figures are based on the assumption that all the sensor data is accumulated in a single location for processing. In a network of sensors where only pairs of sensors sharing an edge in the network graph can exchange data directly, performance loss relative to a fully connected network can be assessed by using maximum-entropy values in place of inter-node data for sensors that cannot directly share data. In initial work [2] we have been able to empirically ascertain the equivalent increase in signal strength (i.e., in dB) corresponding to adding links in the network graph. We are proceeding to study the relative value of various links in the network topology, with the objective of being able to determine precisely which information links are most valuable in terms of overall detection performance.

5.2 Distributed information fusion

Depending on the fusion algorithm implemented, fusion of information across a network of sensors can either enhance or degrade the value of information collected at each sensor. Furthermore, centralized fusion at a fusion center is often not possible due to the limited communications bandwidth and throughput of the network. Thus the study of effective distributed information fusion methods is a key part of the MURI grant's activities. Our progress in distributed fusion falls in three areas: information gathering and graphical models; non-commutative information theory for dimensionality reduction; and distributed inference in sensor networks.

5.2.1 Information gathering and graphical models

Contributors:	Hero UM, Fisher MIT
Publications:	[8], [15]

Graphical models provide a very natural mathematical framework for developing distributed fusion algorithms and studying their performance.

Progress 9: Directed graphical models for distributed PCA in a sensor network (Hero UM): We have applied directed graphical models to accelerate distributed inference in sensor networks [8]. Our previous approaches to graphical modeling for distributed PCA (DPCA) exploited sparsity in the inverse covariance (Wiesel *et al*2010). Under this MURI we have taken a significantly different approach in [8], called directed distributed PCA (DDPCA) that improves on the performance of DPCA. By assuming that the sparsity structure exists in the Cholesky factors of the inverse covariance, which is more natural for generative models of high dimensional time series, we show significant performance and computation/communication gains. The algorithm performs message passing to tie together the locally estimated principal components (PC) of the sensor network. We show that the local estimates converge to the global estimates, i.e., the one that one would find by a centralized PCA. The message passing algorithm is applied to detecting anomalies in networks. Our work establishes that directed graphical models can be used to extract better information from distributed networks for the purposes of anomaly detection and dimensionality reduction.

Progress 10: Submodularity for penalized information measures (Fisher MIT): In the area of information gathering and graphical models we are developing new lower information theoretic bounds on performance based on Ali-Silvey measures. These bounds exploit submodularity and adaptive submodularity and use the intrinsic partial order structure of graphical models. These bounds specify in objective measures that can be used to drive sensor planning algorithms. Submodularity ensures close-to-optimality of single stage planning and therefore avoids the high complexity of multistage optimal policy search. We are also developing a taxonomy of information measures to establish desirable properties of any such a measure in terms of its utility in fusion and plan-ahead sensing. Prior work exploiting the submodular property of conditional mutual information resulted in theoretical performance guarantees and a variety of both off-line and on-line bounds when comparing tractable greedy measurement selection to combinatorial (i.e. intractable) optimal measurement selection in the context of inference in graphical models. These results implicitly assume a homogenous cost structure over the set of all measurement choices. We have extended these bounds to include heterogeneous penalties to measurements so as to better capture real-world information gathering systems. We have established conditions under which submodularity holds for penalized information measures. Consequently, this work extends the use of such bounds to the resource-constrained sensor planning. This work has been published in [15].

5.2.2 Noncommutative information theory

Contributors: Nadakuditi UM, Jordan UCB
Publications: [14], [13], [7]

A major focus of our work is on applying the theory of random matrices to information fusion. We have made progress on developing new spectral measures and large deviation bounds for these non-commutative matrix measurement problems.

Progress 11: Spectral measures and subspace detection from random matrices (Nadakuditi UM): In the area of non-commutative information theory the effort is to establish the fundamental limits on the information that can be extracted from non-commutative observations such as random matrices and tensors. For symmetric matrices these limits are governed by the asymptotic behavior of eigenvalues and eigenvectors of the matrix and specify phase transition thresholds of SNR and matrix dimension for which these eigen-quantities cannot be reliably estimated empirically. Such phase transition thresholds are the key to developing the non-commutative information theory of dimensionality reduction, which is relevant, for example, to variable selection in sensor fusion. To that end, we have developed numerical algorithms for accurately computing the 'free' convolutions of various spectral measures. Since the accuracy of empirical subspace estimates depends on the analytic functions associated with these convolutions (analogous to the role of the Fourier transform/characteristic function in scalar probability theory), the developed algorithms will facilitate performance prediction for algorithms that utilize these subspace estimates. We have established universality of the square-root decay at the edge for free convolutions of compactly supported measures - this will form the basis of universal schemes for signal detection that exploit this characterization of the spectrum [14]. We have also analytically characterized the spectra of graphs with arbitrary expected degree distributions - this facilitates anomaly detection in graph-valued data sets and in spectral clustering for machine learning [13].

Progress 12: Large deviation theory for random matrices and matrix-valued time series (Jordan UCB): Random matrices arise in many statistical applications. The theory of random matrices is developing rapidly, but is still very limited in terms of its applications to statistics. In particular, the theoretical analysis of many learning algorithms requires tail bounds for statistical estimators and there are few such results available for random matrices. Indeed, the non-commutativity of matrices poses significant problems for classical large deviation theory and concentration of measure. In [7] we have shown how to use Stein's method of exchangeable pairs to obtain large deviation results for random matrices. For sums of independent random matrices our theory generalizes classic theory due to Hoeffding, Bernstein, Khintchine and Rosenthal to analogous matrix inequalities. We can also apply this theory to sums of dependent random matrices, thus making it more suitable to analyzing matrix-valued time-series.

5.2.3 Distributed inference in sensor networks

Contributors:	Moses OSU, Ertin OSU, How MIT
Publications:	[23], [11]

We are developing inference algorithms that can be implemented in a distributed manner over a network of sensors. Inference tasks of interest include detection and tracking. Our progress in these areas is reported below.

Progress 13: Aggregating local information for decision-level fusion (Moses OSU): In the area of distributed inference in sensor networks we are analyzing the interplay between local decision, global inference, performance, and communication. As an initial problem, we consider the case where each sensor makes a binary decision on the presence of a signal source and the fusion node combines these to make a more accurate decision [23]. We consider a lossy medium in which signals undergo a range-dependent propagation loss leading to a heterogeneous sensor network model. We analyze how the performance of the system scales with the density of sensor nodes in a random sensor network. Such a scaling analysis has to take into account the increased load on the communications network. Therefore, we analyze the detection performance as a function of sensor density subject to a constant network bandwidth constraint. As the number of sensors per unit area is increased, the average number of messages per sensor has to be scaled down to satisfy the network constraint. We consider two alternative methods of limiting bandwidth of individual nodes: random duty-cycling and increasing the local thresholds. First we consider a simple fusion rule of counting the number of detections in the sensor network. For this global fusion rule and modeling the sensor network as a stationary Poisson Point Process, the test statistic is also Poisson distributed. Thus, the system performance, in terms of Type I and Type II errors, is completely characterized by the conditional means of the count statistic. We showed that the policy of desensitizing sensors (i.e., increasing their thresholds) outperforms random duty-cycling sensor nodes while increasing density to maintain a desired false alarm probability at the fusion node. Next, we extend our results to show that desensitizing the sensors is also beneficial when optimal fusion rules (in contrast to counting rule) are utilized at the fusion center. For the asymptotic regime we also derived the relation between the average information provided by an individual sensor node and the global information available to the fusion center under network constraints.

Progress 14: Adaptive VOI based Algorithms for Efficient Distributed Information Fusion (How MIT): Many distributed information fusion algorithms have recently investigated the notion of *consensus* for state/parameter estimation, e.g. work by Olfati *et al*and Boyd *et al.* However, due reaching consensus using these algorithms can be communication intensive. In particular, it is often the case that not all agents have valuable information to contribute at all times (e.g., the updated state estimate after new measurements are collected may not be sufficiently different than the preceding estimate). To address the inefficiency of these communication intensive algorithms, we have focused on efficient and adaptive techniques for distributed data fusion that reduces network communication cost by restricting communications to those that have high information value. In [11] we showed that this framework can be used for distributed estimation of hyperparameters of a conjugate prior distribution by taking into account the *Value-Of-Information* (VOI) as determined by an appropriate metric. Agents that have a VOI that exceeds a threshold are termed as *informative* agents, and only informative agents communicate their information. A significant contribution of the work in [11] is an online adaptive VOI based distributed hyperparameter estimation algorithm that adjusts the threshold adaptively to strike a balance between communication required and accuracy of the posterior estimate. Communication cost analysis indicates that the proposed algorithm significantly outperforms the Hyperparameter

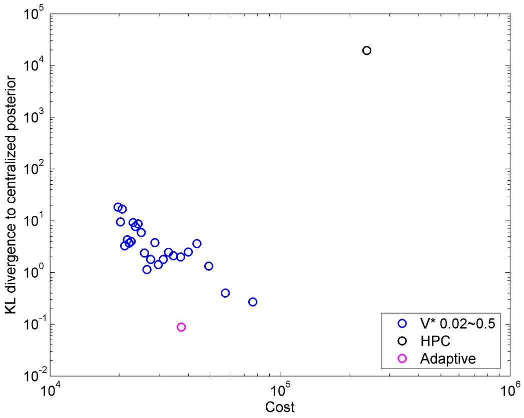

Figure 2: *Performance Comparison of different hyperparameter estimation algorithms for estimating average room temperature at a given time of the day from the Intel dataset. The total communication cost incurred is on the X-axis and KL-divergence to centralized (ideal) posterior is plotted on the Y axis. Adaptive distributed fusion algorithm which adaptively adjusts the VOI threshold V^* outperforms HPC both on accounts of accuracy and cost.*

Consensus (HPC) algorithm (Fraser *et al*2010).
Over a period of time $[0, T]$, the communication
cost of the HPC algorithm is directly proportional to $T \sum_{i=1}^{N} C_i$, where N is the number of agents in the network, and C_i is the cost of sending one message. For a network with a fixed number of agents, the communication cost with HPC increases linearly with time. The communication cost of the new VOI based algorithm is proportional to $\sum_{t=1}^{T} |\nu[t]| \sum_{i=1}^{N} C_i$, where $|\nu[t]|$ is the number of informative agents that send out messages at time t. For a fixed VOI threshold, it can be shown that $|\nu[t]| \to 0$ over time when estimating the hyperparameters of a static distribution. Therefore, the number of informative agents drops to zero over time, leading to a drop in the rate of growth of communication cost. However, the long term estimation performance of the algorithm may suffer if the VOI threshold is too high. Therefore, a time-dependent trade-off exists between the desirable amount of communication and the accuracy of the estimates. Numerical simulations indicate that the adaptive VOI based distributed hyperparameter estimation algorithm incurred approximately $1/8^{th}$ the communication cost of HPC, while arriving at a very close estimate of the hyperparameters (the difference between KL divergence error to centralized posterior of HPC and adaptive was 0.0325 nats). The algorithm was also tested on experimental data by using the Intel temperature dataset (Guestrin 2005), where it resulted in an estimate within 0.06% of the centralized average room temperature at a given time of the day while incurring only $1/10^{th}$ the communication cost.

Figure 2 compares the performance of different hyperparameter estimation algorithms for estimating average room temperature at a given time of the day from the Intel dataset. The total communication cost incurred is on the X axis and KL-divergence to centralized (ideal) posterior is plotted on the Y axis. The blue circles show the performance of VOI based distributed information fusion algorithms for different VOI thresholds V^*. The tradeoff between cost and accuracy is evident from the spread of the points for different values of V^*. The adaptive distributed fusion algorithm dynamically adjusts V^* based on the instantaneous communication load and can be seen to outperform non-adaptive VOI base algorithms and HPC both on accounts of accuracy and cost.

These results indicate the feasibility and suitability of VOI based distributed parameter More recent work has concentrated on the theoretical properties of the proposed distributed hyperparameter estimation algorithm. In particular we have established strong results concerning almost sure convergence of the communication cost and the estimation error to zero for distributions in the exponential family [12]. This work contributes significantly to the goal of this MURI because it uses VOI concepts to provide a more efficient framework for performing distributed parameter estimation than existing consensus or particle filter based approaches, e.g., Hall *et al*, Berg *et al*, Durrant Whyte *et al*, and others. Furthermore, it is significant to the distributed estimation literature because it extends the notion of *censoring* marginally useful information in a centralized estimation framework, e.g., Giannakis *et al*, Willsky *et al*, and others. to VOI based self-censoring in an distributed estimation framework. It is also a significant contribution to the study of distributed hyperparameter estimation as the developed algorithms have been shown to greatly reduce communication cost without compromising the accuracy of hyperparameter estimates for distribution in the exponential family. The algorithms discussed here, and their possible variants, would translate to significant resource savings in real-world distributed sensing applications by preventing irrelevant and marginally useful information from cluttering the network. This work therefore makes progress towards the objective of developing the next generation efficient and accurate information extraction systems.

5.3 Active information exploitation for resource management

The flow of information through a controllable sensor system affects decisions on control actions that can enhance target detection or tracking performance. The active information exploitation thrust lies at the interface of estimation and control and we have made progress in three subareas: information-driven sensing; robust adaptive planning of sensing actions; and active learning in Bayes nets.

5.3.1 Information-driven sensing

Contributors:	Ertin OSU, Cochran ASU
Publications:	[4], [10]

Information can be used as a metric for sensing in situations where there are non-classical information flows. Non-classical information flow occurs when there is an adversary who tries to hinder the system's attempts to maximize the flow of information from the target to the decision maker. Information flow is also non-classical when control actions are introduced into the process of sensing, processing and decision making. In the latter case, we think that differential geometry of the statistical manifold can capture the effect of these control actions on the value of information. We report two areas of progress: adversarial information structures and information geometric planning.

Progress 15: Adversarial information structures (Ertin OSU): In the area of adversarial information structures we are investigating the degree to which an intelligent target can reduce information collection efficiency of the sensor system. We are developing information-theoretic uncertainty measures, such as generalized entropy, to provide a foundation for quantifying value of information in adversarial situations. This quantification will translate into a bound on guaranteed detection/estimation performance and may allow adversarial modeling to be gainfully incorporated into information-driven planning algorithms. We consider the problem of sensor selection for a binary hypothesis testing problem when the conditional density of the sensor readings can be affected by an adversary [4]. A typical application of the proposed setup is surveillance with spatially distributed sensors, where the adversary is changing locations to evade detection. We model the sensor selection problem as a game between two players with opposing objectives. The observer is choosing an open loop randomized strategy to choose sensor observations that maximizes probability of detection, whereas the target is using an open loop randomized control strategy over the available evading actions to minimize the probability of being detected. The payoffs are specified by the asymptotic detection probability under a false alarm constraint. We prove the existence of the Nash equilibrium of this surveillance game. The Nash equilbrium of this zero-sum game provides optimal strategies for surveillance and evasion and the value of the game quantifies the guaranteed performance of the surveillance system. We characterize the optimal min-max strategies and the value of the game. We derive robust sensor selection strategies which provide max-min performance guarantees for detection probability.

Progress 16: Value of information on statistical manifolds (Cochran ASU): In the area of information-geometric trajectory planning we are applying methods of differential geometry to better capture the flow of information over the planning horizon [10]. The enabling observation in this approach is that a parameter estimation problem based on sensor data imposes a Riemannian structure on the parameter manifold via the Fisher information. Managing sensors thus entails navigation on the space of all Riemannian metrics on the parameter manifold, which is itself a Riemannian manifold. A particular sensor scheduling or navigation policy specifies a time evolving curve on a physically feasible submanifold of this manifold of sensor configurations, which captures the intrinsic geometry of the information collected. Through this formalism, the Fisher-Riemann metric, which is equal to the incremental Kullback-Leibler distance at two successive time points, enables specification of geodesic curves of minimal "information distance" over the manifold feasible sensor configurations. We are proceeding to study the quantification of information corresponding to differential-geometric properties of trajectories in this manifold.

An operator-theoretic framework for design of waveforms for adaptive radar applications, initially described in [1], is anticipated to provide a suitable framework for instantiating the information-geometric resource scheduling theory being developed under this thrust.

5.3.2 Robust adaptive planning of sensing actions

Contributors: Hero UM, Moses OSU, Ertin OSU
Publications: [22], [23]

By closing the loop between sensing, processing and decisions one can exploit measurements and models to significantly enhance performance under resource constraints. Planning ahead using predictive models results in sensor actions that are informed by previous measurements and that use all available resources most efficiently. We have made progress in two areas of robust adaptive sensor planning described below.

Progress 17: Multistage adaptive estimation of sparse signals under energy constraints (Hero UM): We have developed a new method for robust scheduling over a large number of sensor actions under constraints on effort (time, energy, resources). The method uses sequential adaptive processing and a sparsity assumption: only a small number of sensor actions will provide information gain. This problem arises in wide area search and tracking, sensor selection, waveform selection, and other relevant scheduling problems. A simple surrogate cost function is proposed that trades off false positives, false negatives and energy in a manner similar to our previous approaches to adaptive resource allocation policies, i.e., ARAP by Bashan *et al.* However, these previous methods did not come with theoretical guarantees on the task-related value of information of these policies. In our recent paper [22] we adopt a framework of open-loop feedback control (OLFC) and show that optimal multistage scheduling can be accomplished using dynamic programming under broad conditions. Furthermore, we show that this policy results in information gains that monotonically increase at each stage of the multistage planning algorithm. For the problem of estimating the amplitude of a signal with unknown but sparse support, implementation of the multistage policy has significant performance gains as compared to the state-of-the-art distilled sensing algorithm of Nowak *et al.* Furthermore, these gains approach the optimal oracle bound as the SNR increases.

Progress 18: Information-based sensor selection for target-tracking networks (Ertin OSU): We are developing information-based sensor selection strategies that retain past sensor measurements for target tracking. Sensor selection strategies for distributed tracking typically estimate the expected information gain of sensor observations to be made at the next time step and interrogate sensor nodes with the highest expected information gain to maximize tracking performance. In contrast, we study a distributed sensor network where nodes maintain a finite buffer of their past sensor observations. As a result, the fusion center can interrogate nodes for current as well as past measurements. Using the information form of the Kalman Filter we quantify the expected information gain of past and future measurements which have not yet been integrated into the state estimate of the tracker. We show that when the sensor coverage is sparse, past measurements can provide higher expected information gain than current measurements at times when the track uncertainty is large compared to single node coverage. Using the derived expected information gain metrics we design a novel sensor management strategy that can optimize sensor interrogations over the available history of measurements existing in the network.

5.3.3 Active learning in Bayes nets

Contributors:	Jordan UCB, Soatto UCLA
Publications:	[21], [9], [20]

In the area of active learning in Bayes nets we are investigating two avenues. Navigation and sensor planning in ground-air control architectures must provide optimal coordination of heterogeneous cooperating platforms in the presence of uncertainty. We are developing robust task allocation policies which will be tested on the MIT testbed in How's lab. We are also investigating ways to tractably quantify how well a candidate measurement will reduce uncertainty in a given subset of Bayes net variables. We plan on investigating this in the context of the MIT testbed to establish that we can improve inference, guidance and control performance in autonomous coupled sensing systems.

Progress 19: Bayesian active learning in sensor networks (Jordan UCB): One major area of research on active learning involves the use of crowdsourcing platforms like Amazon Mechanical Turk to provide actively sampled data. A major difficulty with crowdsourcing data is bias in worker responses. We have developed a Bayesian framework for coping with such bias [21]. Traditionally the effect of worker bias is mitigated using simple data curation techniques; e.g., taking majority votes of workers to decide on a label for an object. The implicit assumption used here is that all labels are being randomly generated from one latent true distribution. We instead model bias as a result of shared random effects. This allows us to analyze complex bias patterns arising in situations where the labeling task is hard or ambiguous. Also instead of going through the steps of data curation and learning, this model lets us combine the two into one single step. We intend to use this for data acquisition in large and noisy sensor networks, where data collected from neighboring sensors can be thought of as workers (sharing a number of random effects) providing labels for the same object.

Progress 20: Value of information for safe exploration by maneuvering platforms (Jordan UCB): In the area of path planning and platform maneuvering, we have developed a new framework where the value of information depends both on the amount by which it could improve and task performance and on the likelihood that the exploration agent be able to return to the start state. This type of exploration, which we call *safe exploration* [9], is instinctual to humans, and we would like automated agents to replicate this behavior. Our method applies to Markov Decision Processes, which can model a broad range of planning problems, and works by restricting the space of allowed exploration policies to only those that preserve ergodicity with high probability. To clarify, if either of these allowed exploration policies is interrupted at a random recall time, the agent will be able to return to the start state by following a return policy that our method also computes. In theory, our framework can extend any exploration method, since it does not specify how to choose among the allowed exploration policies. In practice, we shave shown that the ensuing constrained optimization problem can be solved efficiently for exploration methods that rely on adding reward bonuses to insufficiently explored states. Our terrain mapping experiments show the perhaps counter-intuitive fact that exploration becomes more efficient after adding our safety constrain in environments where the agent might become stuck when exploring greedily.

Progress 21: Information-maximizing control with visibility phenomena (Soatto UCLA): In [20] we have formalized the problem of information-maximizing control in the presence of topological uncertainty due to visibility phenomena (occlusions). We have shown that visibility can be computed efficiently, and that the resulting info-max control results in an optimal control problem where the state-space is the (infinite-dimensional) visibility functional. We have shown that this approach enables, in principle, a solution that is *linear* in the horizon (exploration length) rather than exponential as it is in the discretized equivalent partially-observable Markov decision process (POMDP). Unfortunately, however, the price to pay is an exponential memory complexity that makes the solution of the problem impossible for all but the most trivial cases. Therefore, in [20] we have proposed surrogate information measures for information gain due to visibility, and proved an upper bound in the exploration length (that is, exploration is proved to terminate in a finite number of steps, and the number of steps is bounded as a function of the complexity of the environment).

6 Publications

The following is a list of papers, theses, and other publications of research supported in whole or in part by this project.

[1] D. Cochran, S. D. Howard, and B. Moran, "Operator-theoretic modeling and waveform design for radar in the presence of doppler," in *Proceedings of IEEE-AESS RadarCon*, pp. 774–777, May 2012.

[2] D. Cochran, S. D. Howard, B. Moran, and H. A. Schmitt, "Maximum-entropy surrogation in network signal detection," in *Proceedings of the IEEE Statistical Signal Processing Workshop*, August 2012.

[3] J. C. Duchi, L. Mackey, and M. I. Jordan, "The Asymptotics of Ranking Algorithms," *ArXiv e-prints*, 2012.

[4] E. Ertin, "Sensor selection in adversarial setting," in *IEEE Statistical Signal Processing Workshop (SSP2012)*, Ann Arbor, MI, Aug 2012.

[5] G. Georgiadis, A. Ayvaci, and S. Soatto, “Actionable saliency detection,” in *Proc. of the IEEE Intl. Conf. on Comp. Vis. and Patt. Recog.*, 2012.

[6] A. Kleiner, A. Talwalkar, P. Sarkar, and M. I. Jordan, “The Big Data Bootstrap,” in *Proceedings of the 29th International Conference on Machine Learning (ICML)*, Edinburgh, UK, 2012.

[7] L. Mackey, M. I. Jordan, R. Y. Chen, B. Farrell, and J. A. Tropp, “Matrix Concentration Inequalities via the Method of Exchangeable Pairs,” *ArXiv e-prints*, p. 29, January 2012.

[8] Z. Meng, A. Wiesel, and A. Hero, “Distributed principal component analysis on networks via directed graphical models,” in *IEEE Intl Conf. on Speech, Acoustics and Signal Processing (ICASSP)*, Prague, April 2012.

[9] T. M. Moldovan and P. Abbeel, “Safe Exploration in Markov Decision Processes,” in *Proceedings of the 29th International Conference on Machine Learning (ICML)*, Edinburgh, UK, 2012.

[10] B. Moran, S. D. Howard, and D. Cochran, “An information-geometric approach to sensor scheduling,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 5261–5264, April 2012.

[11] B. Mu, G. Chowdhary, and J. P. How, “Efficient distributed information fusion using value of information based censoring,” Aerospace control laboratory technical report, Massachusetts Institute of Technology, http://hdl.handle.net/1721.1/71875, July 2012.

[12] B. Mu, G. Chowdhary, and J. P. How, “Efficient distributed information fusion using adaptive decentralized censoring algorithms,” in *American Control Conference (ACC)*, Washington DC, July 2013 (in preparation), IEEE.

[13] R. Nadakuditi and M. Newman, “Spectra of random graphs with arbitrary expected degrees,” *Arxiv preprint arXiv:1208.1275*, 2012. http://arxiv.org/abs/1208.1275.

[14] S. Olver and R. Nadakuditi, “Numerical computation of convolutions in free probability theory,” *Arxiv preprint arXiv:1203.1958*, 2012. http://arxiv.org/abs/1203.1958.

[15] G. Papachristoudis and J. W. Fisher III, “Theoretical guarantees on penalized information gathering,” in *Proc. IEEE Workshop on Statistical Signal Processing*, August 2012.

[16] A. Ravichandran and S. Soatto, “Long-range spatio-temporal modeling with application to fire detection,” *Proc. of the Eur. Conf. on Comp. Vision*, 2012.

[17] T. Tsiligkaridis and A. Hero, “Sparse covariance estimation under sparse kronecker product structure,” in *IEEE Intl Conf. on Speech, Acoustics and Signal Processing (ICASSP)*, Prague, April 2012.

[18] T. Tsiligkaridis, A. Hero, and S. Zhou, “Convergence properties of kronecker graphical lasso algorithms,” *Arxiv preprint arXiv:1204.0585*, April 3 2012.

[19] T. Tsiligkaridis, A. Hero, and S. Zhou, "Maximum-entropy surrogation in network signal detection," in *Proceedings of the IEEE Statistical Signal Processing Workshop*, Ann Arbor MI, August 2012.

[20] L. Valente, R. Tsai, and S. Soatto, "Information gathering control via exploratory path panning," *Proc. of the Conf. on Information Sciences and Systems (CISS)*, March 2012.

[21] F. L. Wauthier and M. I. Jordan, "Bayesian Bias Mitigation for Crowdsourcing," in *Advances in Neural Information Processing Systems 24*, J. Shawe-Taylor, R. S. Zemel, P. Bartlett, F. C. N. Pereira, and K. Q. Weinberger, editors, pp. 1800–1808, 2011.

[22] D. Wei and A. Hero, "Multistage adaptive estimation of sparse signals," in *Proceedings of the IEEE Statistical Signal Processing Workshop*, Ann Arbor MI, August 2012.

[23] G. T. Whipps, E. Ertin, and R. L. Moses, "Threshold considerations in distributed detection in a network of sensors," in *Proceedings of SPIE Ground/Air Multisensor Interoperability, Integration, and Networking for Persistent ISR III*, volume 8389, Baltimore, MD, May 2012.

[24] Z. Zhang, S. Wang, D. Liu, and M. I. Jordan, "EP-GIG Priors and Applications in Bayesian Sparse Learning," *Journal of Machine Learning Research*, vol. 13, pp. 2031–2061, 2012.

7 Statistics

1. Submissions or publications under ARO sponsorship during this reporting period. List the title of each and give the total number for each of the following categories:

   a. Papers published in peer-reviewed journals

      1. Z. Zhang, S. Wang, D. Liu, and M. I. Jordan, "EP-GIG Priors and Applications in Bayesian Sparse Learning," *Journal of Machine Learning Research*, vol. 13, pp. 2031–2061, 2012.

   b. Papers published in non-peer-reviewed journals

      1. J. C. Duchi, L. Mackey, and M. I. Jordan, "The Asymptotics of Ranking Algorithms," *ArXiv e-prints*, 2012.
      2. L. Mackey, M. I. Jordan, R. Y. Chen, B. Farrell, and J. A. Tropp, "Matrix Concentration Inequalities via the Method of Exchangeable Pairs," *ArXiv e-prints*, January 2012.
      3. B. Mu, G. Chowdhary, and J. P. How, "Efficient distributed information fusion using value of information based censoring," Aerospace control laboratory technical report, Massachusetts Institute of Technology, http://hdl.handle.net/1721.1/71875, July 2012.
      4. R. Nadakuditi and M. Newman, "Spectra of random graphs with arbitrary expected degrees," *Arxiv e-prints*, 2012.
      5. S. Olver and R. Nadakuditi, "Numerical computation of convolutions in free probability theory," *Arxiv e-prints*, 2012.
      6. T. Tsiligkaridis, A. Hero, and S. Zhou, "Convergence properties of Kronecker graphical lasso algorithms," *Arxiv e-prints*, 2012.

   c. Presentations

      i. Presentations at meetings, but not published in Conference Proceedings

         1. D. Cochran, "Quantifying the value of information sharing in sensor network synchronization," presentation in ARO Workshop on Sensor information estimation and exploitation, (Ann Arbor, MI), April 2012.
         2. E. Ertin, "Sensor Management in Adversarial Settings," presentation in ARO Workshop on Sensor information estimation and exploitation, (Ann Arbor, MI), April 2012.
         3. J. Fisher, "Performance Guarantees for Penalized Information Rewards," presentation in ARO Workshop on Sensor information estimation and exploitation, (Ann Arbor, MI), April 2012.
         4. A. Hero, "Value of information for sample-poor data collection regimes," presentation in ARO Workshop on Sensor information estimation and exploitation, (Ann Arbor, MI), April 2012.
         5. J. How, "Efficient Distributed Information Fusion using Value of Information based Censoring," presentation in ARO Workshop on Sensor information estimation and exploitation, (Ann Arbor, MI), April 2012.

6. M. Jordan, "Matrix completion and matrix concentration," presentation in ARO Workshop on Sensor information estimation and exploitation, (Ann Arbor, MI), April 2012.
7. R. Nadakuditi, "Some surprises in the approximation of low rank matrix-valued random variables," presentation in ARO Workshop on Sensor information estimation and exploitation, (Ann Arbor, MI), April 2012.
8. S. Soatto, "The geometry of the Lambert-ambient model," presentation in ARO Workshop on Sensor information estimation and exploitation, (Ann Arbor, MI), April 2012.
9. G. T. Whipps, "A Case for Being Insensitive," presentation in ARO Student Workshop, (Ann Arbor, MI), April 2012.
10. D. Wei and A. O. Hero, "Multistage adaptive estimation of sparse signals," presentation in ARO Student Workshop, (Ann Arbor, MI), April 2012.
11. T. Tsiligkaridis, A. O. Hero, S. Zhou, "On convergence of Kronecker graphical Lasso algorithms", presentation in ARO Student Workshop, (Ann Arbor, MI), April 2012.
12. T. M. Moldovan, P. Abbeel, M. I. Jordan, "Safety in Markov decision processes" presentation in ARO Student Workshop, (Ann Arbor, MI), April 2012.
13. Z. Meng, A. Wiesel, A. O. Hero, "Distributed principal component analysis in directed graphical models" presentation in ARO Student Workshop, (Ann Arbor, MI), April 2012.
14. G. Georgiadis, A. Ayvaci, S. Soatto, "Actionable saliency detection" presentation in ARO Student Workshop, (Ann Arbor, MI), April 2012.
15. N. Asendorf and R. R. Nadakuditi, "Improving and characterizing the performance of stochastic matched subspace detectors when using noisy estimated subspaces," presentation in ARO Student Workshop, (Ann Arbor, MI), April 2012.
16. E. Ertin, "Compressive Illumination Waveforms for High Resolution Radar Sensing," presentation in SIAM Conference on Imaging Science, (Philadelphia, PA), May 2012.
17. T. Tsiligkardis and A. Hero, "Kronecker Graphical Lasso," presentation in SIAM Conference on Imaging Science, (Philadelphia, PA), May 2012.
18. S Soatto, ICVSS Summer School, July 2012.
19. A. Hero, "Value-centered Information Theory for Adaptive Learning, Inference, Tracking, and Exploitation," presented at ARL, Jan. 2012.

ii. Non-Peer-Reviewed Conference Proceeding publications (other than abstracts)

iii. Peer-Reviewed Conference Proceeding publications (other than abstracts)

1. D. Cochran, S. D. Howard, B. Moran, and H. A. Schmitt, "Maximum-entropy surrogation in network signal detection," in *Proceedings of the IEEE Statistical Signal Processing Workshop*, August 2012.
2. D. Cochran, S. D. Howard, and B. Moran, "Operator-theoretic modeling and waveform design for radar in the presence of Doppler," in Proceedings of IEEE-AESS RadarCon, pp. 774-777, May 2012.

3. E. Ertin, “Sensor selection in adversarial setting,” in *IEEE Statistical Signal Processing Workshop (SSP2012)*, Ann Arbor, MI, Aug 2012.
4. G. Georgiadis, A. Ayvaci, and S. Soatto, “Actionable saliency detection,” in *Proc. of the IEEE Intl. Conf. on Comp. Vis. and Patt. Recog.*, 2012.
5. A. Kleiner, A. Talwalkar, P. Sarkar, and M. I. Jordan, “The Big Data Bootstrap,” in *Proceedings of the 29th International Conference on Machine Learning (ICML)*, Edinburgh, UK, 2012.
6. Z. Meng, A. Wiesel, and A. Hero, “Distributed principal component analysis on networks via directed graphical models,” in *IEEE Intl Conf. on Speech, Acoustics and Signal Processing (ICASSP)*, Prague, April 2012.
7. T. M. Moldovan and P. Abbeel, “Safe Exploration in Markov Decision Processes,” in *Proceedings of the 29th International Conference on Machine Learning (ICML)*, Edinburgh, UK, 2012.
8. B. Moran, S. D. Howard, and D. Cochran, “An information-geometric approach to sensor scheduling,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 5261–5264, April 2012.
9. B. Mu, G. Chowdhary, and J. P. How, “Efficient distributed information fusion using adaptive decentralized censoring algorithms,” in *American Control Conference (ACC)*, Washington DC, July 2013 (in preparation), IEEE.
10. G. Papachristoudis and J. W. Fisher III, “Theoretical guarantees on penalized information gathering,” in *Proc. IEEE Workshop on Statistical Signal Processing*, August 2012.
11. A. Ravichandran and S. Soatto, “Long-Range Spatio-Temporal Modeling with Application to Fire Detection,” *Proc. of the Eur. Conf. on Comp. Vision*, 2012.
12. T. Tsiligkaridis and A. Hero, “Sparse covariance estimation under sparse Kronecker product structure,” in *IEEE Intl Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Prague, April 2012.
13. T. Tsiligkaridis, A. Hero, and S. Zhou, “Convergence properties of Kronecker graphical lasso algorithms,”,” in *Proceedings of the IEEE Statistical Signal Processing Workshop*, Ann Arbor MI, August 2012.
14. L. Valente, R. Tsai, and S. Soatto, “Information gathering control via exploratory path panning,” *Proc. of the Conf. on Information Sciences and Systems (CISS)*, March 2012.
15. F. L. Wauthier and M. I. Jordan, “Bayesian Bias Mitigation for Crowdsourcing,” in *Advances in Neural Information Processing Systems 24*, J. Shawe-Taylor, R. S. Zemel, P. Bartlett, F. C. N. Pereira, and K. Q. Weinberger, editors, pp. 1800–1808, 2011.
16. D. Wei and A. Hero, “Multistage adaptive estimation of sparse signals,” in *Proceedings of the IEEE Statistical Signal Processing Workshop*, Ann Arbor MI, August 2012.
17. G. T. Whipps, E. Ertin, and R. L. Moses, “Threshold considerations in distributed detection in a network of sensors,” in *Proceedings of SPIE Ground/Air*

Multisensor Interoperability, Integration, and Networking for Persistent ISR III, volume 8389, Baltimore, MD, May 2012.

d. Manuscripts

e. Books

f. Honor and Awards

   1. Co-PI Hero received the 2011 Rackham Distinguished Faculty Achievement Award, Univ of Michigan.

g. Title of Patents Disclosed during the reporting period

h. Patents Awarded during the reporting period

2. Student/Supported Personnel Metrics for this Reporting Period

   a. Graduate Students

      1. Doctoral students
         - (a) ASU student Utku Ilkturk supported at 25% annualized FTE
         - (b) OSU student Nithin Sugavanatham supported at 25% annualized FTE
         - (c) OSU student Siddarth Baskar supported at 25% annualized FTE
         - (d) OSU student Gene Whipps supported at 0% annualized FTE (Whipps is a researcher at ARL on temporary assignment to OSU to complete his PhD relevant to this MURI)
         - (e) MIT student Georgious Papachristoudis supported at 50% FTE
         - (f) MIT student Beipeng Mu supported at 50% FTE
         - (g) UM student Zhaoshi Meng supported at 16.5% FTE
         - (h) UM student Hamed Firouzi supported at 16.5% FTE
         - (i) UM student Nick Azendorf supported at 16.5% FTE
         - (j) UM student Theodoros Tsiligkaridis supported at 12.5% FTE + ARL 6 weeks
         - (k) UM student Tzu-Yu Liu supported at 8% FTE
         - (l) UM student Se Un Park supported at 5% FTE
         - (m) UM student Tianpei Xie supported at ARL 12 weeks
         - (n) UCLA student Tai-Hee Lee supported at 15% FTE
         - (o) UCLA student Georgios Georgiadis supported at 15% FTE
         - (p) UCLA student Nikos Karianakis supported at 7.5% FTE
         - (q) UCLA student Jingming Dong supported at 7.5% FTE
         - (r) UCLA student Jonathon Shih supported at 7.5% FTE
      2. Masters students

   b. Post Doctorates

      1. UM postdoc Dennis Wei supported at 58% annualized FTE
      2. MIT postdoc Girish Chowdhary supported at 15% annualized FTE
      3. UCB postdoc Venkat Chandrasekaran supported at 50% annualized FTE
      4. UCLA postdoc Wang Chao-Hui supported at 7.5% annualized FTE

c. Faculty
    1. ASU faculty and co-PI Cochran supported at 12.5% annualized FTE
    2. OSU faculty and co-PI Ertin supported at 15% annualized FTE
    3. MIT faculty and co-PI Fisher supported at 20% annualized FTE
    4. MIT faculty and co-PI How supported at 10% annualized FTE
    5. UCB faculty and co-PI Jordan supported at 12.5% annualized FTE
    6. UM faculty and co-PI Nadakuditi supported at 12.5% annualized FTE
    7. UCLA faculty and co-PI Soatto supported at 2.5% annualized FTE

d. Undergraduate Students
    1. OSU student Christopher Dean (50 %)

e. Graduating Undergraduate Metrics (funded by this agreement and graduating during this reporting period):
    i. Number who graduated during this period
    ii. Number who graduated during this period with a degree in science, mathematics, engineering, or technology fields
    iii. Number who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields
    iv. Number who achieved a 3.5 GPA to 4.0 (4.0 max scale)
    v. Number funded by a DoD funded Center of Excellence grant for Education, Research and Engineering
    vi. Number who intend to work for the Department of Defense
    vii. Number who will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields

f. Masters Degrees Awarded (Name of each, Total #)

g. Ph.D.s Awarded (Name of each, Total #)
    1. UCLA doctoral student Tai-Hee Lee

h. Other Research staff (Name of each, FTE)

3. Technology transfer (any specific interactions or developments which would constitute technology transfer of the research results). Examples include patents, initiation of a start-up company based on research results, interactions with industry/Army R&D Laboratories or transfer of information which might impact the development of products.

    1. ASU doctoral student Utku Ilkturk spent six weeks at ARL during Summer 2012, supported by this award.
    2. UM doctoral student Ted Tsiligkardis spent six weeks at ARL during Summer 2012, supported by this award.
    3. UM doctoral student Tian Pei Xie spent 3 months at ARL during Summer 2012, supported by this award.
    4. co-PI Randolph Moses served on the Technical Advisory Board for the Computational & Information Sciences Directorate of the U.S. Army Research Laboratory.

5. co-PI Ertin and supported doctoral student Whipps visited ARL in March 2012, started a research collaboration with N. Srour and L. Kaplan on anomaly detection in distributed sensor networks.
6. co-PIs Cochran, Fisher and Hero visited ARL in January 2012 to initiate collaborations with ARL researchers N. Nasrabadi, T. Moore, and B. Sadler.