INTERIM PROGRESS REPORT

Period covered by report: 1 August 2013 to 31 July 2014

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
Jonathan How, MIT
Michael Jordan, UC Berkeley
Randolph L. Moses, Ohio State U
Raj Rao Nadakuditi, U of Michigan
Stefano Soatto, UCLA
Alan Willsky, MIT
Angela Yu, UCSD

ARO proposal number: W911NF-11-1-0391

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

Contents

Abstract

This interdisciplinary project is developing 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 twenty-six areas organized around three main thrusts: (1) learning and representation of high dimensional data, (2) distributed information fusion, and (3) 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 further 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 tasks 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 that accounts for human intervention.

3 Scientific barriers

This research addresses several challenges:

1. Reliable value-of-information (VoI) measures for active multi-modal sensing systems are not available. Existing approaches to learning and representation of information do 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-dimensional spaces, 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 collection and delivery of a wide range of data at different times, spatial locations, and often with severe 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 multi-modal 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 they usually require careful tuning when used as objective functions to 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 (e.g., 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 (e.g., 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 information measure as a function of the end task or the uncertainty in the environment.

4. Information collection systems very often involve human intervention at some point in the collection process. Examples are annotation through Mechanical Turk, validation of contextual data, or curation of relations that have been imputed by machine into database. A basic challenge is how to mathematically model human-human and human-machine interaction in such as way as to be predictive of the value the intervention. Mathematical modeling is challenging since it must account for fatigue, latency, and biases that a human may unwittingly contribute to the corpus. There has been very little theory developed for human-in-the-loop processing for adaptive sensing that accounts for these factors and uses human cognition models from experimental psychology. This past year we have pursued several research directions in this area, described below.

4 Significance

The significance of this research is that it addresses the longstanding problem of defining, assessing, predicting, 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 is summarized by the following five 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. The research will uncover new strategies for involving a human-in-the-loop and assessing the intrinsic value of such involvement for different sensing and situational awareness tasks.

5 Specific accomplishments over the period 8/1/13 — 7/31/14

Our efforts remain organized around the three research thrusts defined in the original proposal: (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 of presentation, in what follows we associate each reported progress and accomplishment with one of these thrusts.

5.1 Information-driven structure learning and representation

Learning and feature representation are at the front-end of the data collection system and feed the downstream functions of fusion and resource planning. By tying the learning and feature representation directly to information we can better understand information bottlenecks, limiting factors on performance, and evaluate the value of information delivered function relative to the task. This year our accomplishments are organized under three topic areas: i) learning in graphical models, ii) trade-offs between complexity and performance, and iii) representation of information for video.

5.1.1 Learning in graphical models

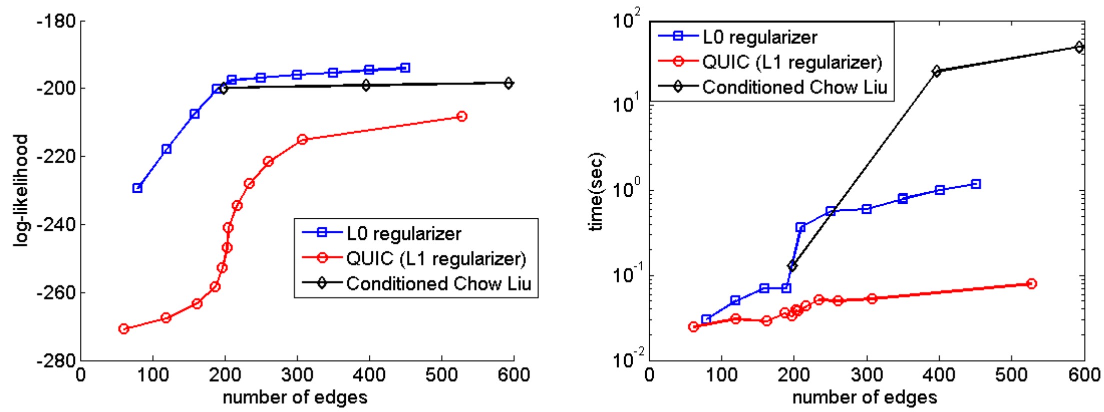
Contributors: Jon How (MIT), Alfred Hero (UM) and John Fisher (MIT)

Publications: [22], [17], [8]

Graphical models are a parsimonious class of models that capture dependency between large sets of variables. Gaussian graphical models (GGMs) are a subclass that compress information in multidimensional data through a model that assumed sparsity of the inverse covariance. When sparse inverse covariance is a valid assumption, such models are much more efficient representations of information than standard unstructured models. We have developed new efficient and fast converging algorithms for estimating GGMs under a sparse inverse assumption. For situations where the inverse is not sparse we have developed theory for a generalized GGM, called a latent variable GGM, for which the inverse covariance is not itself sparse but, when conditioned on a few latent variables, the conditional inverse covariance is sparse. We have also established the optimality of a single pass distributed estimation algorithms, introduced by us and discussed as progress last year, that does not require message passing. Finally, we have developed a tractable polynomial-complexity algorithm for Bayesian inference over latent structures. Below we report these four advances as: i) learning sparse GGMs; ii) learning latent variable GGMs; iii) distributed learning of GGMs; iv) learning latent variable structures.

Progress 1: Learning sparse Gaussian Graphical Models (How MIT):
GGM are compact probabilistic representations of the conditional dependence of Gaussian random variables. It is widely used in modeling signals, images, audios, etc. Learning a GGM has two aspects: learning the structure, i.e., how variables are connected to each other; and learning the parameters, i.e., the correlations between connected variables. Learning a GGM has traditionally been done by maximizing the l_1-regularized likelihood. l_1-regularized likelihood is easy to optimize because it is continuous and convex, but does not give sparse graph structures for high likely parameter estimates. In this work, we used a l_0-regularized ML cost function instead, which is then reduced to a mixed integer programming (MIP) problem. However, the l_0-regularized MIP problem is NP-hard, so exact solutions are intractable. To obtain a tractable solution, a greedy algorithm is used to learn edges sequentially. It is proved that when the new edge links two unconnected components of the graph, the parameter update can be achieved in constant time, thus the greedy algorithm scales well on sparse graphs where the number of cycles is small compared to the total number of edges. We also proved that when the graph is acyclic, our algorithm recovers Chow-Liu algorithm thus gives the optimal graph structure and parameter estimates. As shown in Figure 1 (a), the likelihood of the new proposed l_0 regularized method dominates that of QUIC (l_1 regularized method). It is also better than Chow-Liu in that it does not constrain the network to be acyclic. Figure 1 (b) compares the running time of the l_0 regularized method, QUIC and Chow-Liu. It shows that it does not significantly increase the running time when the graph is sparse. This work contributes to the goals of this MURI in that it has higher likelihood, thus more information with same amount of edges in GMMs than previously proposed methods. It offers an efficient way of building graphical models from complex, high-dimension data.

Progress 2: Learning Latent Variable Gaussian Graphical Models (Hero UM):
GGM have been widely used in many high-dimensional applications ranging from biological and financial data to recommender systems. Sparsity in a GGM plays a central role both statistically and computationally. Unfortunately, real-world data often does not fit well to sparse graphical models. We have developed a new family of latent variable Gaussian graphical models (LVGGM), where the model is conditionally sparse given latent variables, but marginally non-sparse. In LVGGM, the



(a) Likelihood (b) Run time

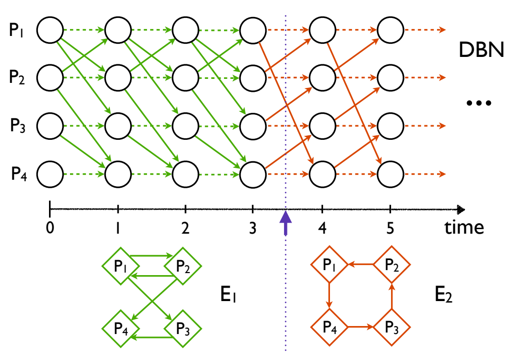
Figure 1: Performance comparison of the l_0 regularized approach, QUIC and Conditional Chow-Liu. The log likelihood of l_0 method dominates the others, and it does not significantly increase the run time when the graph is sparse.

inverse covariance matrix has a low-rank plus sparse structure, similarly to the Chandresekaran-Parillo-Sangavi-Willsky model, and can be learned in a regularized maximum likelihood framework. We have derived parameter estimation error bounds under mild conditions in the high-dimensional setting. Numerical experiments show excellent consistency between our theoretical and empirical results. Our theory can be used to quantify the value of information conveyed by knowledge that the inverse covariance is conditionally sparse. This work was published in [17].

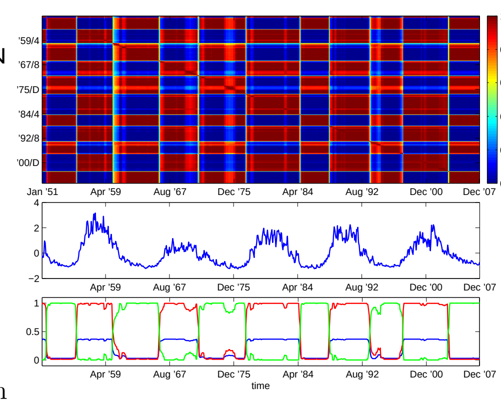
Progress 3: Marginal Likelihoods for Distributed Estimation of Graphical Model Parameters (Hero UM)

We have considered the estimation of GGM parameters when data collection and computation are distributed over multiple locations. We proposed an alternative framework for distributed parameter estimation based on maximizing marginal likelihoods. Each node independently estimates local parameters through solving a low-dimensional convex optimization with data collected from its extended local neighborhood. The local estimates are then combined into a global estimate without iterative message-passing. This year we have obtained a complete asymptotic analysis of the proposed estimator, establishing consistency and rate of convergence in Frobenius norm. Numerical experiments validate the rate of convergence and demonstrate that the decentralized estimator performs as well as the centralized maximum likelihood estimator. The significance of this result to the topic of this MURI is that, under a GGM, the value of local information is almost as high as the value of global information. This paper, which extends the work in [20] reported in the last IPR, was published in an IEEE conference proceedings [18] and will soon appear in the IEEE TRANSACTIONS ON SIGNAL PROCESSING [19]. The paper won a Best Student Paper award (2nd place) at the conference. This work has been followed up by at least two other groups in machine learning and statistics where it has been shown that this property, discovered by us in [20], actually applied to a much more general class that included non-Gaussian continuous and discrete multinomial models.

Progress 4: Learning Latent Variable Structures (Fisher MIT):
We introduce a Bayesian discrete-time framework for switching-interaction analysis under uncertainty, in which latent interactions, switching pattern and object states and dynamics are inferred from noisy (and possibly missing) observations of these objects. We propose reasoning over full posterior distribution of these latent variables as a means of combating and characterizing uncertainty. The inference procedure allows for exact probabilistic reasoning over a super-exponential number of relational graphs in polynomial time. The resulting representation is suitable for exploratory pattern discovery and post-analysis by human experts. It also encompasses straightfoward mechanisms for inclusion of complexity penalties over structures. This framework is based on a fully-Bayesian learning of the structure of a switching dynamic Bayesian network (DBN) and utilizes a state-space approach to allow for noisy observations and missing data. It generalizes two previous approaches: the autoregressive switching interaction model of Siracusa et al., which does not allow for observation noise, and the switching linear dynamic system model of Fox et al., which does not infer structural relations among objects. Posterior samples are obtained via a Gibbs sampling procedure, which is particularly efficient in the case of linear Gaussian dynamics and observation models. We demonstrate the utility of our framework on a controlled human-generated data, and climate data. One result from analysis of a well known climate data set is shown in Figure 2 where dependencies between various climate indices are discovered by the algorithm. The full algorithm and empirical results were presented at the AI Statistics Workshop in 2014 [8].



(a) Switching State-space Interaction Model (SSIM)



(b) Analysis of Climate Data

Figure 2: (a) The inference procedure is applied to dynamically varying vector-valued time-series with the goal of reasoning over the distribution of edges (e.g. E_1 versus E_2 (b) Analysis of the climate data using SSIM model. Top row is the switching-state pairwise probability matrix. Middle row is the Solar flux time series. Bottom row are the posterior probabilities of edges: Nino12 $\rightarrow$ GMT (blue), Nino12 $\rightarrow$ Nino4 (red), Nino12 $\rightarrow$ Nino34 (green).

5.1.2 Trade-offs between complexity and performance

Contributors: Michael Jordan (UCB), Emre Ertin (OSU), and John Fisher (MIT)
Publications: [40], [29], [25], [26]

One of the central foci of this MURI has been establishing and quantifying fundamental tradeoffs between performance and constraints such as bias and variance, model complexity, sample complexity, algorithmic complexity, and communication complexity. Such tradeoffs, which come in the form of phase transitions delineating the achievable region of performance, provide a theoretical underpinning of the value of information under such constraints. This year we have continued to explore tradeoffs and report on three advances: i) tradeoffs between computation and statistical estimation performance; ii) tradeoffs between memory (quantization/compression) and sequential detection performance; iii) computationally efficient evaluation information rewards for complex information plans in Markov chains, trees, and poly-trees.

Progress 5: Computation/statistics tradeoffs (Jordan UCB) We have continued our work on the interface between statistical complexity and algorithmic complexity. Our recent work has focused on the minimax prediction risk for sparse linear regression and has aimed to understand whether it is possible to obtain minimax optimality within the class of polynomial-time algorithms [40]. Under a standard assumption in complexity theory (NP not in P/poly), we have demonstrated the gap between the minimax performance that can be achieved by polynomial-time algorithms and that achieved by optimal algorithms. In particular, when the design matrix is ill-conditioned, the minimax prediction loss achievable by polynomial-time algorithms can be substantially greater than that of an optimal algorithm. This result is the first known gap between polynomial and optimal algorithms for sparse linear regression, and does not depend on conjectures in average-case complexity.

Progress 6: Quantizing information for sequential decisions (Ertin OSU)
Distributed sensors can be used for detection and monitoring for applications ranging from perimeter control, to situational awareness, to structural failure warning systems. In large scale applications, it is desirable that the sensors use algorithms and devices having low complexity. To detect low signal-to-noise ratio events sensors have to perform temporal integration of sensor readings. Sequential decision procedures rely on computing, aggregating and communicating likelihood information at high precision which might be unsuitable for low-power sensor nodes with limited computation and communication capability. In our earlier work [29], we considered design of quantized likelihood algorithms in the form of finite state machines suitable for implementation in low complexity devices. We derived necessary and sufficient conditions for optimal likelihood quantization for sequential testing. Based on these results, we introduced an iterative algorithm based on policy iteration. In our recent work we focused on theoretical analysis of the quantized decision making using nonlinear renewal theory. We show that simple finite state machine decision rule with small number of states can approximate optimal sequential test performance arbitrary closely and uniform quantization of truncated log-likelihoods is asymptotically optimal for large number of quantization levels. This work has been published in [30].

Over the next year we will consider the problem of optimal information aggregation for sequential decision that combine quantization and linear projection. For a classification problem with known class conditional densities sum of log likelihood ratios provides an optimal strategy of information accumulation to perform sequential decision tasks. Learning strategies for information accumulation from training samples is a challenging problem with few known results. Results based on classical density estimation methods performs poorly with high dimensional data. Indirect methods that rely on likelihood ratio and related divergence measure estimators such as Adaboost has been suggested in the literature with analysis limited to empirical results. We propose an alternative strategy based on design of quantizers and linear projectors to map the training samples to a low dimensional space where density estimation can be performed reliably. We plan to analyze performance of the proposed scheme analytically in the case of Gaussian signals with low rank structure and empirically using standard data sets in comparison with non-parametric methods of likelihood estimation.

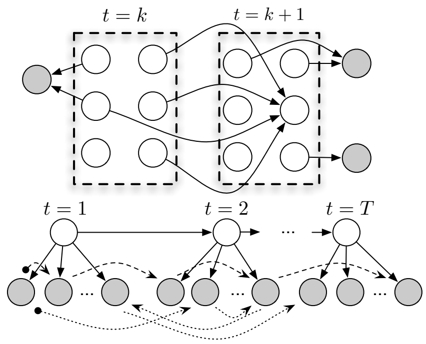
Progress 7: Efficient Information Planning in Markov Chains (Fisher MIT):
Previously, we showed that when conditional independence holds and as a consequence of the submodular property of mutual information, information planning using greedy measurement selection is guaranteed to be within a computable factor of the optimal (though intractable) selection plan. The underlying constraints are cast as selections over multiple sets of measurements. In specifying the plan, a *walk* sequence (i.e., an ordering of the sets), is specified. While the guarantees hold for all orderings (satisfying the selection constraints), the resulting information reward varies depending on the sequence order. We have shown that by taking advantage of the sparsity of the measurement process, the complexity of examining multiple sequences and computing the expected information reward can be dramatically reduced. Our method leads to substantial savings especially in large-scale models, with an abundance of measurements and a finite budget of constructing a planning strategy. We additionally demonstrate that working with the information form reduces the computational load to the absolutely necessary computations. Figure 3 shows empirical analysis of the computational complexity of different orders of selecting measurements, i.e., different *walk sequences*, and suggests how this could help in forming a measurement schedule. The underlying analysis generalizes to trees and poly-trees. This work is currently under review [25]. Additionally, the underlying analysis has been extended to an incremental belief propagation (BP) inference procedure that yields orders of magnitude reduction of computational complexity as compared to standard BP [26].

5.1.3 Representation of information for video

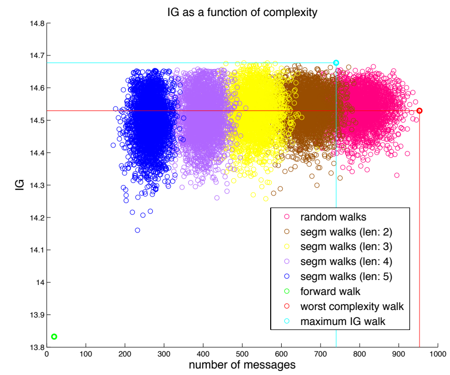
Contributors: Stefano Soatto (UCLA) and Alfred Hero (UM)
Publications: [16], [13], [12]

Videos and imagery are increasingly prevalent in defense technology and are deployed in systems ranging from autonomous navigation platforms, and wide area EO sensors, and distributed surveillance camera systems. The value of information provided by videos and imagery can be significantly higher than other sensing if the information can be easily and accurately extracted for use downstream in fusion and control, e.g., robot navigation and/or sensor management. Representation of the information in video streams is a prerequisite for design of effective information extraction algorithms. We report two areas of progress in video and image representation and learning: i) Kronecker PCA for video streams; ii) value-of-information for video segmentation.

Progress 8: Kronecker-PCA for video streams (Hero UM)
We have applied the space versus time Kronecker product decomposition developed in the first year of this MURI to represent information in video sequences. Unlike principal components analysis (PCA), the proposed Kronecker-PCA decomposition extracts lower dimensional structure that



(a) Sparse Markov Chain



(b) Information Reward vs. Computational Complexity

Figure 3: (a) Notional illustration of a Markov Chain where the graphical structure is sparse in both the measurement model and the diffusion between states. (b) The resulting algorithm allows for efficient evaluation of information rewards over multiple sequences. The plot shows that the variability of the information reward (for our example) is somewhat independent of the walk complexity. Consequently, moderate complexity walks can be analyzed to yield high information rewards.

is embedded in the spatio-temporal structure of the covariance matrix of the video. We have found that the compressibility of video data is significantly enhanced under this Kronecker-PCA representation as compared to standard algebraic PCA representations of video. This significantly reducing the number of samples required for estimation of covariance parameters and facilitates their application to video classification of objects evolving over both space and time. To allow a smooth tradeoff between the reduction in the number of parameters (to reduce estimation variance) and the accuracy of the covariance approximation (affecting estimation bias), we introduced a diagonally loaded modification of the sum of Kronecker products representation [12]. We derived a Cramér-Rao bound (CRB) on the minimum attainable mean squared predictor coefficient estimation error for unbiased estimators of Kronecker structured covariance matrices. The method was to human activity data and demonstrated advantages relative to previous approaches that do not use a Kronecker approximation of the covariance. Over the coming year we anticipate demonstrating the Kronecker-PCA framework on the software radio testbed that is being built by co-PI Ertin under a DURIP.

Progress 9: VoI for video segmentation (Soatto UCLA)

In [16], we have developed an active inference scheme to perform video segmentation, according to a VoI criterion. Video segmentation consists in the labeling of every pixel of every frame in a video with one of a number of class labels, under the assumption that one has at his disposal a battery of detectors that can, for each class, return a probability that such a class is present in the portion of the scene that projects at that pixel.

While one could clearly run a detector for every class (tens, in our case) at every pixel (millions) of every frame (hundreds), this leads to a computational complexity that is unmanageable in any practical application. Instead, we have developed a technique to select *which frames* (time), *which* *pixels* (space), and *which labels* (class) to test, so as to minimize the information loss (uncertainty reduction) compared to the "paragon" scenario of running detectors for each class, at each spatial location at each instant of time.

While in principle one expects (smallest) performance reduction by such (time, space, class) subset selection, we have found that our scheme actually *improves* performance compared to the paragon scenario. This is due to the fact that we exploit temporal, spatial, and class consistency. We have shown that such context enables running a small subset of detectors, at a small subset of locations and frames, so with a computational cost of 10% of the paragon, we can actually obtain a performance improvement, as measured by standard classification criteria.

Our approach has been used both in the presence of an "oracle" (a human annotator), to propagate labels performed in an image to temporally adjacent ones, as well as to select the "best" (in the sense of VoI) frame to submit to the human annotator, and also to perform fully automated video segmentation in the presence of noisy detectors. The results have been presented at the latest CVPR [16].

One of the challenges in extending these techniques to unstructured video (for instance, that captured by a drone, rather than a purposeful video captured by a human and posted on the web), is the significant nuisance variability due to scaling and visibility, including limitations to visibility due to finite field of view, and occlusions.

In [13] we have tackled scale, by proposing a correspondence method for shapes that factors out nuisance variability due to scale. This is done by generating an equivalence class of descriptors, whose orbits characterize a (planar) shape up to an arbitrary change of scale. This work concludes a long line of work, commenced in prior phases of this project, on multi-scale integral invariant signatures, where invariance to geometric transformations is achieved *not* by differential quantities (such as various notions of curvature), but by integral quantities that are still local (hence robust to nuisance variability due to occlusions) but not as sensitive to fine-scale structure due to sensor noise.

5.2 Distributed information fusion

Accurate aggregation of information at multiple sensors is a key part of the value of information proposition we are studying. The information at a single sensor may have little or no value until matched with information from another sensor, e.g., when the objective is to extract correlation from the sensors for the purposes of target localization or clutter abatement. Subspace processing and dimension reduction are widely used methods for information aggregation and our MURI is working on minimizing any associated loss of information due to decentralized processing, mismodeling error, bandwidth-limited inter-sensor communications, and other factors. We report progress in distributed fusion along three axes: (i) Decentralized learning and local information aggregation; (ii) Subspace processing and fusion of information; and (iii) robust information-driven fusion.

5.2.1 Decentralized learning and local information aggregation

Contributors: Michael Jordan (OSU), Emre Ertin (OSU) and Randy Moses (OSU)
Publications: [7], [36], [35], [34]

In a large networks of sensors centralized learning and fusion of information is impractical due to limited bandwidth interconnectivity between sensors and a fusion center that prevents global information aggregation. An alternative is decentralized learning where sensors extract features or estimates and share this information with their neighbors. Several advances have been made this year on decentralized learning and information fusion: i) lower bounds on decentralized minimax risk that specify the minimum amount of information sharing required to achieve the centralized minimax risk; ii) decentralized learning using a mixture of factor analyzers; iii) decentralized decision-making in the presence of communication errors due to random access packet collisions.

Progress 10: Lower bounds for the statistical performance of distributed estimation methods (Jordan UCB) In this line of work, we have continued to study the statistical performance of distributed estimation algorithms [7]. We have defined and studied some refinements of the classical minimax risk that apply to distributed settings, comparing to the performance of estimators with access to the entire data. Lower bounds on these quantities provide a precise characterization of the minimum amount of communication required to achieve the centralized minimax risk. We have studied two classes of distributed protocols: one in which machines send messages independently over channels without feedback, and a second allowing for interactive communication, in which a central server broadcasts the messages from a given machine to all other machines. We have established lower bounds for a variety of problems, including location estimation in several families and parameter estimation in different types of regression models. Our results include a novel class of quantitative data-processing inequalities used to characterize the effects of limited communication.

Over the course of the next year we will look at submodularity for decentralized distributed estimation. Many VOI functionals are based on submodular functions or on Lovasz extensions of submodular functions. The minimization of such functionals can exploit the submodular structure and, moreover, can exploit the additive structure that arises in distributed setting. We wish to develop a general theory of convergence for the minimization of additive submodular functions,establishing matching upper and lower bounds for distributed algorithms.

Progress 11: Decentralized learning of a mixture of factor analyzers (Moses, Ertin OSU)
In our recent work [36], we developed a decentralized manifold learning method with a potentially reduced data bandwidth need, and which results in a global appearance manifold model shared by all sensor nodes. A spatially distributed sensor network can be used to construct a rich appearance model for targets in their common field-of-view. These models can then be used to identify previously seen objects if they reappear in the network at a later time. As an example, consider a network of cameras capturing images of an object from different but possibly overlapping aspects as the object traverses through the network's field of view. The ensemble of images captured by the network forms a low-dimensional nonlinear manifold in the high-dimensional ambient space of images. One approach to appearance modeling would be to construct independent models of a

local data manifold at each sensor and share it across the network. However, such an ensemble of models suffers from discretization of the aspect space and poor parameter estimates as the number of unknown parameters necessarily scale linearly with the number of sensor nodes. Alternatively, the sensor nodes can collaborate to construct a joint model for the image ensemble. The parameter estimates of the joint model will improve with the number of sensor nodes, since the number of unknown parameters in the model is intrinsic to the object and fixed, whereas the measurements scale linearly with the number of sensor nodes. The straightforward method of pooling images to a central location for joint model construction will require large and likely impractical network bandwidth.

We model the overall statistics as a mixture of factor analyzers (MFA) and derive a consensus-based decentralized expectation maximization (EM) algorithm for learning model parameters. Gossip-based methods are appropriate where sensor nodes that share a common view of a target are not necessarily network neighbors. The MFA model is a probabilistic and generative one, and can be used for dimensionality reduction, manifold learning, and signal recovery from compressed sensing. We consider a more general MFA model suitable for modeling data observed by heterogeneous sensor nodes differing in their aspect angle with respect to the object. Specifically, we assume observations are drawn from the mixture density with mixture probabilities which can vary across the different sensor nodes. In the case of learning a data manifold, the MFA model is a linearization of a (potentially) nonlinear structure. We extend the EM algorithm for the MFA model to the case of a spatially-distributed sensor network with goals of distributing computations across the network and being robust to individual node failures (e.g., losing connectivity to a central node in centralized or distributed systems). Our work incorporates a low-dimensional structure which is key to accurately modeling high-dimensional data observed by a network of sensors and whose relevant characteristics lie on a common low-dimensional manifold structure.

Progress 12: Aggregating local information under communication constraints 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. In our earlier work [35] we analyzed how the performance of a large scale sensor network scales with the density of sensor nodes in a random sensor network, while considering a random access channel model with collisions resulting in packet losses. We analyzed numerically the detection performance as a function of sensor density subject to a constant network bandwidth constraint. Under this model we showed that detection performance improves with increasing sensor density while satisfying constant network bandwidth by desensitizing sensors (i.e., increasing local decision thresholds) and satisfying a global false alarm probability.

In our recent work [34] we extended our previous work to a random target signal model and derive Neyman-Pearson-optimal decision rules. We provided conditions where local and global decision rules do not need to know the target signal distributions. We analyzed analytically and numerically how the performance scales with density of the sensor network and the number of communications slots in the random access model. We showed that detection performance improves with increasing sensor density, despite an increase in the probability of a collision per communications slot, while satisfying a constant network bandwidth and satisfying a global false alarm probability. Further-more, we showed that the detection performance under the random access channel asymptotes to a perfect channel model as the number of communications slots increases. Lastly, we provided a bound on the confidence interval of the receiver operating characteristic (ROC) curve to account for variability in performance across realizations of the random sensor network and target signal.

5.2.2 Subspace processing and fusion of information

Contributors: Michael Jordan (UCB), Raj Nadakuditi (UM), and Emre Ertin (OSU)
Publications: [39], [23], [21], [3], [11], [9], [11]

Dimension reduction is at the heart of the information fusion function of data collection systems as it extracts the space containing common information residing in different components of the data. Dimension reduction should depend on the definition of the task, e.g., classification, parameter estimation, or tracking, which determines the value of the information contained in the subspace. Overestimation of the dimension of this subspace leads to high sensitivity to noise while underestimation of the subspace dimension leads to bias due to omission of important information carrying components. Spectral methods have been used in machine learning and signal processing to accurately determine the correct subspace dimension and perform dimension reduction. Three areas of progress are reported this year: i) spectral methods for designing classifiers with noisy labels provided through crowdsourcing; ii) spectral methods for determining dimension in subspace processing applied to imaging under noisy conditions; iii) spectral manifold learning methods for compensating clutter perturbations wide area SAR sensors.

Progress 13: Optimal convergence rates via combination of spectral methods and EM (Jordan UCB) Spectral methods have become a popular alternative to maximum likelihood methods in recent years. Whereas maximum likelihood methods often have an unknown runtime, it is often possible to guarantee polynomial-time complexity for spectral methods by relying on the singular value decomposition. Statistically, however, spectral methods are method-of-moments estimators and their statistical efficiency may be poorer than that of maximum likelihood. We have studied these issues in the setting of crowdsourcing, where the goal is to effectively collect labels at low cost [39]. The Dawid-Skene estimator has been widely used for inferring the true labels from the noisy labels provided by non-expert crowdsourcing workers. However, since the estimator maximizes a non-convex log-likelihood function, it is hard to justify its performance theoretically. We have proposed a two-stage efficient algorithm for multi-class crowd labeling problems. The first stage uses the spectral method to obtain an initial estimate of parameters. The second stage then refines the estimation by optimizing the objective function of the Dawid-Skene estimator via the EM algorithm. We show that this algorithm achieves the optimal convergence rate up to a logarithmic factor.

Next year we will look at deriving optimal rates of convergence as a benchmark for this problem. We have shown that it is possible to combine spectral methods and EM and obtain optimal convergence rates in a particular crowdsourcing application. It is of major interest to explore whether a properly initialized one-step EM algorithm can achieve the optimal rate for other latent variable models such as latent Dirichlet allocation or other mixed membership models.

Progress 14: Spectral measures and subspace detection from random matrices (Nadakuditi UM)

In the area of non-commutative information theory, we are seeking to establish 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 they specify phase transition thresholds of SNR and matrix dimension for which these eigen-quantities cannot be reliably estimated empirically. Such phase transition thresholds are key for developing the non-commutative information theory of dimensionality reduction, which is relevant, for example, to variable selection in sensor fusion.

Last year we made substantial progress in developing data-driven algorithms for low-rank signal matrix denoising using non-commutative (or free) probability theory. Specifically, we developed an algorithm [23] for denoising a low-rank signal matrix buried in noise by optimal singular value shrinkage. The algorithm[1] explicitly utilizes information in the “noise portion” of the singular value spectrum to compute these shrinkage coefficients and returns an estimate of the approximation MSE that is provably consistent and that can serve as a new VoI metric.

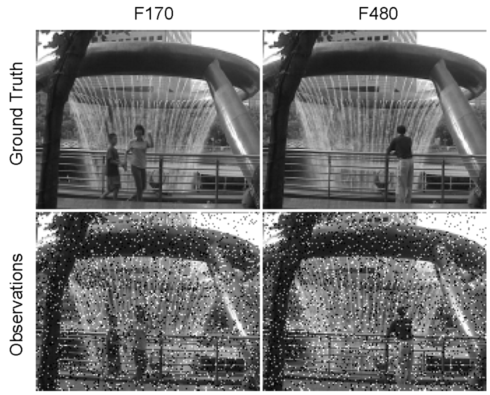
This year, we looked at computer vision and image processing tasks where these ideas could be applied. We looked at the problem of low-rank-plus-sparse decomposition or robust PCA [21]. We developed a method that utilizes OptShrink to estimate and denoise low-rank part with with an ℓ_1 estimator for estimating and denoising the sparse part. We theoretically and empirically showed that the method yields superior background subtraction relative to the state-of-the-art approach that utilizes a nuclear norm induced singular value thresholding operator for the low-rank denoising part. See Figure 4 for an illustration of the algorithm’s superior performance. We have initiated and have made progress in collaborations with co-PIs Cochran and Hero on extensions of this work to the sensing and detection of correlated signals in multi-modal signal processing problems that are corrupted with outliers. We are actively collaborating with MURI co-PIs to investigate application of the robust PCA algorithm. In the upcoming year, we plan to test our algorithms on additional real-world datasets. This work

Progress 15: Performance bounds for inference using high dimensional data (Ertin OSU)

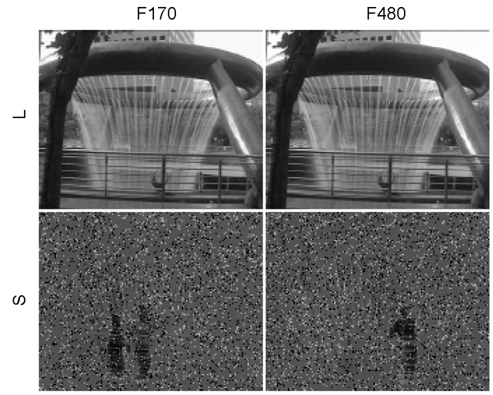
Many sensor systems such as camera network or EO/RF sensors mounted on airborne platforms are able to interrogate a scene persistently over a large range of aspect angles. For many modalities the target signature varies significantly with target pose [3, 11]. Learning and exploiting the additional information provided by wide-aspect target signatures is key to developing successful automatic target recognition algorithms and characterizing their performance. Characterizing achievable performance in target classification and pose estimation tasks is complex due to the unknown noise statistics in the high dimensional sensor data space. While performing feature extraction step provides certain noise immunity, background returns leads to inclusion of non-target features and occlusion of some of the target features. The resulting perturbations in the high dimensional data are extremely hard to parameterize with non-uniform spatial correlations.

In our early work [9] we employed manifold learning based dimensionality reduction for modeling

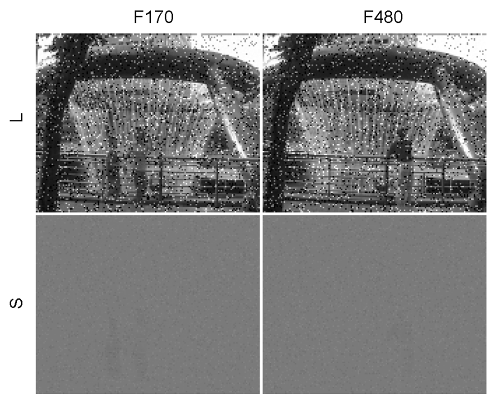
[1] Software available at `www.eecs.umich.edu/~rajnrao/optshrink`



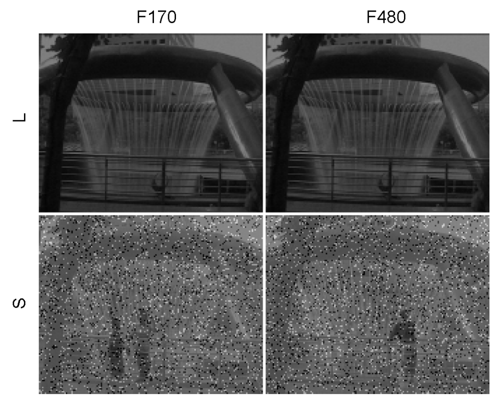
(a) Top: ground truth data, Y. Bottom: observed data, $\tilde{X}$.



(b) OptShink based approach with $r = 1$ and $\lambda_S = 0.0035$. NRMSE = 9.5%.



(c) SVT-based updates with $\lambda_L = 6.5$ and $\lambda_S = 0.0035$. NRMSE = 38.4%.



(d) SVT-based updates with $\lambda_L = 300$ and $\lambda_S = 0.0035$. NRMSE = 42.3%.

Figure 4: Simulation results for the fountain dataset with parameters $\delta = 0.0025$, $\tau_k = 0.5$, $p = 0.15$, and $K = 0.5$. The row labels L and S denote the low-rank and sparse components, respectively, returned by each algorithm. The column labels denote the frame number (i.e., column of L and S) that is displayed. Each subfigure has the same intensity scale. NRMSE values are reported for the low-rank components using output of the SVT-based updates with $p = 0$ as ground truth.

and learning the clutter perturbations in the low dimensional embedding space. In the embedding space, correlated perturbations are projected to small number of dimensions multivariate normal distributions, which are well approximated with Gaussian random vectors. For the signal model, the geometry of the low-dimensional embedding is learned to compute gradients and differential area elements. To validate this promising direction we performed simulation experiments with data from wide-angle SAR sensors. Next Cramér-Rao Bound (CRB) analysis is employed to quantify information provided by the wide-aspect target signatures. Our performance bounds learned from the data are in close agreement with state-of-the art ATR algorithms developed in independent work. Recently [10] we studied random projection matrices for estimating Fisher Information, by characterizing information loss due to the random projection. We show that average taken over random projection matrices lead to an information loss with a constant factor. As a result the bias due to the random projection can be corrected to provide an estimator for Fisher Information and a method to compute achievable performance in pose estimation tasks. This work quantifies the loss in value of information due to subspace projection and uses Fisher information as a proxy for task-related performance.

5.2.3 Robust information-driven fusion

Contributors: Alfred Hero (UM), Emre Ertin (OSU), and John Fisher (MIT)
Publications: [37],

An information fusion criterion that lacks robustness to model mismatch may perform poorly when deploying sensing algorithms in uncertain environments. More importantly, in terms of MURI goals, if not accounted for, model mismatch will cause the computed value-of-information to be inaccurate and possibly lead to violation of the performance guarantees and error control levels that have been designed into the system. This year we report progress on three fronts: i) fusion with unreliable information; ii) robust sensing matrix design for signal-dependent clutter; iii) fusion of multi-modal data sources with imperfect sensor models. These have been applied to fusion of multi-modality data in the presence of possible sensor failures and remote sensing when the statistical distribution of the radar clutter return lies in an convex uncertainty set.

Progress 16: Fusion with unreliable information (Hero UM)
We have been addressing the problem of sensor fusion in the presence of unreliable information sources. This problem is of particular importance in multimodality sensing systems when there may be sensor failures that affect the reliability of a classifier working on the aggregated data. We have formulated this in the Bayesian framework of maximum entropy discrimination (Jaakola, 2001) and have developed a very flexible approach to fusion with unreliable information. In this approach one attempts to recover the maximum value of information (relative entropy) from the sensor data subject to a constraint on task-specific performance of a failure-resistant classifier that uses this information to classify. The unreliability of each sensor is incorporated into the model as an unknown latent variable (reliable, not-reliable) that must be estimated along with the class label. The fusion strategy was applied to intruder classification in the "footstep" dataset collected by ARL. This data includes acoustic, infrared, and seismic sensors and exhibits a significant number of sensor failures. The optimized relative entropy function is used as a measure of the value of information from an intermittent sensor system. A paper on our work appeared at an IEEE conference [37]. This work has been performed in collaboration with Nasser Nasrabadi at ARL.

Progress 17: Robust sensing matrix design for Gaussian mixture model signals in signal dependent clutter (Ertin OSU)
We have developed an alternative sensing matrix design strategy based on the Gaussian mixture signals in colored Gaussian clutter. This model is commonly adapted in remote sensing applications, since clutter returns –from non-target objects and background– depend on the waveform used in interrogation. The sensing matrix design problem has been considered earlier in a Bayesian setting. We consider optimal matrix designs with robust mean square error performance when the mixture coefficients are unknown. The performance of a sensing matrix is given by the value of the min-max game between the estimator designer and nature choosing the mixture coefficients form a fixed class of estimators. Examples of estimator class include Linear Estimators and Empirical

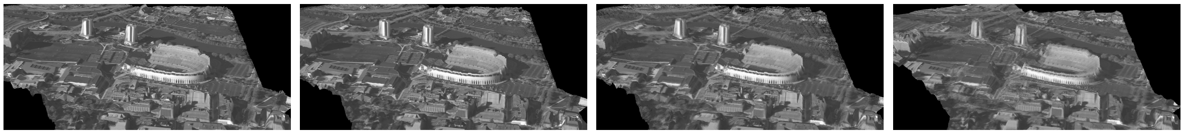


Figure 5: Reconstruction as a function of number of images used. From left to right: 49, 24, 10 and 5 images (fixed world geometry and LiDAR, same number of optimizations run for all). Combining LiDAR allows one to use significantly fewer images for comparable reconstruction

Bayes Estimators. For each estimator class we show that the estimator and the mixture coefficients form a regular pair, with regularity conditions established in the robust filter design literature. Therefore the least favorable prior and the robust estimator form a a saddle point and the order of optimization can be exchanged. Using the envelope theorem we obtain a closed form gradient of the value function with respect to the sensing matrix, which can be used in a gradient descent optimization algorithm. We note that even for the class of linear estimators the value function is non-convex and only local optimality can be achieved. Next, we show that our results apply to the case of gaussian signals in clutter with mixture density. We illustrate the results on the canonical problem of rank-1 target classification in clutter. This work was published in [].

Progress 18: Multi-modal Fusion of LiDAR and WAMI Data (Fisher MIT):
We developed an integrated probabilistic model for multi-modal fusion of aerial imagery, LiDAR data, and (optional) GPS measurements using approximate sensor models. The model allows for analysis and dense reconstruction (in terms of both geometry and appearance) of large 3D scenes. An advantage of the approach is that it explicitly models uncertainty and allows for missing data. Empirical results demonstrate that inclusion of LiDAR data renders the information content of multiple images redundant, consequently, as compared with image-only based methods, dense reconstructions of complex urban scenes are feasible with significantly fewer observations. Moreover, the proposed model allows one to estimate absolute scale and orientation, and reason about other aspects of the scene, e.g., detection of moving objects. As formulated, the model lends itself to massively-parallel computing. We exploit this in an efficient inference scheme that utilizes both general purpose and domain-specific hardware components. We demonstrate results on large-scale reconstruction of urban terrain from LiDAR and aerial photography data. Figure 5 shows reconstruction results using WAMI data collected over a large urban city. This work was presented in the recent CVPR conference [4].

5.3 Active information exploitation for resource management

The active information exploitation thrust completes the feedback loop from acquisition, learning and fusion to control of sensing resources. In active information exploitation one takes a sensing action based on prior measurements and sensing actions. This active feedback of information to control sensing actions is one of the aspects of our project that differentiates it much of the prior work on quality of information. A key component to making effective use of feedback is the specification of suitable proxies for the value of information delivered by each potential sensing action. Another component, which we have made progress on, is the possible role of humans in this feedback loop. Another area of progress is laying the foundations for an information geometric theory of actively controlled sensing systems. These components of progress are described below.

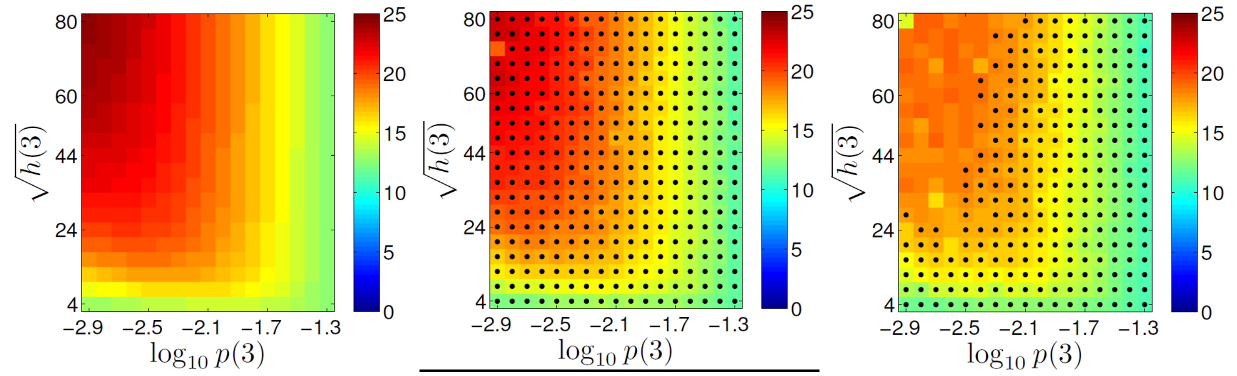
5.3.1 VoI proxies for mission dependent resource management

Contributors: Alfred Hero UM, Jon How MIT, Emre Ertin OSU, Doug Cochran ASU
Publications: [24], [33]. [27], [14], [6]

Ideally the value of information acquired from an action policy would be predicted from the available data from which a best policy could be determined. This ideal problem is intractable in general since the VoI would need to be predicted over all possible action sequences. An alternative is to define simpler proxies for VoI that lead to tractable sub-optimal policies whose performance can be analysed mathematically. We have made progress in the area of defining and analyzing good proxies: 1) proxies for multi-class target search; 2) asymptotic performance limits of single class target search proxies; and 3) proxies for value of information when there is an adversary who is trying to minimize the amount information acquired.

Progress 19: Mission-weighted adaptive search for multi-class targets (Hero UM, How MIT)
Our research has shown that significant performance improvements can be obtained in sparse target detection problems by adaptively allocating sensing resources. Especially, targets have various mission values in many cases. For example, tanks are more important than cars in surveillance problems. Therefore, treating targets equally as has been done by previous work may lead to inefficient use of resources, particularly when they are used for collecting information from low-value targets. This year, we generalized previous work on adaptive sensing to (a) include multiple classes of targets with different mission value and (b) account for heterogeneous sensor models. Upper and lower bounds on performance are provided by defining an oracle policy which knows the locations and classes of targets, (i.e., the upper bound), and the global uniform (GU) policy which allocates resource uniformly across the whole scene (i.e., the lower bound). Figure 6(a) shows the performance gain of the oracle over the GU policy. It can be seen that significant gains are possible when high-importance targets are sparse (low $p(3)$) and have high mission value (high $h(3)$). We also developed new optimization policies that allocate a limited resource budget to simultaneously locate, classify and estimate a sparse number of targets embedded in a large space. Moreover, we explicitly considered three heterogeneous sensors models: global adaptive (GA), which can allocate arbitrary resource over subsets of the whole space; global uniform (GU), which can allocate resources uniformly across the scene; and local adaptive (LA), which can allocate fixed amount resource to one location. Figure 6 shows that the policy that uses a single GA policy closely mirrors the performance of the oracle (<3dB difference is indicated by black dots). Global adaptive sensors have the most agility but can be expensive in reality. The performance of a policy that uses a mixture of cheaper GU and LA sensors is shown in Figure 6(c) from [24]. The plot shows that the policy performs close to the oracle, except for a few cases with high mission value $h(3)$ but low probability $p(3)$.

Next year we plan on looking at human-in-the-loop generalizations of this strategy. Machines are good at processing large volumes of data and organizing them in a pre-defined structure, humans can provide special insights in the information contained in data and help improve the



(a) Gain oracle vs. GU (b) Gain GA vs. GU (c) Gain GULA vs. GU

Figure 6: Performance gain of the full-oracle, GA and GU/LA policies with respect to the GU policy at 20 dB SNR. Gains are given by varying class value weight $h(3)$ and prior probability $p(3)$. Black dots indicate that the policy is within 3 dB of the oracle policy. The GA policy closely mirrors the oracle in almost all cases. The GU/LA policy also performs closely with the oracle except for a few cases with high $h(3)$ and low $p(3)$.

performance of machine learning algorithms. For example, as has been shown in the active learning literature, having humans add only a few informative labels can significantly speed up the binary classification process. However, humans typically have limited capacity to process data as compared to machines, therefore it is important to design mechanisms by which the learning processes can generate particularly informative questions to ask the human collaborators. This is an interesting VoI question that has elements of planning, operator modeling, and information quantification. Our work in this area next year will include humans in the learning of mixture models. More specifically, we will study how to evaluate the informativeness of labels, thus machines can ask questions only on data points that are informative. And how human behaviors, such as errors, will affect this interaction between humans and machines.

Progress 20: Performance limits for convex proxies in VoI search (Hero UM)
We continue to make progress on VoI-driven strategies for the wide area search problem. As explained in the previous report, this problem arises in wide-area search and tracking, sensor selection, waveform selection, and other relevant scheduling problems. Fundamental theory for the target localization performance under a convex proxy for VoI has been developed. Unlike other proxies for VoI, our convex proxy is sufficiently simple to analyze to obtain closed form analytical characterizations of the optimal policy, the associated VoI, and the associated exploration vs exploitation tradeoff. Our results were published in conference paper [33] and we have submitted an extended version to a journal. This work will impact collaborative work with co-PI How on designing convex VoI proxies for wide search applications where target classification is the primary goal and correct identification of different targets may have different value payoffs relative to the mission (5.3.1).

Progress 21: Sensor management using Adversarial information structures (Ertin OSU)
In this effort we are interested in developing information-theoretic uncertainty measures to provide

a foundation for quantifying value of information in adversarial situations. In turn this measures are used to devise information driven sensor management/placement algorithms applicable to adversarial scenarios. We study asymptotic error exponents in a surveillance game between an observer with a sensor system and an adversary trying to avoid detection. In this setting the fusion center can access only a subset of sensors at any time in order to satisfy bandwidth or energy constraints while maximizing a performance metric for an inference task such as tracking or detection. In our earlier work we derived optimal strategies for the target and the observer and found the Nash Equilibrium of the surveillance game and characterized the value function as the VOI measure applicable to this adversarial setting.

In [27] we modeled the sensor management problem as a two stage game between the observer and the target: in the first stage sensor locations are chosen by the observer, in the second stage the observer and the target play the surveillance game that we studied in our earlier work. The value function of the surveillance game is parameterized as a function of the sensor configuration. Next, the problem of finding an optimal configuration for the sensors is formulated as an optimization problem of the value function achieved at the saddle point of the surveillance game. We provide an iterative algorithm for computing the value function of the second stage game and a computationally efficient gradient based method for sensor placement optimization in the first stage game. We show that sensor placement problem is a non-convex problem with only guaranteed convergence to local maxima. We provide heuristic reinitialization methods based on VoI assigned to subsets of sensors for efficient search of the sensor configuration space. Our empirical results show that for the class of problems where optimal sensor configurations can be computed in closed-form, the proposed iterative algorithm with informed reinitialization step achieves the global maximum.

Progress 22: Value of information sharing in networked systems (Cochran ASU)
This vein of our work is seeking to quantify the value of sharing information in a class of detection and estimation problems involving multiple networked sensors. We have examined the relative performance of such systems when data shared on links between sensor nodes in the network graph is replaced by proxy data obtained by an entropy maximization procedure constrained by the actual data on other links. This work is providing insight into how the passing of information in a sensor network can be prioritized when communication resources are constrained. In contrast to existing work in the context of communication networks, value of communication between cooperating nodes is measured here with respect to sensing objectives (detection or estimation performance) rather than measures of data throughput [14, 6].

5.3.2 Human-in-the-loop distributed search

Contributors: Angela Yu UCSD, Alfred Hero UM
Publications: [38], [1], [2], [31]

A human observer can provide essential contextual information to help automated sensing algorithms to perform estimation, tracking, classification and situational awareneness, among other tasks. Good computational and mathematical models for human-interaction systems are not widely available, especially in the context of collaborative estimation and competitive foraging, areas that we have addressed with new theory, simulation, and human experiments. We continue to develop mathematical models for human-human and human-machine interaction that are relevant to the human's added-value to the value of information. In particular, we have extended to a decentralized framework our prior work on centralized "twenty questions" where a controller asks questions of agents (humans and/or machines) to collaboratively localize a target in a noisy image. We have also made significant progress in computational modeling of human cognition in cooperative search problems - establishing that human decisionmaking is well modeled by Markov process with 'leaky" memory. These two areas are described below.

Progress 23: Computational models of human cognition in cooperative search problems (Yu UCSD)
This new activity started under this MURI in summer 2013 when Angela Yu joined the project team. Her research group focuses on understanding the computational processes underlying human cognition, in particular how the brain represents and seeks out information from the environment as it tries to achieve behavioral goals in varying contexts. She takes a multi-pronged approach of understanding human behavior by incorporating state-of-the-art techniques in Bayesian inference and Markov decision processes as modeling tools, developing tractable approximations motivated by human behavioral data, designing and executing human experiments that serve as testbed for model assumptions and scientific hypotheses, and predicting behavior in novel task settings. Specific applications of her work within the context of the VOI MURI includes developing reduced models of human behavior, inferring "true" human confidence in judgment and decision-making, optimizing situation-dependent choice of human experts, and designing individualized training to optimize human performance.

Over the past year, Co-PI Yu's group has made significant progress on three areas of human information foraging and decision-making. The first area is human active learning behavior in a multi-arm bandit task [38]: it was found that humans have systematic "forget" past experiences in a manner consistent with believing that the world is changeable in a Markov fashion, and that they trade off exploration and exploitation smartly but inexpensively by assuming a partly myopic exploration strategy (known as knowledge gradient). The second area is human active sensing in a target search task, in which humans are again found to have a "leaky" memory consistent with assuming the environmental statistics to undergo Markovian changes, and a decision-strategy that is near-optimal in taking into account external and internal costs such as the cost of sampling (or time), the cost of sensor repositioning, and the cost of inaccurate search outcomes [1]. The third area is human competitive foraging behavior. So far, we have developed a myopic Markov decision process model to account for individual decision-making and group dynamics in a competitive foraging context [2]. We have explored how individual behavior and group dynamics would change as a function of the level of communication among individuals, the amount of knowledge individuals have about the environmental reward distribution, and the absence/presence of dynamical changes in environmental statistics.

Over the next year, we plan to deepen and unify these three areas of research. We plan to incorporate the modeling framework and experimental findings from the active sensing project to expand the study of human trade-off of exploration and exploitation in the multi-arm bandit task. In particular, we plan to introduce a notion for "travel cost" that is equivalent to the "sensor repositioning" cost in the active sensing problem, as well as the possibility of continuously valued hypothesis space (as opposed to discrete options) and the possibility of non-stationarity in environmental statistics. This will also allow us to move toward a more realistic foraging framework, and thus evolving the framework to better account for competition in a multi-agent context. We also plan to experimentally test some of the assumptions and predictions we have developed so far in the modeling work for competitive foraging. In general, we will be developing both the theoretical modeling aspect as well as the experimentation aspect over the next year, where we expect the modeling results will help constrain experimental design, and the experimental findings will help refine the next iteration of models.

Progress 24: Decentralized cooperative human-machine tracking (Hero UM)
Last year a new effort was started that explores the potential improvement in VoI when a human is included in the processing loop. We adopted a 20-questions framework where a human and a computer cooperate to locate the position of a target based on sensing inputs; e.g., noisy imagery or ranging data. This work appeared early this year as [32]. This year we extended this framework to decentralized 20 questions for multiple players/agents in a network. This extension combines the model from our previous 20 questions work in addition to a social learning model for information sharing. The proposed framework provides a flexible and tractable mathematical model for information gathering in active decentralized parameter estimation systems. A paper on this was submitted to a journal [31]. Over the coming year we plan on refining this decentralized model in two important directions: 1) extension of the target state estimation problem to classification and 2) adoption of a more realistic model for human cognition. This work is in collaboration with Brian Sadler at ARL.

5.3.3 Information geometric foundations

Contributors: Douglas Cochran (ASU) and Alfred Hero (UM)
Publications: [5],[15],[28]

The geometric viewpoint has been very useful in simplifying and unifying many problems in signal processing. For example, in linear prediction and estimation problems the normal equations provide a intuitive interpretation of the minimum distance properties of Wiener and Kalman filters in a Hilbert space. However, this viewpoint only applied to linear operations on the data and does not account for the full data probability distribution. An alternative viewpoint, that we have been developing in this MURI project, uses the manifold of probability distributions to describe the effect of control actions. In this geometry minimum distances between distributions can be quantified through the KL distance and the associated Fisher-Rao Riemannian metric even though the manifold is non-linear. This information geometric viewpoint leads to interesting characterizations of value of information relative to non-linear tasks such as detection, classification, and prediction with active sensor control. Two areas of progress are reported: 1) information-driven sensor planning and navigation a statistical manifold; 2) intrinsic Fisher information.

Progress 25: Information-driven sensor planning: navigating a statistical manifold (Cochran ASU and Hero UM)
Many adaptive sensing and sensor management strategies seek to determine a sequence of sensor

actions that successively optimizes an objective function. Frequently the goal is to adjust a sensor to best estimate a partially observed state variable, for example, the objective function may be the final mean-squared state estimation error. Information-driven sensor planning strategies adopt an objective function that measures the accumulation of information as defined by a suitable metric, such as Fisher information, Bhattacharyya affinity, or Kullback-Leibler divergence. These information measures are defined on the space of probability distributions of data acquired by the sensor, and there is a distribution in this space corresponding to each sensor configuration. Hence, sensor planning can be posed as a problem of optimally navigating over a statistical manifold of probability distributions. This information-geometric perspective presents new insights into adaptive sensing and sensor management. This work was published in [5]. Future work will be to explore whether the equivalence between estimation and detection, proved by information geometrical arguments in [5], extends to other probability laws different from the multinomial distribution explored last year.

Progress 26: Intrinsic Fisher Information (Cochran ASU)
Work reported in the past two years has examined the possibility of using information-geometric characterization of VoI as a basis for control of sensing and other information collection processes. This perspective draws upon classical estimation theory for situations in which a collection of conditional probability density functions $p(x|\theta)$ for a random variable X is parameterized by $\theta \in M$ with M a smooth manifold. In this setting, the Fisher information for the problem of estimating θ from collected data x induces a Riemannian metric on M, and this metric quantifies the value of the collected data x in a way that allows optimal (geodesic) trajectories for data collection to be identified. This vein of work was extended during the past year to define and incorporate *intrinsic* Fisher information on a manifold, which does not depend on an underlying estimation problem. The De Bruijn identity relates the intrinsic Fisher information to the Kullback-Leibler divergence, which is the rate exponent governing speed of convergence of the type I and type II errors of an optimal Neyman-Pearson binary hypothesis test. The advantage of the intrinsic Fisher information is that it is invariant to the choice of connection on M and is in fact the Riemannian metric on this space.

Building on this perspective, our work this year has generalized intrinsic Fisher information and de Bruijn's identity from $\mathbb{R}^n$ to the setting of a Riemannian manifold X. In this setting, there is a Fisher information which is a right-invariant metric on the manifold of diffeomorphisms $\mathcal{D}(X)$ from X to X and naturally generalizes the VoI properties of classical Fisher information exploited in our earlier work. We have also generalized de Bruijn's identity to the manifold setting and shown that it relates intrinsic information for a dynamic family of probability densities evolving according to a heat equation to the rate at which entropy is increasing.

In 1966, the Russian mathematician V. I. Arnol'd made substantial advance in the understanding of the flow and dynamics of incompressible, inviscid fluid flow on a Riemannian manifold M by explaining how such flows can be viewed as geodesic flows on $\mathcal{D}(M)$. Through our work, we seek to establish new understanding of information flow and dynamics analogous to this classical result in topological fluid dynamics. This work will appear this fall in [15].

6 Future research plans and anticipated scientific accomplishments

Future research plans in the individual projects are discussed in the context of each project in Sec. 5. Here we discuss some of the broader research plans.

As a direct result of our MURI a fundamental theory of value-of-information is emerging for adaptive sensing, distributed fusion, and information exploitation systems. The value of information is quantified as the incremental change in an objective function due to adding a sample, adding an information source, or degrading the quality of available information, e.g. by quantization or approximation to reduce complexity. Despite the difficulty of this problem, we have been able to establish mathematical bounds and limits on the value of information for several important problems in learning, fusion and control for adaptive sensing. This development of theory is accompanied by new algorithms that provably and practically outperform the state-of-the-art. Several members of our MURI team are working on a manuscript that will lay out these advances in defining value-of-information as a crucial part of the science of sensing.

We will complement the emergent and general theory with advances in specific applications of interest. 1) Kronecker product representations of spatio-temporal process covariance in video and other imaging applications. 2) Refinement of the theory of cooperative human-machine processing that includes decentralized groups of human-machine decisionmakers and its application to target search and classification. 3) Theory of human-in-the-loop processing that accounts for privacy and performs predictive inference of human behavior from partially observed data, in particular preference ranking data and web search data. 4) Random matrix and non-commutative information theory approaches to signal subspace detection and information fusion that can extend the signal detection threshold by several dBs; 5) Exploitation of information sharing in decentralized sensor networks. 6) Value-of-information driven fusion and sensor planning that accounts for mission-dependent rewards. 7) Scalable and accurate proxies for value-of-information. 8) The use of submodularity and adaptive submodularity for bounding the loss due to use of greedy sensor fusion and planning. 9) Enhancing the value-of-information by adaptive sensing strategies including navigation, waveform selection, and beam-scheduling for radar, sonar and active vision modalities.

There are several collaborative projects among the MURI co-PIs that will be pursued this coming year. These include the following:

1. **Software defined radar testbed** (Ongoing). A collaborative DURIP grant for building X-band radars was awarded to us (PI is Emre Ertin) in mid-2013. The primary purpose of these software defined radars is to provide an experimental testbed for MURI researchers. Small radars (breadbox size) and a larger radar (rack mountable) are currently in the final stages of construction and testing. We anticipate that these radars will be ready to be used by other MURI researchers by early fall 2014. As soon as the larger radar is ready it will be shipped to ASU where OSU, Michigan and ASU researchers will collaborate to implement their algorithms.

2. **Mission-aware VoI planning** (Ongoing). The UM (Hero) and MIT (How) groups have adapted UM's convex optimization approach to mission-aware planning problems being pursued in How's laboratory. MIT and UM students have extended the convex approach to the

mission-aware classification problem where identification of different targets have different associated value (See description in Progress 5.3.1). A journal paper is in final stages of preparation and we may perform experiment in How's lab to validate the findings on real experimental data.

3. **Information geometric value of information for sensor planning** (Ongoing). UM (Hero) and ASU (Cochran) had been independently looking at formulation of value-of-information for sensor planning using information geometry. The advantage of information geometric theory of VoI is that it may provide the same simple design rules that come with other geometric theory, such as the projection theorem associated with optimal Wiener filtering, Kalman estimation, and matched filtering for detection. Two different but complementary theories of information geometry for sensor planning have emerged – one a dynamic theory based on adaptive filtrations and the other a static theory based on the geometry of sufficient statistics (probability distributions and likelihood ratios). We made progress on this collaborative project last year (See description in Progress 5.3.3).

4. **Improving information collection via human-in-the-loop processing**. As seen in Progress 5.3.2 and Progress 5.3.2, several on the team have been studying aspects of human-in-the-loop processing. A postdoc has been hired this year at UM to focus on extending human-in-the-loop methods for target tracking to more general problems including detection, classification, and navigation. We anticipate further interactions with UCSD on human cognition models and experimental validation of these models for this problem.

5. **Data-driven Henze-Penrose estimation of Fisher information** (New). A discussion between Hero and Ertin last year sparked a new idea for determining Fisher information by estimation of locally perturbed f-divergences. Hero, in collaboration with ASU researcher Visar Berisha, has made progress on theory and we anticipate having a paper out on this next year.

7 Publications

[1] S. Ahmad, H. Huang, and A. J. Yu, "Context-sensitive active sensing in humans," *Advances in Neural Information Processing Systems*, vol. 26, , 2013.

[2] S. Ahmad and A. J. Yu, "A socially aware Bayesian model for competitive foraging," *Proceedings of the Cognitive Science Society Conference*, 2014.

[3] J. Ash, E. Ertin, L. C. Potter, and E. Zelnio, "Wide-angle synthetic aperture radar imaging: Models and algorithms for anisotropic scattering," *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 16–26, 2014.

[4] R. Cabezas, O. Freifeld, G. Rosman, and J. W. Fisher III, "Aerial reconstructions via probabilistic data fusion," in *IEEE Computer Vision and Pattern Recognition Conference on Computer Vision*, June 2014.

[5] D. Cochran and A. O. Hero, "Information-driven sensor planning: Navigating a statistical manifold," in *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 1049–1052, 2013.

[6] L. Crider and D. Cochran, "Effects of network topology on the conditional distributions of surrogated generalized coherence estimates," in *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*, 2014. (to appear).

[7] J. C. Duchi, M. I. Jordan, M. J. Wainwright, and Y. Zhang, "Information-theoretic lower bounds for distributed statistical estimation with communication constraints," *arXiv preprint arXiv:1405.0782*, 2014.

[8] Z. Dzunic and J. Fisher III, "Bayesian switching interaction analysis under uncertainty," in *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, pp. 220–228, 2014.

[9] E. Ertin, "Manifold learning methods for wide-angle SAR ATR," in *International Conference on Radar (RADAR)*, pp. 500–504, 2013.

[10] E. Ertin, "Information geometry of radar targets," in *Information Theory and Applications Workshop*, San Diego, CA, 2014.

[11] E. Ertin, "Three dimensional imaging of vehicles from sparse apertures in urban environment," in *Compressive Sensing for Urban Radars*, M. Amin, editor, CRC Press, 2014.

[12] K. Greenewald, T. Tsiligkaridis, and A. O. Hero, "Kronecker sum decompositions of space-time data," in *Proceedings of the 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 65–68, 2013.

[13] B. W. Hong and S. Soatto, "Shape matching using multiscale integral invariants," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, July 2014.

[14] S. D. Howard, D. Cochran, W. Moran, and F. Cohen, "Estimation and registration on graphs," *IEEE Transactions on Signal Processing.* (in review).

[15] S. D. Howard, W. Moran, and D. Cochran, "Intrinsic Fisher information on manifolds," in *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*, 2014. (to appear).

[16] V. Karasev, A. Ravichandran, and S. Soatto, "Active frame, location, and detector selection for automated and manual video annotation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2014.

[17] Z. Meng, B. Eriksson, and A. O. Hero, "Learning latent variable Gaussian graphical models," in *Proceedings of the International Conference on Machine Learning (ICML)*, Beijing, 2014. (arXiv preprint arXiv:1406.2721).

[18] Z. Meng, D. Wei, A. O. Hero, and A. Wiesel, "Marginal likelihoods for distributed estimation of graphical model parameters," in *Proceedings of the 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 73–76, 2013.

[19] Z. Meng, D. Wei, A. Wiesel, and A. O. Hero III, "Marginal likelihoods for distributed parameter estimation of Gaussian graphical models," *ArXiv e-prints*, March 2013.

[20] Z. Meng, A. Wiesel, and A. O. Hero, "Distributed principal component analysis on networks via directed graphical models," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vancouver, 2012.

[21] B. Moore, R. R. Nadakuditi, and J. Fessler, "Improved robust PCA using low-rank denoising with optimal singular value shrinkage," in *Proceedings of the IEEE Statistical Signal Processing Workshop*, July 2014.

[22] B. Mu and J. P. How, "Learning sparse Gaussian graphical model with l_0-regularization," in *Advances in Neural Information Processing Systems 27*, 2014. (in review).

[23] R. R. Nadakuditi, "OptShrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 3002–3018, May 2014.

[24] G. E. Newstadt, B. Mu, D. L. Wei, A. O. Hero, and J. P. How, "Mission-weighted adaptive search for multi-class targets," *IEEE Transactions on Signal Processing*, 2014. (in preparation).

[25] G. Papachristoudis and J. W. Fisher III, "Efficient information planning in Markov chains," Technical report, June 2014. (in review).

[26] G. Papachristoudis and J. W. Fisher III, "Incremental belief propagation in Markov random trees," Technical report, June 2014. (in review).

[27] N. Sugavanam and E. Ertin, "Sensor selection and placement in adversarial environment," in *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 185–188, December 2013.

[28] S. Suvorova, B. Moran, S. D. Howard, and D. Cochran, "Control of sensing by navigation on information gradients," in *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 197–200, December 2013.

[29] D. Teng and E. Ertin, "Optimal quantization of likelihood for low complexity sequential testing," in *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 675–678, December 2013.

[30] D. Teng and E. Ertin, "Optimal quantization of likelihood for sequential detection," *IEEE Transactions on Signal Processing*, 2014. (in review).

[31] T. Tsiligkaridis, B. M. Sadler, and A. O. Hero, "Blind collaborative 20 questions for target localization," in *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 161–164, 2013.

[32] T. Tsiligkaridis, B. M. Sadler, and A. O. Hero, "Collaborative 20 questions for target localization," *IEEE Transactions on Information Theory,*, vol. 60, no. 4, pp. 2233–2252, April 2014.

[33] D. Wei and A. O. Hero, "A performance guarantee for adaptive estimation of sparse signals," in *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 189–192, 2013.

[34] G. T. Whipps, E. Ertin, and R. L. Moses, "Distributed detection of binary decisions with collisions in a large, random network," *IEEE Transactions on Signal Processing*. (in review).

[35] G. T. Whipps, E. Ertin, and R. L. Moses, "Distributed detection with collisions in a random, single-hop wireless sensor network," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, May 2013.

[36] G. T. Whipps, E. Ertin, and R. L. Moses, "A consensus-based decentralized EM for a mixture of factor analyzers," in *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, September 2014.

[37] T. Xie, N. Nasrabadi, and A. O. Hero, "Learning to classify with possible sensor failures," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Florence, May 2014.

[38] S. Zhang and A. J. Yu, "Forgetful Bayes and myopic planning: Human learning and decision-making in a bandit setting," *Advances in Neural Information Processing Systems*, vol. 26, , 2013.

[39] Y. Zhang, X. Chen, D. Zhou, and M. I. Jordan, "Spectral methods meet em: A provably optimal algorithm for crowdsourcing," *Submitted to: Neural Information Processing Systems*, 2014.

[40] Y. Zhang, M. Wainwright, and M. I. Jordan, "Lower bounds on the performance of polynomial-time algorithms for sparse linear regression," in *Annual Conference on Computational Learning Theory (arxiv 1402.1918)*, 2014.

8 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 (12)

      1. J. Ash, E. Ertin, L. C. Potter, and E. Zelnio, "Wide-Angle Synthetic Aperture Radar Imaging," *IEEE Signal Processing Magazine*, vol. 31, no. 4, pp. 16–26, 2014.
      2. B. W. Hong and S. Soatto, "Shape Matching using Multiscale Integral Invariants," *IEEE Transactions on Pattern Recognition and Machine Intelligence*, July 2014.
      3. F. Lindsten, M. I. Jordan, and T. Schoen, "Particle Gibbs with ancestral sampling," *Journal of Machine Learning Research*, vol. 15, pp. 2145–2184, 2014.
      4. L. Mackey, M. I. Jordan, R. Chen, B. Farrell, and J. A. Tropp, "Matrix concentration inequalities via the method of exchangeable pairs," *Annals of Probability*, vol. 42, pp. 906–945, 2014.

5. Z. Meng, D. Wei, A. Wiesel, and A. O. Hero, “Marginal Likelihoods for Distributed Parameter Estimation of Gaussian Graphical Models,” *IEEE Transactions on Signal Processing* (in press).
6. B. Mu, G. Chowdhary, and J. P. How, “Efficient distributed sensing using adaptive censoring-based inference,” *Automatica*, vol. 50, no. 6, pp. 1590–1602, 2014.
7. R. R. Nadakuditi, “OptShrink: An Algorithm for Improved Low-Rank Signal Matrix Denoising by Optimal, Data-Driven Singular Value Shrinkage,” *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 3002 - 3018, May 2014.
8. T. Tsiligkaridis, B. M. Sadler, and A. O. Hero, “Collaborative 20 questions for localization,” *IEEE Transactions on Information Theory*, vol. 60, no. 4, pp. 2233–2252, April 2014.
9. T. Tsiligkaridis and A. O. Hero III, “Covariance Estimation in High Dimensions Via Kronecker Product Expansions,” *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 5347–5360, November 2013.
10. D. Wei and A. O. Hero, “Multistage Adaptive Estimation of Sparse Signals,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 5, pp. 783–796, October 2013.
11. X. Zhang, R. R. Nadakuditi, and M. E. J. Newman, “Spectra of random graphs with community structure and arbitrary degrees,” *Physical Review E*, vol. 89, no. 4, 2014.

b. Papers in non-peer-reviewed vehicles

c. Presentations (2)

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

1. Co-PI Ertin gave the presesntation “Information Geometry of Radar Targets,” at the Information Theory and Applications Workshop in San Diego, February 2014.
2. Co-PI Ertin, gave the presentation “Estimation of Information Measures on Target Manifolds,” at the SPIE Conference on Algorithms for Synthetic Aperture Radar Imagery in Baltimore, May 2014.

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

iii. Peer-Reviewed Conference Proceedings publications (21)

1. S. Ahmad and A. J. Yu, “A socially aware Bayesian model for competitive foraging,” *Proceedings of the Cognitive Science Society Conference*, 2014.
2. S. Ahmad, H. Huang, and A. J. Yu, “Context-sensitive active sensing in humans,” *Advances in Neural Information Processing Systems*, vol. 26, 2013.
3. R. Cabezas, O. Freifeld, G. Rosman, and J. W. Fisher III, “Aerial Reconstructions via Probabilistic Data Fusion,” *IEEE Computer Vision and Pattern Recognition Conference on Computer Vision*, June 2014.
4. D. Cochran and A. O. Hero III, “Information-driven sensor planning: Navigating a statistical manifold,” *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 1049–1052, 2013.
5. Z. Dzunic and J. Fisher III, “Bayesian switching interaction analysis under uncertainty,” in *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, pp. 220–228, 2014.

6. E. Ertin, "Manifold learning methods for wide-angle SAR ATR," *Proceedings of the IEEE International Conference on Radar*, pp. 500–504, 2013.
7. K. Greenewald, T. Tsiligkaridis, and A. O. Hero III, "Kronecker sum decompositions of space-time data," *Proceedings of the IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 65–68, 2013.
8. V. Karasev, A. Ravichandran, and S. Soatto, "Active Frame, Location, and Detector Selection for Automated and Manual Video Annotation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2014.
9. G. Marjanovic and A. O. Hero III, "On lq Estimation of Sparse Inverse Covariance," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014.
10. Z. Meng, D. Wei, A. O. Hero III, A. Wiesel, "Marginal likelihoods for distributed estimation of graphical model parameters," *Proceedings of the IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 73–76, 2013. (Received a Best Student Paper Award)
11. Z. Meng, B. Eriksson, and A. O. Hero III, "Learning Latent Variable Gaussian Graphical Models," *Proceedings of the International Conference on Machine Learning (ICML)*, 2014.
12. B. Moore, R. R. Nadakuditi and J. Fessler, "Improved robust PCA using low-rank denoising with optimal singular value shrinkage," *Proceedings of the IEEE Statistical Signal Processing Workshop*, July 2014.
13. N. Sugavanam and E. Ertin, "Sensor selection and placement in adversarial environment," *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 185–188, 2013.
14. S. Suvorova, D. Cochran, B. Moran, and S. D. Howard, "Control of sensing by navigation on information gradients," *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 197–200, 2013.
15. D. Teng and E. Ertin, "Optimal quantization of likelihood for low complexity sequential testing," *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 675–678, 2013.
16. T. Tsiligkaridis, B. M. Sadler, and A. O. Hero III, "Blind Collaborative 20 Questions for Target Localization," *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 161–164, 2013.
17. D. Wei and A. O. Hero III, "A Performance Guarantee for Adaptive Estimation of Sparse Signals," *Proceedings of the IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 189–192, 2013.
18. G. T. Whipps, E. Ertin, and R. L. Moses, "A Consensus-based Decentralized EM for a Mixture of Factor Analyzers," *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, September 2014.
19. T. Xie, N. Nasrabadi and A. O. Hero III, "Learning to classify with possible sensor failures," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014.

20. S. Zhang and A. J. Yu, "Forgetful Bayes and myopic planning: Human learning and decision-making in a bandit setting," *Advances in Neural Information Processing Systems*, vol. 26, 2013.
21. Y. Zhang, M. Wainwright, and M. I. Jordan, "Lower bounds on the performance of polynomial-time algorithms for sparse linear regression," *Annual Conference on Computational Learning Theory*, 2014.

d. Manuscripts (13)

1. L. Crider and D. Cochran,"Effects of network topology on the conditional distributions of surrogated generalized coherence estimates," *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*, November 2014 (to appear).
2. S. D. Howard, D. Cochran, W. Moran, and F. Cohen, "Estimation and registration on graphs," *IEEE Transactions on Signal Processing* (in review).
3. S. D. Howard and W. Moran and D. Cochran, "Intrinsic Fisher Information on manifolds," *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*, November 2014 (to appear).
4. Z. Meng, D. Wei, A. Wiesel, and A. Hero, "Distributed Learning of Gaussian Graphical Models via Marginal Likelihoods," *IEEE Transactions on Signal Processing* (in press).
5. B. Mu and J. P. How, "Learning Sparse Gaussian Graphical Model with l_0-regularization," *Advances in Neural Information Processing Systems 27* (in review).
6. G. E. Newstadt, D. Wei and A.O. Hero, "Resource-Constrained Adaptive Search and Tracking for Sparse Dynamic Targets," *IEEE Transactions on Signal Processing* ArXiv e-print 1404.2201. (in review).
7. G. E. Newstadt, B. Mu, D. Wei, A. O. Hero, and J. P. How, "Mission-weighted adaptive search for multi-class targets," *IEEE Transactions on Signal Processing* (in preparation).
8. G. Papachristoudis and J. W. Fisher III, "Efficient information planning in Markov chains," Technical report, June 2014 (in review).
9. G. Papachristoudis and J. W. Fisher III, "Incremental belief propagation in Markov random trees," Technical report, June 2014 (in review).
10. T. Tsiligkaridis, B. M. Sadler and A. O. Hero III, "A collaborative 20 questions model for target search with human-machine interaction," *IEEE Transactions on Signal Processing* (in review). Available as arxiv:1306.1922.
11. N. Sugavanam and E. Ertin, "Compressive measurement designs for structured signals in signal dependent noise," *Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (in preparation for submission).
12. G. T. Whipps, E. Ertin, and R. L. Moses, "Distributed Detection of Binary Decisions with Collisions in a Large, Random Network," *IEEE Transactions on Signal Processing* (in review).
13. Y. Zhang, X. Chen, D. Zhou, and M. I. Jordan, "Spectral methods meet EM: A provably optimal algorithm for crowdsourcing," *Neural Information Processing Systems* (in review).

e. Books (2)

1. E. Ertin, “Three Dimensional Imaging of Vehicles from Sparse Apertures in Urban Environment,” Book Chapter in *Compressive Sensing for Urban Radars*, M. Amin, ed., CRC Press, 2014.
2. W. Moran, S. Howard, and D. Cochran, “Positive-operator-valued measures: A general setting for frames,” in *Excursions in Harmonic Analysis*, T. Andrews, R. Balan, W. Czaja, K. Okoudjou, and J. Benedetto, eds., vol. 2, pp. 49–64, Springer 2013.

f. Honors and Awards

1. PI Hero received a Lifetime career award: IEEE Signal Processing Society Technical Achievement Award, given at the 2014 International Conference on Acoustics, Speech, and Signal Processing, Florence Italy, May 2014.
2. PI Hero was co-recipient of a best paper award: IEEE CAMSAP 2013 Best Student Paper Competition, awarded 2nd place for a paper coauthored with former student Zhaoshi Meng and former post-docs Dennis Wei and Ami Wiesel entitled “Marginal Likelihoods for Distributed Estimation of Graphical Model Parameters,” IEEE Computational Advances in Multi-Sensor Adaptive Processing workshop, St Martins, December 2013.
3. PI Hero was co-recipient of a best paper award: IEEE ICIP 2013 Best Paper Award, for a paper co-authored with former student Paul Shearer and colleague Anna Gilbert entitled “Correcting Camera Shake by Incremental Sparse Edge Approximation,” at the 2013 IEEE International Conference on Image Processing, Melbourne Australia, September 2013.
4. PI Hero gave the plenary talk “Small sample community detection in massive data sets,” at the IEEE CAMSAP Workshop, St Martins, December 2013.
5. PI Hero gave the keynote address “Resource constrained adaptive sensing,” New Sensing and Statistical Inference Methods Symposium, IEEE GlobalSIP Conference, Austin TX, December 2013.
6. PI Hero gave the keynote talk “Spatio-temporal graphical models for high dimensional network data,” Network Theory Symposium, IEEE GlobalSIP Conference, Austin TX, December 2013.
7. PI Hero gave the keynote talk “Sparsity regularized image reconstruction” at Quantitative Non-Destructive Evaluation (QNDE), Boise ID, July 2014.
8. PI Hero gave the keynote talk “Correlation mining in large networks with limited samples” at the IEEE International Telecommunications Symposium, Sao Paolo Brazil, August 2014.
9. PI Hero will give the plenary presentation “Correlation mining for imaging and multidimensional signal processing,” at the IEEE International Conference on Image Processing, Paris France, October 2014.
10. Co-PI Jordan was awarded the Rumelhart Prize, a $100,000 annual prize for “fundamental contributions to the theoretical foundations of human cognition”
11. Co-PI Jordan was elected Fellow of the International Society for Bayesian Analysis (ISBA)
12. Co-PI Jordan was Keynote Speaker for the Australian Meeting on Statistical Modelling and Analysis of Big Data, Brisbane, February 2014
13. Co-PI Jordan was Keynote Speaker at Artificial Intelligence and Statistics (AISTATS), Reykjavik, April 2014

14. Co-PI Jordan was Keynote Speaker at the Statistical Society of Canada Annual Meeting, Toronto, May 2014
15. Co-PI Jordan was Keynote Speaker at the Computational Learning Theory Annual Conference (COLT), Barcelona, June 2014
16. Co-PI Jordan was Keynote Speaker at the International Conference on Machine Learning (ICML), Beijing, June 2014
17. Co-PI Jordan was Invited Speaker at the Neural Information Processing Systems (NIPS) Workshop on Discrete Optimization, December 2013
18. Co-PI Jordan was Invited Speaker at the Neural Information Processing Systems (NIPS) Workshop on Big Data Analysis, December 2013

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 (22)

(a) ASU student Shih-Ling Phuong supported at 25% annualized FTE
(b) MIT student Beipeng Mu supported at 50% annualized FTE
(c) MIT student Georgios Paperchristoudis at 50% annualized FTE
(d) MIT student Christopher Dean at 50% annualized FTE
(e) MIT student Randi Cabezas at 50% annualized FTE
(f) MIT student Julian Straub at 50% annualized FTE
(g) OSU student Nithin Sugavanam supported at 50% annualized FTE
(h) OSU student Diyan Teng supported at 50% annualized FTE
(i) OSU student Gene Whipps supported at 0% annualized FTE (Note: Whipps is a researcher at Army Research Laboratory who is on temporary assignment to Ohio State in order to complete his PhD degree. His research is fully aligned with this MURI)
(j) UC Berkeley student Nicholas Boyd
(k) UC Berkeley student John Duchi
(l) UC Berkeley student Fabian Wauthier
(m) UC Berkeley student Yuchen Zhang
(n) UCLA student Georgios Georgiadis supported at 5% annualized FTE
(o) UCLA student Nikos Karianakis supported at 50% annualized FTE
(p) UCLA student Vasiliy Karasev supported at 50% annualized FTE
(q) UCSD student Sheeraz Ahmad supported at 25% annualized FTE
(r) UM student Pin-Yu Chen supported at 50% FTE
(s) UM student Tianpei Xie supported at 50% FTE
(t) UM student Zhaoshi Meng supported at 25% FTE
(u) UM student Raj Tejas Suryaprakash at 50% FTE August, 2013. 25% FTW January 2014 - April, 2014. Total Effort: .25

  (v) UM student Nick Asendorf at 50% FTE September 2013 - April, 2014. Total Effort: .67

2. Masters Students (1)
   (a.) ASU student Kaitlyn Beaudet supported at 20% annualized FTE

b. Post Doctorates (7)
   1. Jie Chen, UM, 50% July 2014
   2. Karl Liechty: 50% 2014 - May, 2014, Total Effort: .20
   3. Oren Freifeld, MIT, 25% annualized FTE
   4. Goran Marjanovic, UM, 50% annualized FTE
   5. Guy Rosman, MIT, 25% annualized FTE
   6. Shunan Zhang, UCSD, 50% annualized FTE
   7. Dennis Wei, UM, 50% August 2013

c. Faculty (9)
   1. Douglas Cochran, ASU (14% FTE)
   2. Emre Ertin, OSU (20% FTE)
   3. Randy Moses, OSU (0% FTE)
   4. John Fisher, MIT (20% FTE)
   5. Alfred Hero, UM (15% September-May 2014)
   6. Jonathan How, MIT (0%)
   7. Michael Jordan, UC Berkeley (0%)
   8. Raj Nadakaduti, UM (0%)
   9. Angela Yu, UCSD (33% FTE)

d. Undergraduate Students (1)
   1. ASU student Lauren Crider supported at 15% annualized FTE

e. Graduating Undergraduate Metrics (funded by this agreement and graduating during this reporting period):

f. Masters Degrees Awarded (1)
   1. Randi Cabezas, MIT

g. Ph.D.s Awarded (3)
   1. Jason Chang, MIT
   2. John Duchi, UC Berkeley
   3. Theodoris Tsiligkaridis, UM

h. Other Research Staff

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. Technology Transitions

2. Student interns at Service Labs
   (a) ASU student Shih-Ling Phuong did an internship at ARL under Ethan Stump in summer 2014.
   (b) UM student Brandon Oselio did an internship at ARL under Lance Kaplan in summer 2014.
   (c) UM student Tianpei Xie did an internship at ARL under Nasser Nasrabadi in summer 2014.
   (d) UM student Ted Tsiligkaridis did an internship at ARL under Brian Sadler in summer 2013.
   (e) UM student Kristjan Greenewald did an internship at the AFRL ATR Center on topics related to this MURI in Summer 2014.
   (f) ASU students Jakob Hansen and Anthony Helmstetter did internships in the AFRL ATR Center on topics relevant to this MURI in Summer 2014.
3. Co-PI interactions with ARL and other Federal research labs
   (a) PI Hero and co-PIs Cochran, Ertin, Fisher, and Yu visited ARL in June 2014 where they briefed ARL researchers on research under this grant and engaged in technical discussions. ARL researchers engaged in this visit included Lance Kaplan, Brian Sadler, Nasser Nasrabadi, and Ethan Stump.
   (b) PI Hero visited AFRL in Aug. 2013 and again in Dec. 2013 where he spoke with Ed Zelnio and Jeff Simmons about MURI related topics.
4. Other relevant co-PI activities on national committees
   (a) PI Hero served on the National Academy of Sciences Committee on Applied and Theoretical Statistics (2011-present)
   (b) Co-PI Moses served on the National Academy of Sciences Panel on Information Science at the Army Research Laboratory (2011-present).
5. Internal MURI visits
   (a) Co-PI Cochran visited UM for collaborative work on this project with PI Hero and co-PI Nadakuditi (January 2014).
   (b) PI Hero visited ASU for collaboration with Co-PI Doug Cochran (February 2014).
   (c) PI Hero visited MIT for collaboration with Co-PIs John Fisher and Jon How (July 2014).
   (d) PI Hero visited CalTech for collaboration with co-PI Jordan's former post-doc Venkat Chandrasekharan (April 2014).
   (e) Co-PI Jon How's student Beipeng Mu visited UM to work with PI Hero (July-August 2013).
   (f) PI Hero's student Greg Newstadt visited MIT to work with Co-PI Jon How (August 2013).