INTERIM PROGRESS REPORT

Period covered by report: 1 August 2014 to 31 July 2015

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 a new measure of VoI that is empirically estimatable from the length of a multi-colored minimal spanning tree to a convergence results on the popular ADMM optimization algorithm used in many relevant machine learning applications. In the distributed information fusion thrust, progress is reported in collaborative fusion over sensor networks to dimensionality reduction and subspace tracking. In the active information exploitation thrust, progress is reported in new convex proxies for VoI in wide area search problems to the use of VoI for two-stage navigation of autonomous robot vehicles. Our future plans are to continue to develop fundamental theory for VoI, to develop algorithms that optimize VoI for agile sensing applications, and to validate our results on real data, including a software defined radar testbed developed under a DURIP.

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 8/1/14 — 7/31/15

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: John Fisher (MIT)

Publications: [1]

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: Inference in Sparse Graphical Models (Fisher MIT):

Many estimation problems can be cast as inference in graphical models, where nodes represent variables of interest and edges between them indicate dependence relations. Naïve inference may have exponential complexity in the number of variables. Message passing algorithms, such as BP reduce the complexity significantly. Despite the fact that BP performs exact inference only on trees, it is often applied to loopy graphs (for which it is approximate) due to its computational efficiency.

We consider the problem of inference in large-scale models. Such models which arise, for example, in complex spatio-temporal phenomena, may be utilized in multiple settings. It is often the case that only a subset of latent variables is of interest for different applications which may vary from instance to instance. Additionally, the set of available measurements may vary with use or become available at different points in time. The latter is common for any sequential estimation problem. In such situations, general-purpose inference algorithms, such as BP may utilize many unnecessary computations when only a small subset is desired. The complexity of such approaches becomes prohibitive as the size of graph increases, e.g., due to constant re-evaluation of messages. There exist several examples that fall into this category of problems. Patient monitoring provides one such practical example. Large-scale systems may monitor the health status of many patients; however, different physicians limit their interest to patients under their immediate care. Temperature monitoring sensors provide data over time and space, but sensitive areas (e.g., server room) may require more careful examination for the timely response in case of abnormal behavior. Lastly, in computational biology, the effects of mutations are explored (*computational mutagenesis*), with each putative mutation resulting in a very similar problem.

This motivates methods for problems where measurements are added incrementally and the interest is in a subset of node marginals at a given time point or the MAP sequence of the full latent graph. This is the problem of *adaptive inference*, where the goal is to take advantage of previously computed quantities instead of performing inference from scratch. In these cases, standard BP results in many redundant computations. Consequently, we develop an adaptive inference approach [1] which avoids redundant computations and whose average-case performance shows significantly lower complexity compared to BP. The main idea is to send only messages between the node where a measurement has been obtained from (w_ℓ) and the node whose marginal is of interest (v_ℓ). The correctness of this approach is guaranteed by propagating messages between consecutive measurement nodes $w_{\ell-1}, w_\ell$ at every iteration. As a result, we only send the necessary messages to guarantee that the incoming messages to the node of interest v_ℓ are correct. We call this minimal messaging schedule *adaptive BP*. We show that it gives exact results on trees (as standard BP) and provide an extension for Gaussian loopy graphs that still guarantees exactness in the evaluation of marginals.

The proposed method requires a preprocessing step of $\mathcal{O}(N \log N)$ time, where N is the number of latent nodes. In the worst case, when relative distance between consecutive "measurement" nodes is approximately the tree diameter and the diameter is on the order of N (highly unbalanced tree), the performance is comparable – yet still faster to– standard BP. However, for height-balanced trees worst-case performance results in $\mathcal{O}(\log N)$ messages per update as compared to $\mathcal{O}(N)$ for standard BP. In the worst case, if distance of consecutive nodes is very small, the computation of the node marginal is obtained in constant time per iteration. We provide an extension of the method for MAP inference and for Gaussian loopy MRFs and show how it can be used to suggest nearly optimal measurement schedules. This method was compared to alternative approaches and the regimes in which one approach may have advantages over the other were examined. We also empirically demonstrated the performance of our method in a variety of synthetic datasets, as well for two real applications.

5.1.2 Estimation and optimization of VoI metrics

Contributors: Alfred Hero (UM), Michael Jordan (UCB)
Publications: [2], [3]

This year we have made progress on the estimation and optimization of objective functions that can be used for evaluating the value of information, distributed prediction, and maximum likelihood estimators. Two major advances are reported: 1) empirical data-driven estimation of entropic VoI measures and 2) convergence of ADMM optimizers. ADMM stands for alternating direction method of multipliers and is one of the most widely used methods for constrained optimizaation in machine learning, signal processing and statistical inference.

Progress 2: Empirical data-driven estimation of entropic VoI measures (Hero UM)
This year we report a completely new approach to estimation of entropy-based VoI measures from observational data in any dimension. These measures include α-entropies and f-divergences that can be related directly to VoI measures such as estimator MSE, detection ROC curves, or classifier confusion matrices. Our approach eliminates the need for mathematical models, statistical probability densities, or computation of complicated integral representations of MSE or probability of error. The general concept is based on building entropic graphs over the data set from which asymptotically consistent estimators of certain f-divergences can be extracted without the need to estimate the underlying model or probability density. We have applied this approach to empiri-

cal estimation of the Fisher Information. The Fisher information matrix (FIM) is a foundational concept in statistical signal processing. The FIM depends on the probability distribution, assumed to belong to a smooth parametric family. Traditional approaches to estimating the FIM require estimating the probability distribution function (PDF), or its parameters, along with its gradient or Hessian. However, in many practical situations the PDF of the data is not known but the statistician has access to an observation sample for any parameter value. We proposes a method of estimating the FIM directly from sampled data that does not require knowledge of the underlying PDF. The method is based on non-parametric estimation of an f-divergence over a local neighborhood of the parameter space and a relation between curvature of the f-divergence and the FIM. Thus we obtained an empirical estimator of the FIM that does not require density estimation and is asymptotically consistent. We empirically evaluated the validity of our approach in the context of learning the optimal tuning parameters of hearing aids from numerical data. This work was reported earlier this year in the IEEE Signal Processing Letters [2] and was presented as an oral presentation at the the 2015 IEEE ICASSP meeting.

Progress 3: Convergence of ADMM optimizers (Jordan UCB)
We provide a new proof of the linear convergence of the alternating direction method of multipliers (ADMM) when one of the objective terms is strongly convex. Our proof is based on reducing algorithm convergence to verifying the stability of a dynamical system. This approach generalizes a number of existing results and obviates any assumptions about specific choices of algorithm parameters. On a numerical example, we demonstrate that minimizing the derived bound on the convergence rate provides a practical approach to selecting algorithm parameters for particular ADMM instances. We complement our upper bound by constructing a nearly-matching lower bound on the worst-case rate of convergence. This work will be published in the Proceedings of the upcoming Intl Confernence on Machine Learning (ICML) [3].

5.1.3 Representation of information for visual sensors

Contributors: Stefano Soatto (UCLA)
Publications: [4], [5], [6], [7], [8], [9]

While multimodality sensing is a major theme of our MURI, including radar, acoustic, seismic, and soft contextual information, visual data is of particular interest due to its high complexity, potential value for scene understanding and navigation. We report on several advances including: learning visual representations, domain-size pooling, visual textures and shape scaling.

Progress 4: Learning Visual Representations (Soatto UCLA)
For tasks involving interaction with physical space, the inference of a "representation" of the surrounding space from data should be guided by maximizing task-specific information. The representation (a function of past data) then plays the role of "memory" or "state" of the system, and should ideally be as informative as the data for the purpose of the task at hand, but have bounded complexity. Furthermore, the task informs the inference of a representation by determining what aspects of the data formation process is relevant. Otherwise, *nuisance factors* shoud be discounted in the representation, as well as in the computation of task-specific information. Ideally, a representation should be a minimal sufficient statistic that is invariant to nuisance factors.

It has long been believed that these two requirements (sufficiency and invariance) are conflicting, and therefore there is a tradeoff between the two. Indeed, when invariance is achieved by averaging statistics [10, 11], it comes at the expense of discriminative power, or at an information loss for classification tasks. However, we have recently shown [4] that this is not necessarily the case, and strict invariance to nuisance factors that are composition of independent group actions can be achieved by a combination of profiling (maximization) and marginalization of likelihood functions.

Progress 5: Domain-Size Pooling (Soatto UCLA)

This helps explaining the empirical success of "domain-size pooling" [5] that consists of averaging empirical likelihood functions (histograms) on domains of different sizes, which would seem at first counter-productive as one aggregates statistics from different regions. Of course, there is no "free lunch" and the price to pay to achieve invariance without loss of discriminative power is the need to lift the representation from sample statistics to the (infinite-dimensional) space of likelihood functions. Such a space has necessarily to be approximated, for instance non-parametrically in [6] but with a very crude model assuming (conditional) independence of gradient orientations at each pixel, which reduces the (local) representation to the concatenation of one-dimensional histograms. More accurate estimates of minimal sufficient and invariant statistics requires aggregation in high-dimensional spaces, which is one of the foci of Thrust 1 (learning and representation of high dimensional data).

Domain-size pooling consists of the aggregation of image statistics not just in spatial neighborhoods, as customary in the design of local representations of image data, but also in scale neighborhoods, which has been shown to improve the performance on image matching in benchmark datasets (the Oxford Matching Dataset) by up to 39% (mean-average precision).

The theoretical justification for performing domain-size pooling is rooted in sampling theory, and would at first seem to go counter to the teaching of harmonic analysis and the "uncertainty principle" that predicates the size of the spatial domain should be matched to the spatial frequency of the signal being represented. But while such a link is sensible for data transmission and storage tasks, it is not when the *task* is *correspondence* of different sensory data, for instance imaging data where the size of the visible domain depends on occlusion and other spatial properties of the scene (shape) and have little to do with its reflectance property (appearance). Yet, such a link has been a mainstay in local image descriptors for a quarter century, following the theory of scale-space, that was rooted in harmonic analysis and wavelets.

Progress 6: Visual Textures (Soatto UCLA)

Another special case of relevance for local representations is when the underlying scene radiance (informally the "appearance") can be considered to be a sample from a stationary process. This is when the portion of the region is known as a *texture*. In this case, aggregating statistics in high dimensions is still costly, but justified since the statistics are spatially homogeneous, so once inferred from a neighborhood of a pixel, they can be extended to entire regions [7].

Progress 7: Shape Scaling (Soatto UCLA)

For tasks requiring correspondence of remote sensory modalities, *scaling phenomena* must be managed in a task-specific representation. Domain-size pooling is a way of managing scale for local photometric correspondence, relevant for instance in EO image matching. For *geometric correspo-nence*, relevant for instance in shape matching or range imaging, the issue of scale remains critical, and some progress has been described [8], where we have introduced shape signatures (invariant statistics) that aggregate data across different scale.

Of course, even an optimal representation can be rather uninformative if the data provided is such, as any representation can be at most as informative as the data from which it is computed (data processing inequality). This brings into focus the problem of active inference, or experiment design, which is necessary not just to guarantee that the representation is "best" for the task *given the data available*, but that the data is gathered in such a way that the performance in the task (or the task-relevant information) is maximized. This is addressed in Thrust 3 of this MURI (active information exploitation).

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: Emre Ertin (OSU), Randy Moses (OSU), Al Hero (UM)
Publications: [12], [13], [14], [15], [16], [17]

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. For a successful decentralized learning protocol the sensors reach a consensus about the state, class, or other latent variable after a sufficient amount of information sharing. Several advances have been made this year on decentralized learning and information fusion: i) decentralized learning of a mixture of factor analyzers; ii) Aggregating local information under communication constraints for decision-level fusion; iii) decentralized cooperative target tracking.

Progress 8: Decentralized learning of a mixture of factor analyzers (Ertin and Moses OSU)
In this research thrust, 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 [12]. 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. 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. 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.

We have adopted the MFA model in our recent work in modeling Synthetic Aperture Radar (SAR) target signatures for performance prediction [13]. Specifically, we study the problem of target identification from SAR imagery. Target classification using SAR imagery is a challenging problem due to large variations of target signature as the target aspect angle changes. Previous work on modeling wide angle SAR imagery has shown that point features, ex- tracted from scattering center locations, result in a high dimensional feature vector that lies on a low dimensional manifold. We employ MFA models for these target manifolds to analyze classification per- formance as a function of Signal-to-noise ratio (SNR) and Bandwidth. We employ Mixture of Factor Analyzers (MoFA) models to approximate the target manifold locally, and use error bounds for the estimation and analysis of classification error performance. We compare our performance predictions with the empirical performance of practical classifiers using simulated wideband SAR signatures of civilian vehicles.

Progress 9: 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 previous work [14] we considered 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. Furthermore, 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.

In our recent work [15], we consider the problem of distributed detection of a radioactive source using a network of emission count sensors. Sensor nodes observe their environment and a central fusion node attempts to detect a change in the joint probability distribution due to the appearance of a hazardous source at an unknown time and location. We consider a minimax-type distributed change- point detection problem that minimizes detection delay for a desired false alarm rate. A statistical model of the radiation source detection problem is formulated where sensors observations are correlated with non-identical distributions. We first derive a centralized detection algorithm that is asymptotically optimal for vanishing false alarm rate. Then we analyze the performance loss, as measured by the detection latency, when sensor counts are quantized at each sensor node. The detection latency of the centralized rule provides a lower bound on performance for the proposed distributed method. The empirical results indicate that the distributed detection strategy provides a reasonable tradeoff between latency and information bandwidth.

Progress 10: Decentralized cooperative tracking (Hero UM)
We continue to develop the 20-questions framework for tracking where multiple agents cooperate to locate the position of a target based on sensing inputs; e.g., noisy imagery or ranging data. Last year we reported on our collaborative 20 questions framework with a human response model, which was published in the IEEE Information Theory Transactions [17] this year. This year we report on a networked multi-agent extension of this work that applies to a large network of agents that iteratively converge on target location estimates using message passing in a decentralized 20 questions framework. The social learning model for information sharing is used and we proved convergence and concensus of our distributed algorithm under this model. This work has been published in the IEEE Signal Processing Transactions [16]. Over the coming year we have the following plans. 1) to establish tight bounds on entropy-based approaches to 20-questions like the ones taken in [17] and [16]; 2) to use these bounds to predict when this entropy-minimization approach will have slow convergence; 3) to develop a new approach to collaborative 20-questions model based on minimizing a weighted entropy criterion that is related to unequal error protection in channel coding theory. We will then apply this relation to define improved collaborative 20-questions approaches that have better convergence rates. This work is in collaboration with Brian Sadler at ARL.

5.2.2 Subspace processing and fusion of information

Contributors: Raj Nadakuditi (UM), Michael Jordan (UCB), Alfred Hero (UM),
Publications: [18], [19],[20],[21], [22],[23], [24]

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 measures and subspace detection from random matrices; ii) generalized matrix rank estimation.

Progress 11: Spectral measures and subspace detection from random matrices (Nadakuditi UM)
In the area of non-commutative information theory, we have established 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.

These non-commutative value of information type metrics can help quantify the "informativeness" of an information sources. This past year, we focused on utilizing these metrics to improve the fusion of multiple sources, where each source is assumed to have different SNR. The key idea that we exploit is that signals of interest or targets will occupy different low dimensional subspaces for each modality the *expressiveness* of a modality depends on the target signature for that modality (e.g. hyperspectral versus EO/IR). A combination of these probing multiple modalities yields optimal detection or classification performance.

The technical challenge is to automatically compute the weighting coefficient that is to be assigned to each modality modalities with greater informational content should receive a higher weight while modalities with lower information content should receive a lower weight or not be used at all.

Last year we made substantial progress in developing data-driven algorithm (OptFuse) [18] which computes the optimal linear combination of the signal-plus-noise matrices that produces the most accurate estimate of the latent signal subspace. The algorithm builds on the OptShrink algorithm [25] developed in the previous reporting year for denoising a low-rank signal matrix buried in noise by optimal singular value shrinkage. OptFuse also explicitly utilizes information in the "noise portion" of the singular value spectrum to compute these optimal linear weighting coefficients and returns an estimate of the approximation MSE that is provably consistent and that can serve as a new VoI metric.

These ideas can be applied to the fusion of graph-valued signals. We have initiated collaborations with ARL (Dr. Ananthram Swami and Dr. Terrence Moore) on inference from time-varying, multi-modal graphs. A PhD student (Himanshu Nayar) spent the summer of 2015 at ARL furthering this collaboration. We are actively collaborating with MURI co-PIs to investigate applications of the multi-modal fusion algorithm. In the upcoming year, we plan to test our algorithms on real-world datasets generated using the DURIP testbed. Other successes in the past year include a new algorithm for estimation of low-rank matrices with Kronecker structure [19], new algorithms for passive bistatic radar detection [20, 21], finite sample performance analysis of the Tucker HOSVD algorithm [26] and the MUSIC DOA algorithm [22, 23] under noise and missing data assumptions.

Progress 12: Generalized matrix rank estimation (Jordan UCB)
The estimation of matrix rank is an important precursor to subspace processing and tracking and s at the core of most data fusion algorithms. We have studied the following generalized matrix rank estimation problem: given an $n \times n$ matrix and a constant $c > 0$, estimate the number of eigenvalues that are greater than c. In the distributed setting, the matrix of interest is the sum of m matrices held by separate machines. We show that any deterministic algorithm solving this problem must communicate $\Omega(n)$ bits, which is order-equivalent to transmitting the whole matrix. In contrast, we propose a randomized algorithm that communicates only $O(n)$ bits. The upper bound is matched by an $\Omega(n)$ lower bound on the randomized communication complexity. We demonstrate the practical effectiveness of the proposed algorithm with numerical experiments. This work will appear at the Intl Conference on Machine Learning (ICML) [24].

5.2.3 Robust information-driven fusion

Contributors: Alfred Hero (UM), Emre Ertin (OSU)
Publications: [27], [28, 29]

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 i) information fusion with partial and unreliable information, and ii) learning to aggregate information for sequential inference. The former is a parametric Bayesian approach while the latter is a non-parametric approach to fusion.

Progress 13: Information fusion with partial and unreliable information (Hero UM)
In collaboration with Nasser Nasrabadi at ARL, we have continued to address the important problem of sensor fusion in the presence of partial or unreliable information. Last year we reported a new approach published in the 2014 IEEE ICASSP proceedings [30] that applied minimum entropy discrimination (MED) to multimodality sensing systems when training data may be contaminated by sensor failures. This year we report progress on applying a similar MED approach to fusing partial information collected from multiple views of an information source when there are many unlabeled samples. We extend the MED approach, developed by Jaakola (2001), to perform semi-supervised information fusion by combining the multiple view data in such a way so as to maximize

the value of information (relative entropy) subject to a constraint on classifier performance. The problem is formulated under the multi-view learning framework and a Consensus-based Multi-View Maximum Entropy Discrimination (CMV-MED) algorithm is proposed. By iteratively maximizing the stochastic agreement between multiple classifiers on the unlabeled dataset, the algorithm simultaneously learns multiple high accuracy classifiers. We demonstrate that our proposed method can yield improved performance over previous multi-view learning approaches by comparing performance on three real multi-sensor data sets, including the ARL footstep data evaluated in our previous work [30]. In the coming year we will be preparing journal papers on this work and applying the work to data acquired with OSU's software radar testbed. This work was presented as an oral presentation and was published in the 2015 IEEE ICASSP Proceedings [27]. This is joint work performed in collaboration with Nasser Nasrabadi at ARL.

Progress 14: Learning to aggregate information for sequential inference (Ertin OSU) Sequential decision strategies outperform their fixed sample size counterparts in achieving same decision risk using less number of samples on the average. Even when the cost of samples are not a major concern, sequential techniques can be used to reduce the computational cost of obtaining relevant information from a data sample. Thus sequential test is still a method of great potential in any time sensitive scenario. For example, in many computer vision problems, more sophisticated feature is usually expensive and slow to obtain even though they provide higher accuracy. Therefore cascading classifier is widely used due to their sequential nature. For the case of known class conditional densities accumulating likelihood statistics and comparing with fixed thresholds minimizes the average stopping time under fixed error constraints. In our recent research [28, 29], we consider the case where the class conditional densities generating the data is unknown and sequential decision rule has to be learned directly from labeled data samples. While there exists plethora of supervised learning algorithms to learn fixed sample test rules using parametric and non-parametric forms, there exist relatively few algorithms designed to learn to perform sequential classification. Unlike the single sample classification problems where only the decision boundary is critical, sequential decision rules require a mapping from sample space to a state space for aggregation of evidence and making stopping rules. First, using Martingale theory we derive an upper bound on stopping performance of learned likelihood ratio function estimators. Next, we show that the problem of minimizing this upper-bound can be posed as a convex optimization problem using a Reproducing Kernel Hilbert Space representation for the log-density ratio function. The resulting binary sequential classifier is tested on synthetic and real world data sets comparing its performance to previously suggested approaches for density ratio estimation. Our empirical results show that the classifier trained with the modified error metric tailored for sequential inference achieves smaller average sampling cost than previous classifiers proposed in the literature for the same error rate.

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.

[31] this past year. Last year we reported progress on fundamental limits specifying the value of information gained by using such convex proxies for planning. This year we have published a complete analysis of VoI in IEEE Transactions on Information Theory [32] for the case of target detection and localization. In collaboration with co-PI John How, we have used this analysis to motivate a convex proxy approach to sensor planning for multi-class targets where successful detection of different classes of targets have different value payoffs relative to the mission. A conference paper on this multi-class convex proxy approach was published in a special session on information-centric fusion in the proceedings of the IEEE Fusion conference [33]

5.3.1 Optimization approaches to sequential planning and navigation

Contributors:	Michael Jordan (UCB), Alfred Hero (UM), John How (MIT), John Fisher (MIT), Doug Cochran (ASU)
Publications:	[34], [31], [32], [33], [35], [36], [37], [38], [39].

Sequential multistage planning and decisionmaking is capable of significant performance gains relative to static offline approaches to resource constrained sensing. Using feedback of information at the current stage to improve the action taken to collect information at the next stage can reduce learning delay and time to detect a target. This year progress is reported in: convergence of greedy optimizers of submodular functions, two-stage focused inference for navigation, and convex proxies for VoI-driven search.

Progress 15: Convergence of greedy optimizers of submodular functions (Jordan UCB)
Submodular functions describe a variety of discrete problems in machine learning, signal processing, computer vision and, in particular, multistage sensor planning using mutual information as the planning criterion. However, minimizing submodular functions poses a number of algorithmic challenges. Recent work introduced an easy-to-use, parallelizable algorithm for minimizing submodular functions that decompose as the sum of "simple" submodular functions. Empirically, this algorithm performs extremely well, but no theoretical analysis was given. In this paper, we show that the algorithm converges linearly, and we provide upper and lower bounds on the rate of convergence. Our proof relies on the geometry of submodular polyhedra and draws on results from spectral graph theory. This work was published in the Proceedings of the 2014 Conference on Neural Information Processing Systems (NIPS) [34].

Progress 16: Two-stage focused inference for navigation (How MIT)
Sensor breakthroughs in the past decade have greatly improved the capability of robots to gather data about the environments. These technologies have enabled many new capabilities for mobile robots operating in complicated, partially-known worlds, but they also introduce new challenges on processing data and extract valuable information. This work considers scenarios in which mobile robots operate in uncertain and GPS-denied environments, and thus must autonomously build a map that can be used to achieve some mission objective. As the robot continues to explore

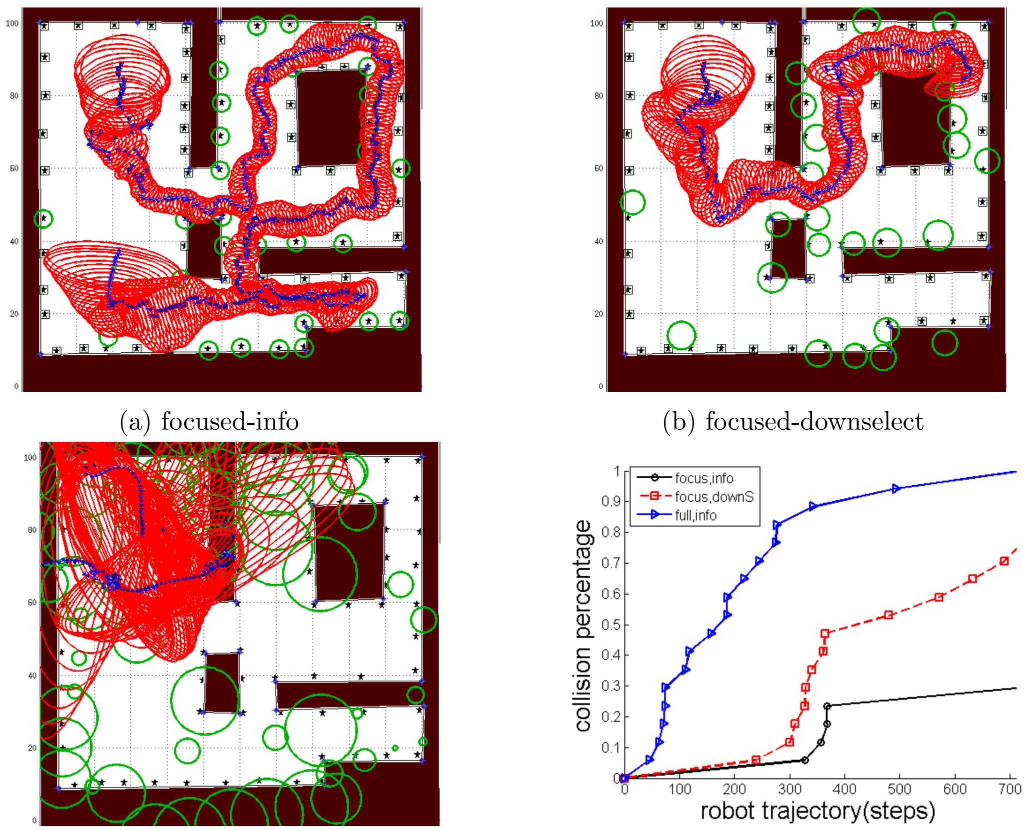
more of the environment, the data collected and the features/variables that must be kept track off in the environment will increase. Thus for long-term operations for a robot that is resource-constrained, some hard decisions will have to be made about which variables (i.e. landmarks) and data (measurements from a chosen landmark) to process - there is likely going to be too much data for the robot to be able to process it all. This research tackles the problem of choosing which data is likely to be the most useful to help achieve the mission task, which is a key VoI question. Furthermore this work presents the first known approach to tackle *both* variable growth and data growth.

The approach has two stages. First, a subset of the variables (focused variables) is selected that is most useful for a particular task. Second, a task-agnostic and principled method (focused inference) is proposed to select a subset of the measurements that maximizes the information over the focused variables. The approach is then applied to the specific task of robot navigation in an obstacle-rich environment. In this case, variables are locations of landmarks and data is measurements of landmark locations. A set of focused landmarks are selected to minimize the probability of collision and then a set of measurements are selected to best localize those landmarks. In the experiment, we compare 3 approaches:

1. *focused-info*: the proposed two-stage approach that selects focused landmarks in terms of minimizing collision probability and selects measurements based on information gain over focused landmarks.
2. *focused-downselect*: select focused landmarks but downsampling measurements of these focused landmarks.
3. *full-info* use all landmarks but select measurements based on information gain over all landmarks.

Fig. 1a–1c shows sample navigation trials with the map built with the three different algorithms. Fig. 1d shows the overall probabilities of collision obtained from all trials. The trials are stopped whenever there is an actual collision with an obstacle. In the newly developed *focused-info* case, the procedure picks the landmarks that contribute the most to reducing the robot's uncertainty in task-important regions (narrow passages) and exerts the computational budget to process the data that helps reduce the uncertainty of these focused landmarks. In case *full-info*, the measurements are selected to maximize information on all available landmarks. As a result, only a limited amount of resources are spent on each landmark. In particular, very little resource is expended to reduce the uncertainty in the important landmarks and the method failed to recover a meaningful map for navigation. In the *focused-downselect* case, more computational resources are spent on the focused landmarks, thus the map is more accurate than *full-info*, but the measurement selection is not based on how much they contribute to uncertainty reduction, so bad choices are made and the landmark positions are much less accurate than *focused-info*. The key point to note about the proposed approach in Fig. 1a is that the robot uncertainty is preferentially reduced (smaller uncertainty ellipses) in the areas of the environment where the corridors are tight and there is a higher chance of collision.

In the real-world experiment, we ran a Pioneer robot in a cluttered office space. Fig. 2a shows the floor plan of the environment. AprilTags were put up to create an initial pool of landmarks. The



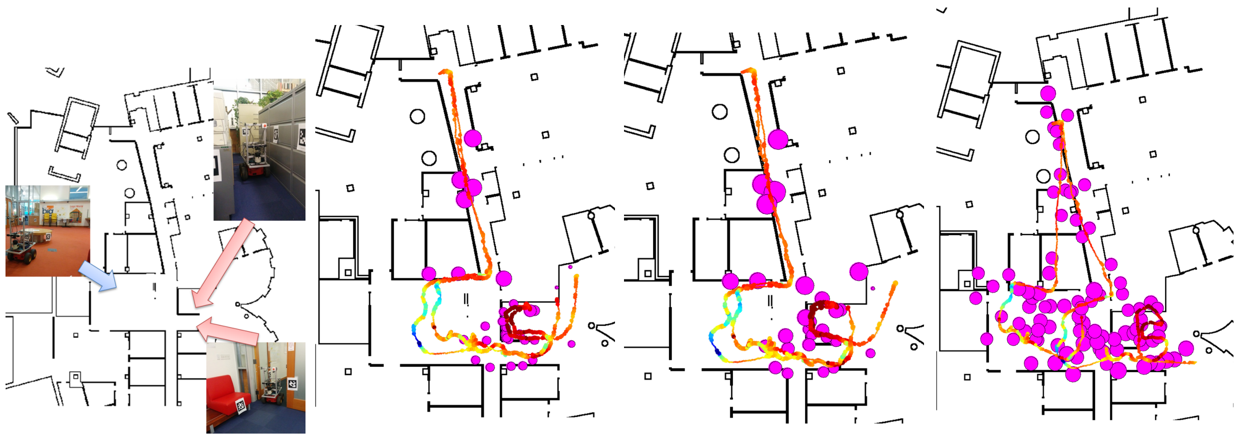
(a) focused-info (b) focused-downselect

(c) full-info (d) Collision probability in Monte Carlo sim.

Figure 1: Comparison of map building results and navigation with different maps. Green circles represent learned landmark positions with their size representing uncertainty. Blue lines are the nominal trajectories each robot wants to follow with red circles representing pose uncertainty. The two-stage focused approach has much more accurate landmark estimates and much less uncertainty in narrow passages, thus lower collision probability compared to the unfocused case

odometry measurements are obtained by matching visual features in consecutive RGBD images. Fig. 2 compares mapping results of the same cases. The rebuilt robot trajectory is shown with a color map, where the red color on the trajectory indicates the risky (close to obstacles) regions and blue indicates the safer regions. Magenta circles represent landmarks with the size representing its uncertainty. The *focused-info* approach (Fig. 2b) can concentrate the measurements on the narrow passage and door way, resulting in less uncertainty there. The other approaches scatter the measurements across different landmarks, and thus have much higher landmark uncertainty in narrow passages.

Both simulation and hardware experiment showed the the proposed approach can build a reduced map that retains the important information about the environment. The reduced map is much smaller than the original full map with all available variables and measurements, thus make it much more memory and computationally efficient for navigation. These results have been reported in the conference publication [35].



(a) floor map (b) focused-info (c) focused-downselect (d) full-info

Figure 2: Mapping results. Color line represent robots risk of collision. Magenta circles represent landmarks with the size representing its uncertainty. The proposed two-stage approach (focused-info) outperforms either only selecting measurements (full-info) or only selecting landmarks (focused-downselect).

Our future research will further explore our new approach of combining model learning and trajectory planning as an integrated problem. With a partial map learned from the places the robot has visited, the new approach samples feasible points in the partial map, and computes a path from the feasible points that gives maximum future information on the robot tasks. Our new approach explicitly accounts for the fact that when the robot moves to the boundary, it can observe new features from previously unknown regions, thus gain new information about the environment. The information in observed features and new features is quantified with a unified metric. Therefore, the proposed approach can compare the benefits between exploring unknown regions and revisiting explored regions with the same metric, and balance between them.

Progress 17: Convex proxies for VoI-driven search (Hero UM)
Results on convex VoI-driven strategies for the wide area search problem have been developed and published this year. Convex proxies are very effective since they are easily optimized over the planning space, e.g., the amount radar energy allocated to a particular region of a field of view or the particular sensing modality selected. We continue to use a simple convex proxy that captures detection and localization performance in a natural manner. This convex proxy is very simple to analyze and we can often obtain closed form analytical characterizations of the optimal policy, the associated VoI, and the associated exploration vs exploitation tradeoff. This has been applied to adaptive search and tracking of sparse dynamic targets under resource constraints, published in the IEEE Trans. of Signal Processing [31] this past year. Last year we reported progress on fundamental limits specifying the value of information gained by using such convex proxies for planning. This year we have published a complete analysis of VoI in IEEE Transactions on Information Theory [32] for the case of target detection and localization. In collaboration with co-PI John How, we have used this analysis to motivate a convex proxy approach to sensor planning for multi-class targets where successful detection of different classes of targets have different value payoffs relative to the mission. A conference paper on this multi-class convex proxy approach was published in a special

session on information-centric fusion in the proceedings of the IEEE Fusion conference [33] and a full paper on this approach has been accepted in the IEEE Trans on Signal Processing [40]. This work is the fruit of collaboration between MIT and UM.

5.3.2 Information-based approaches to sequential planning and exploitation

Contributors: John Fisher (MIT), Doug Cochran (ASU)
Publications: [36], [37], [38], [39].

Progress 18: Efficient Information Planning in Gaussian MRFs (Fisher MIT):

In the sequential setting, complexity issues arise when planning multiple time-steps ahead. Specifically, the complexity of active planning methods is combinatorial in the number of sensing actions and exponential in the planning horizon. A commonly utilized choice of information measure, mutual information (MI), is *submodular* when the measurements are statistically independent conditioned on the quantity of interest. As such, tractable greedy selection methods are guaranteed to be within a factor of the optimal (though, intractable) selection. Previous analysis provides guarantees for greedy selection for the more general case of inference in graphical models when measurements are divided into subsets with local constraints on subset selection and when the latent variable structure may not be fully specified a priori (e.g., inference in Markov chains for streaming data). It is important to emphasize that this analysis is about the basis for *planning* the sensing actions. Inference proceeds *after* having selected a plan. Consequently, a first step is to evaluate the reward of the prospective plan.

An important, often neglected, aspect of information-based approaches, however, is the computational cost of evaluating a given plan. While the bounds for greedy selection hold for *any* plan subject to the same constraints, one is free to reorder the sequence in which subsets are considered. Some reorderings have significantly higher information rewards than others. A simple example occurs in a Markov chain where at each node one may choose k out of N measurements. A naïve plan considers each node in order (greedily selecting k out of N available measurements at each node). Alternatively, one may consider nodes in random order selecting a single measurement (from those that have not already been selected), but ensuring each node is considered k times. Evaluating the information reward of the naïve plan has significantly lower computational complexity than the random plan, but the random plan will often have significantly higher information reward. Thus, there is motivation to expend computational resources for exploring multiple plans subject to the same constraints. Furthermore, when exploring multiple plans, the plan with lowest reward provides the lowest upper bound on the optimal plan yielding a tighter performance guarantee as compared to the greedy plan with highest reward.

This work [36, 37] considers the computational complexity of evaluating information rewards for measurement selection in Gaussian models. In such models, complexity depends on the number of latent variables, the number of measurements to be explored and the visitation order. We show speedups up to a thousand times by taking advantage of sparsity in the measurement process without changing the outcome of the greedy algorithm. In addition, we demonstrate that by working with the information form of Gaussian, we can provide the sufficient statistics at every step with much reduced computation. We achieve that by deploying a variant of belief propagation that is more suitable for adaptive inference settings. The results of the later technique are exact for Markov chains, trees, and poly-trees. This analysis is particularly useful for large-scale models, since the evaluation of information rewards poses a major computational bottleneck. Additionally, we demonstrate empirically that both the information reward and evaluation complexity are largely decoupled and as such, exploration of low-complexity walks yields high information rewards and tighter upper bounds.

Progress 19: 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. Last year's activity in this area 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 [39]. This year's work developed a theory of gauge-invariant data registration for networks that enables alignment of the local coordinate systems at individual sensors in the the network. This facilitates fusion of data across the network and enables multi-sensor detection and estimation algorithms that require appropriately aligned data.

The new developments marry the mathematical machinery of connections on principal bundles (gauge theory) with statistical estimation theory [38]. Sensor data may reside in data spaces that are naturally parameterized by nonlinear manifolds (e.g., measurements of the three-dimensional orientation of a target are parameterized by the special orthogonal group $SO(3)$), so the estimation component entails probability distributions on such structures (i.e., Lie groups). Estimators use sensor data that allow approximate local alignment between adjacent nodes in the network graph to be deduced. From this local data, they estimate *global* gauge transformations that will align all nodes in the network to a common coordinate system.

Framing this class of alignment problems in terms of statistical estimation theory enables the use of information-geometric quantities (e.g., Fisher information) and methods to analyze and bound algorithm performance in estimation of the desired gauge transformations. In particular, performance is seen to depend both on the accuracy with which alignment of coordinates *locally* at neighboring nodes in the network graph can be ascertained and on the *global* topology of the network graph. Quantifying how network topology affects performance provides insight into how the passing of information in a sensor network can be prioritized when communication resources are constrained or can be allocated: If one edge of the network graph must be sacrificed, which one will have least impact on global performance? If one link can be added, where will it have the most impact? Importantly, and in contrast to existing work in the context of communication networks, *value of information sharing* between nodes is measured here with respect to sensing objectives rather than measures of data throughput.

5.3.3 Human-in-the-loop distributed search

Contributors: Angela Yu UCSD, John Fisher MIT
Publications: [41], [42], [43], [44]

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. This year we report progress in computational models of human cognition for cooperative search problems.

Progress 20: Computational models of human cognition in cooperative search problems (Yu UCSD)

This continuing activity seeks to understand 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.

In the past year, we have continued to investigate, via modeling and behavioral experiments, the value of information in human cognition. We have focused on two main problems, multi-arm bandit task and multi-attribute preference choice. In the multi-arm bandit task, we find that humans do not utilize the computationally intense optimal algorithm, but rather adopt heuristic algorithms that effectively trade off informational gain (exploration) and immediate reward (exploitation). Intriguingly, we find that the best algorithm for capturing bandit behavior in young adults (university students) is the knowledge gradient algorithm [41], which greedily computes the value of exploration, but that older adult's behavior are best captured by the softmax algorithm [42], which stochastically deviates from optimal exploitation with no explicit representation of the value of exploration). It remains to be seen whether this age-related difference replicates, and, if so, why aging affects the representation and utilization of the value of information in humans in this way.

Separately, we have been analyzing human behavior in multi-attribute preference choice (e.g. buying a car as a function of safety and efficiency), where, curiously, the relative preference between two options has been shown to systematically depend on the presence/absence of a third option. We have developed and iteratively refined a Bayesian statistical inference model to demonstrate that this apparently irrational behavior actually naturally arises from humans using the available options as a source of information regarding the relative utilities of two attribute dimensions, and the unavailability of some options (sold out) as a source of social information with regard to the attribute values [45]. In the past year, we developed a better model to account for individual differences in this task [43]. Currently, we are trying to develop a model of active preference discovery, whereby we assume that consumption outcomes inform humans of their own intrinsic valuation of different attributes and therefore consumption choices reflect not only exploitative value but also incorporate a longer-term strategy of value exploration.

Progress 21: Boosting Crowdsourcing with Expert Labels: Local vs. Global Effects (Fisher MIT)

Crowdsourcing has emerged as a powerful approach for collecting data and information at large scales. The idea is to outsource tasks that are easy for humans but difficult for machines, sending them to online "crowd" workers who are given a relatively small incentive. Crowdsourcing has been widely used in many application and scientific domains, including machine learning, human-computer interaction and social forecasting, to name only a few.

A major challenge in crowdsourcing is quality control. The (often anonymous) crowd workers have unknown and highly diverse levels of expertise, making it a critical problem to evaluate workers' performance and optimally combine their labels. In addition, human judgments are inherently noisy, often with significant individual biases; this is especially common in the estimates of continuous quantities, such as probabilities, product prices, and point spreads in sports, where people tend to give under- or over-estimates based on their personal experience. In these cases, it is necessary to calibrate the crowdsourcing results by incorporating some ground truth information or accurate labels from domain experts.

Because the expert or true labels are often much more expensive than the labels from the crowd, this raises an important problem of understanding the values of these valuable resources and hence making optimal use of them. In this work [44], we study the optimal allocation of the true labels to best calibrate the crowd labels for estimating continuous quantities. We frame the problem into a minimization of a conditional variance criterion, and establish its monotonic submodularity, enabling efficient approximation via greedy selection. We observe that our greedy selection rule decomposes into two terms that reflect a trade-off between a local effect and a global effect, where the local effect encourages acquiring true labels from the most uncertain items, which improve the performance in a local, myopic fashion, while the global effect favors the most "influential" items, whose true labels provide valuable information for decreasing the uncertainty on their associated workers' performance, and significantly improve the prediction of all the other items via a snowball effect. We show that it is critical to consider the global effect when allocating the true labels, especially in the initial stage when the number of acquired true labels is small, and the uncertainty on the workers' performance is relatively large.

5.4 Related activities: DURIP Software Defined Radar Testbed

A collaborative DURIP grant for building X-band radars was awarded to OSU (PI 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) have been designed, constructed and field tested. These radars have been deployed at ARL Aberdeen proving grounds, as described below, and have been scheduled for use by MURI researchers beginning in August 2015. Here we report on progress on i) the software defined radar testbed, and ii)structured sensing matrix designs for high resolution radar, which can be deployed in future versions of the software defined radar testbed.

Progress 22: Software defined radar testbed (Ertin OSU)
Microwave radar systems are crucial components of any standoff sensor system due to their all-weather capabilities and proven performance for tracking, imaging, and situational awareness. MIMO radar systems which can transmit independent waveforms on multiple antennas have been suggested for improving detection, parameter estimation and clutter suppression capabilities. Under a DURIP grant we have built a collaborative research resource based on software defined radar (SDR) platforms that can adaptively modify both transmit waveforms and receive signal-processing tasks in real time [46]. We have focused development of a low power, short range versatile radar system that combines a high speed FPGA digital back-end with sideband digital/analog and analog/digital converters with a custom built RF Frontend. The key idea is software defined radar system is to sample the transmit/receive waveforms using high speed digital/analog and analog/digital converters and to implement key processing stages using programmable digital hardware. This collaborative research resource will be utilized by faculty and students of the Ohio State University, University of Michigan, Massachusetts Institute of Technology and Arizona State University. The testbed consists of 14 Micro SDR Platforms with 2 transmit and 1 receive antennas. We have participated in a recent data collection effort by ARL at the Aberdeen Proving Grounds in July 2015 for characterization of multimodal signature personnel and civilian vehicles. In Fall 2015 we plan to conduct multiple data collection campaigns to support demonstration of algorithms and VOI theory developed under the MURI.

Progress 23: Structured sensing matrix designs for high resolution radar (Ertin OSU)
Radar imaging systems transmit modulated wide- band waveform to achieve high range resolution resulting in high sampling rates at the receiver proportional to the bandwidth of the transmit waveform. Analog processing techniques can be used on receive to reduce the number of measurements to N, the number of potential delay bins. If the scene interrogated by the radar is assumed to be sparse consisting of K point targets, results from compressive sensing suggest that number of measurements can be further reduced to scale with $K \log N$ for stable recovery of a sparse scene from measurements with additive noise. While unstructured random projectors guarantee successful recovery under sparsity constraints, they cannot be implemented in the radar hardware in practice. Recently, structured random Toeplitz and Circulant matrices that result from using stochastic wave- forms in time delay estimation setting have been shown to yield recovery guarantees similar to unstructured sensing matrices. However, the corresponding transmitter and receiver structures have high complexity and large storage requirements. In our recent work [47], we propose an alternative low complexity compressive wideband radar sensor which combines multitone signal chirp waveform on transmit with a receiver that utilizes an analog mixer followed with a uniform sub-Nyquist sampling stage. We derive the recovery guarantees for the resulting structured measurement matrix and sufficient conditions for the number of tones. The only random component of our design is the sparse tone spectrum implementable efficiently in hardware. Our analytical and empirical results show that the performance of our scheme is in par with unstructured random sensing matrices and structured Toeplitz and Circulant matrices with random entries. While the proposed method for sensing matrix design is offline, our results establish a foundation from which to investigate active and adaptive designs based on structure assumptions.

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.

As described above, key elements of a fundamental theory for value-of-information for adaptive sensing, distributed fusion, and information exploitation systems are being developed. Mathematical bounds and limits on the VoI have been established for several important problems in learning, fusion and control for adaptive sensing. These results have been used to develop algorithms that outperform the state-of-the-art.

In the last year of this MURI, several of the co-PIs plan to demonstrate our progress on VoI-driven algorithms and theory using the radar testbed. The testbed has been used to collect data for vehicle and dismount signatures that can supplement other data that we have been using to validate our algorithms, e.g., the ARL footstep data. The testbed will be used to collect data on dynamic interactive multi-agent systems that will allow validation of models for target search and classification, predicting human behavior from partially observed data, signal subspace detection and information fusion, information sharing in emulated decentralized sensor networks, and value-of-information driven fusion and sensor planning that accounts for mission-dependent rewards, among other models.

7 Publications

[1] Georgios Papachristoudis and John W. Fisher III, "Adaptive Belief Propagation," in *International Conference on Machine Learning (ICML)*, July 2015.

[2] Visar Berisha and Alfred O. Hero, "Empirical non-parametric estimation of the Fisher information," *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 988–992, 2015.

[3] Robert Nishihara, Laurent Lessard, Benjamin Recht, Andrew Packard, and Michael I. Jordan, "A general analysis of the convergence of ADMM," *arXiv preprint arXiv:1502.02009*, 2015.

[4] Stefano Soatto, "Visual scene representations: Sufficiency, minimality, invariance and approximations," *Proc. of the Intl. Conf. on Learning Representations; ArXiv: 1411.7676, arXiv preprint arXiv:1411.7676*, 2015.

[5] J. Dong, J. Hernandez, J. Balzer, D. Davis, and S. Soatto, "Multiview feature engineering and learning," in *Proc. of the IEEE Intl. Conf. on Comp. Vis. and Patt. Recog. (CVPR)*, 2015.

[6] J. Dong and S. Soatto, "Domain size pooling in local descriptors: DSP-SIFT," in *Proc. of the IEEE Intl. Conf. on Comp. Vis. and Patt. Recog. (CVPR)*, 2015.

[7] G. Georgiadis, A. Chiuso, and S. Soatto, "Texture representations for image and video texture synthesis," in *Proc. of the IEEE Intl. Conf. on Comp. Vis. and Patt. Recog. (CVPR)*, 2015.

[8] Byung-Woo Hong and Stefano Soatto, "Shape matching using multiscale integral invariants," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 1, pp. 151–160, 2015.

[9] J. Balzer, D. Acevedo-Feliz, S. Soatto, S. Hofer, M. Hadwiger, and J. Beyerer, "Cavlectometry: Towards holistic reconstruction of large mirror objects," *Proc. of 3DVision*, 2014.

[10] J. Bruna and S. Mallat, "Classification with scattering operators," *arXiv preprint arXiv:1011.3023*, 2010.

[11] Fabio Anselmi, Lorenzo Rosasco, and Tomaso Poggio, "On invariance and selectivity in representation learning," *arXiv preprint arXiv:1503.05938*, 2015.

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

[13] T. Abdelrahman and E. Ertin, "Mixture of factor analyzers models of appearance manifolds for resolved SAR targets," in *Proceedings of SPIE 9475, Algorithms for Synthetic Aperture Radar Imagery XXII*, May 2015.

[14] 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*, vol. 63, no. 1, 2015.

[15] G. T. Whipps, E. Ertin, and R. L. Moses, "Distributed sensing for quickest change detection of point radiation sources," in *International Conference on Information Fusion (FUSION)*, July 2015.

[16] Theodoros Tsiligkaridis, Brian M. Sadler, and Alfred O. Hero III, "On decentralized estimation with active queries," *IEEE Transactions on Signal Processing*, vol. 63, no. 10, pp. 2610–2622, 2015.

[17] Theodoros Tsiligkaridis, Brian M. Sadler, and Alfred O. Hero, "Collaborative 20 questions for target localization," *IEEE Transactions on Information Theory*, vol. 60, no. 4, pp. 2233–2252, 2014.

[18] Himanshu Nayar and R. R. Nadakuditi, "OptFuse: Low-rank factor estimation by optimal data-driven linear fusion of multiple signal-plus-noise matrices," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, July 2015.

[19] Raj Tejas Suryaprakash, Brian E. Moore, and Raj Rao Nadakuditi, "Algorithms and performance analysis for estimation of low-rank matrices with kronecker structured singular vectors," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 3776–3780.

[20] Sandeep Gogineni, Pawan Setlur, Muralidhar Rangaswamy, and Raj Rao Nadakuditi, "Random matrix theory inspired passive bistatic radar detection with noisy reference signal," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 2754–2758.

[21] S. Gogineni, P. Setlur, M. Rangaswamy, and R. R. Nadakuditi, "Random matrix theory inspired passive bistatic radar detection of low-rank signals," in *IEEE Radar Conference (RadarCon)*, May 2015, pp. 1656–1659.

[22] Raj Tejas Suryaprakash and Raj Rao Nadakuditi, "The performance of MUSIC in white noise with limited samples and missing data," in *IEEE Radar Conference*, 2014, pp. 0940–0944.

[23] Raj Tejas Suryaprakash and Rajesh Nadakuditi, "Consistency and MSE performance of MUSIC-based DOA of a single source in white noise with randomly missing data," *IEEE Transactions on Signal Processing*, 2015.

[24] Yuchen Zhang, Martin J. Wainwright, and Michael I. Jordan, "Distributed estimation of generalized matrix rank: Efficient algorithms and lower bounds," *To appear in Intl Conf on Machine Learning (ICML) and on arXiv preprint arXiv:1502.01403*, 2015.

[25] 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.

[26] H. Nayar and R. R. Nadakuditi, "Theoretical performance analysis of Tucker higher order SVD in extracting structure from multiple signal-plus-noise matrices," in *Proceedings of the Asilomar Conference on Signals, Systems and Computers*, Nov 2014, pp. 755–759.

[27] Tianpei Xie, Nasser M Nasrabadi, and Alfred O Hero III, "Semi-supervised multi-sensor classification via consensus-based multi-view maximum entropy discrimination," in *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 1936–1940.

[28] Diyan Teng and Emre Ertin, "Learning density ratios for sequential inferences," in *Information Theory and Applications Workshop*, San Diego, CA, 2015.

[29] Diyan Teng and Emre Ertin, "Learning to aggregate information for sequential inferences," in *submitted to Neural Information Processing Systems (in review)*, 2015.

[30] Tianpei Xie, Nasser M. Nasrabadi, and Alfred O. Hero, "Learning to classify with possible sensor failures," in *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 2395–2399.

[31] Gregory Newstadt, Dennis Wei, and Alfred O. Hero, "Adaptive search and tracking of sparse dynamic targets under resource constraints," *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2321–2335, 2015.

[32] Dennis Wei and Alfred O Hero, "Performance guarantees for adaptive estimation of sparse signals," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 2043–2059, 2015.

[33] Beipeng Mu, Gregory E. Newstadt, Dennis L. Wei, Alfred O. Hero, and Jonathan P. How, "Adaptive search for multi-class targets with heterogeneous importance," in *International Conference on Information Fusion*, Washington D.C., July 2015.

[34] Robert Nishihara, Stefanie Jegelka, and Michael I. Jordan, "On the convergence rate of decomposable submodular function minimization," in *Advances in Neural Information Processing Systems*, 2014, pp. 640–648.

[35] Beipeng Mu, Ali akbar Agha-mohammadi, Liam Paull, Matthew Graham, Jonathan How, and John Leonard, "Two-stage focused inference for resource-constrained collision-free navigation," in *Proceedings of Robotics: Science and Systems*, Rome, Italy, July 2015.

[36] Georgios Papachristoudis and John W. Fisher III, "On the Complexity of Information Planning in Gaussian Models," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, April 2015.

[37] Georgios Papachristoudis and John W. Fisher III, "Efficient Information Planning in Gaussian MRFs," in *International Conference on Information Fusion (Fusion)*, July 2015.

[38] S. D. Howard, D. Cochran, and W. Moran, "Gauge-invariant registration in networks," in *Proceedings of the International Conference on Information Fusion*, July 2015.

[39] Lauren Crider and Douglas 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*, November 2014, pp. 465–469.

[40] Gregory E. Newstadt, Beipeng Mu, Dennis Wei, Jonathan P. How, and Alfred O. Hero III, "Resource-constrained adaptive search for sparse multi-class targets with varying importance," *to appear in IEEE Trans on Signal Processing, arXiv preprint arXiv:1409.7808*, 2014.

[41] 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.

[42] K. M. Harlé, S. Zhang, M. Schiff, S. Mackey, M. P. Paulus*, and A. Yu*, "Altered belief formation and reward-maximization strategy in amphetamine dependence: evidence from a bayesian modeling approach," *Frontiers in Psychology*, 2015 (under review), *Paulus and Yu are co-senior authors.

[43] S. Ahmad and A. J. Yu, "A rational model for individual differences in preference choice," *Proceedings of the Cognitive Science Society Conference*, 2015.

[44] Georgios Papachristoudis and John W. Fisher III, "Boosting Crowdsourcing with Expert Labels: Local vs. Global Effects," in *International Conference on Information Fusion (Fusion)*, July 2015.

[45] P. Shenoy and A. J. Yu, "A rational account of contextual effects in preference choice: What makes for a bargain?," *Proceedings of the Thirty-Fifth Annual Conference of the Cognitive Science Society*, 2013.

[46] S. Baskar and E. Ertin, "A software defined radar platform for waveform adaptive MIMO radar research," in *IEEE Radar Conference (RadarCon)*, May 2015, pp. 1590–1594.

[47] N. Sugavanam and E. Ertin, "Recovery guarantees for multifrequency chirp waveforms in compressed radar sensing," *submitted to IEEE JSTSP*, Jul 2015.

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 (10)

      1. Hong, Byung-Woo, and Stefano Soatto. "Shape matching using multiscale integral invariants." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 1 (2015): 151-160.
      2. Raj Tejas Suryaprakash and Rajesh Nadakuditi, "Consistency and MSE performance of MUSIC-based DOA of a single source in white noise with randomly missing data," *IEEE Transactions on Signal Processing*, vol. 63, no. 18, pp. 4756–4770, September 2015.
      3. T. Tsiligkaridis, B. M. Sadler and A. O. Hero III, "On decentralized estimation with active queries," *IEEE Transactions on Signal Processing*, vol. 63, no. 10, pp. 2610-2622, May 2015. Also available as arxiv:1312.7848.
      4. G. Newstadt, D. Wei, A.O. Hero, "Adaptive Search and Tracking of Sparse Dynamic Targets under Resource Constraints," *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2321-2335, May 2015. Availalable as arXiv:1404.2201
      5. V. Berisha and A. O. Hero, "Empirical non-parametric estimation of the Fisher Information," *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 988-992, July 2015. Available as arxiv 1408.1182
      6. D. Wei and A. O. Hero, "Performance Guarantees for Adaptive Estimation of Sparse Signals," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 2043-2059, April 2015. Available as arxiv:1311.6360
      7. K. J. Hsiao, J. Calder and A. O. Hero, "Pareto-depth for Multiple-query Image Retrieval," *IEEE Transactions on Image Processing*, vol. 24, no. 2, pp. 583-594, Feb. 2015. Available as arxiv:1402.5176
      8. Z. Meng, D. Wei, A. Wiesel, A. Hero, "Distributed Learning of Gaussian Graphical Models via Marginal Likelihoods," *IEEE Transactions on Signal Processing*, vol. 62, no. 20, pp. 5425-5438, Oct. 2014. Available as arxiv:1303.2378
      9. 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, Apr 2014. Available as arXiv:1306.1922
      10. 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*, vol. 63, no. 1, 2015.

   b. Papers in non-peer-reviewed vehicles (3)

      1. Kristjan Greenewald, Edmund Zelnio, Alfred O. Hero III, "Kronecker PCA Based Robust SAR STAP," arxiv 1502:07481, Jan 2015.
      2. Yoann Altmann, Steve McLaughlin, Alfred Hero, "Robust Linear Spectral Unmixing using Anomaly Detection," arxiv 1501.03731, Jan 2015.

3. Visar Berisha, Alan Wisler, Alfred O. Hero, and Andreas Spanias, "Empirically Estimable Classification Bounds Based on a New Divergence Measure," arxiv 1412.6534, Jan 2015.

c. Presentations (6)

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

1. D. Teng and E. Ertin, "Learning density ratios for sequential inferences, presentation in Information Theory and Applications Workshop, (San Diego, CA), Feb 2015.
2. Alfred Hero, "Signal Processing for graphs," UM CSL Communications and Signal Processing Seminar, Sept 2014.
3. Alfred Hero, "Correlation mining in high dimension with limited samples." UCLA Dept ECE Distinguished Seminar, May 2015.
4. Alfred Hero, "Cooperative localization in networks," tutorial given at the School of ICASSP, Intl Conf on Acoust., Speech, and Signal Processing, Brisbane Australia, May 2015.
5. Alfred Hero, "Large scale correlation mining," Duke workshop on sensing and analysis of high dimensional data, Duke Univ., Durham July 29, 2015
6. Beipeng Mu, "Focused Information Gathering and an application in SLAM," MIT-LIDS student conference, 2015 (co-PI How's student).

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

iii. Peer-Reviewed Conference Proceedings publications (24)

1. R. Nishihara, L. Lessard, B. Recht, A. Packard and M. I. Jordan. "A general analysis of the convergence of ADMM." International Conference on Machine Learning (ICML), 2015.
2. Y. Zhang, M. J. Wainwright, and M. I. Jordan, "Distributed estimation of generalized matrix rank: Efficient algorithms and lower bounds." International Conference on Machine Learning (ICML), 2015.
3. G. Papachristoudis and J. W. Fisher III, "On the Complexity of Information Planning in Gaussian Models," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2015.
4. G. Papachristoudis and J. W. Fisher III, "Efficient Information Planning in Gaussian MRFs," *International Conference on Information Fusion (Fusion)*, 2015.
5. G. Papachristoudis and J. W. Fisher III, "Adaptive Belief Propagation," *International Conference on Machine Learning (ICML)*, 2015.
6. Q. Liu, A. Ihler, and J. W. Fisher III; "Boosting Crowdsourcing with Expert Labels: Local vs. Global Effects," *International Conference on Information Fusion (Fusion)*, 2015.
7. Himanshu Nayar and R. R. Nadakuditi,"OptFuse: low-rank factor estimation by optimal data-driven linear fusion of multiple signal-plus-noise matrices," Proceedings of the International Conference on Information Fusion, June 2015.
8. Raj Tejas Suryaprakash, Brian E. Moore, and Raj Rao Nadakuditi, "Algorithms and performance analysis for estimation of low-rank matrices with Kronecker structured singular vectors," Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), July 2015. pp. 3776–3780.

9. Sandeep Gogineni, Pawan Setlur, Muralidhar Rangaswamy, and Raj Rao Nadakuditi, "Random matrix theory inspired passive bistatic radar detection with noisy reference signal," Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), July 2015, pp. 2754–2758.
10. S. Gogineni, P. Setlur, M. Rangaswamy, and R. R. Nadakuditi, "Random matrix theory inspired passive bistatic radar detection of low-rank signals," *IEEE Radar Conference (RadarCon)*, May 2015, pp. 1656–1659.
11. H. Nayar and R.R. Nadakuditi, "Theoretical performance analysis of Tucker higher order SVD in extracting structure from multiple signal-plus-noise matrices," *Proceedings of the Asilomar Conference on Signals, Systems and Computers*, Nov 2014, pp. 755–759.
12. Raj Tejas Suryaprakash and Raj Rao Nadakuditi, "The performance of MUSIC in white noise with limited samples and missing data," *IEEE Radar Conference*, 2014, pp. 0940–0944.
13. Ahmad, S. and Yu, A. J., "A rational model for individual differences in preference choice," *Proceedings of the Thirty-Seventh Conference of the Cognitive Science Society*, 2015.
14. Soatto, Stefano, and Alessandro Chiuso "Visual scene representations: Sufficiency, minimality, invariance and approximations." Proc. of the Intl. Conf. on Learning Representations, June 2014. Available as arXiv:1411.7676. (2014).
15. G. Georgiadis, A. Chiuso and S. Soatto, "Texture representations for image and video texture synthesis," Proc. of the IEEE Intl. Conf. on Comp. Vis. and Patt. Recog. (CVPR), 2015.
16. J. Dong and S. Soatto, "Domain size pooling in local descriptors: DSP-SIFT," Proc. of the IEEE Intl. Conf. on Comp. Vis. and Patt. Recog. (CVPR), 2015.
17. J. Dong, J. Hernandez, J. Balzer, D. Davis and S. Soatto, "Multiview feature engineering and learning," Proc. of the IEEE Intl. Conf. on Comp. Vis. and Patt. Recog. (CVPR), 2015.
18. J. Balzer, D. Acevedo-Feliz, S. Soatto, S. Hofer, M. Hadwiger and J. Beyerer, "Cavlectometry: Towards holistic reconstruction of large mirror objects," Proc. of 3DVision, 2014.
19. B. Mu, G. Newstadt, D. Wei, A. O. Hero, and J. P. How, "Adaptive Search for Multi-class Targets with Heterogeneous Importance," Fusion 2015, Washington D.C. Available as arxiv 1409.7808.
20. Yoann Altmann, Steve McLaughlin, and Alfred O. Hero, "Robust linear spectral unmixing using outlier detection," IEEE Intl Conf on Acoustics, Speech, and Signal Processing (ICASSP), Brisbane, April 2015. Available as arxiv 1501.03731.
21. Goran Marjanovic, Magnus Ulfarsson, and Alfred O. Hero, "MIST: L0 sparse linear regression with momentum," IEEE Intl Conf on Acoustics, Speech, and Signal Processing (ICASSP), Brisbane, April 2015. Available as arxiv:1409.7193.
22. Tianpei Xie, Nasser Nasrabadi, Alfred O. Hero, "Multi-sensor classification via consensus-based multi-view maximum entropy discrimination," IEEE Intl Conf on Acoustics, Speech, and Signal Processing (ICASSP), Brisbane, April 2015. Available as arxiv 1507.01269.

23. S. D. Howard, D. Cochran, and W. Moran, "Gauge-invariant registration in networks," *Proceedings of the International Conference on Information Fusion*, July 2015.
24. Lauren Crider and Douglas Cochran, "Effects of network topology on the conditional distributions of surrogated generalized coherence estimates," *Proceedings of the Asilomar Conference on Signals, Systems, and Computers*, pp. 465–469, November 2014.

d. Manuscripts (6)

1. J. C. Duchi, M. I. Jordan, M. J. Wainwright, and A. Wibisono, "Optimal rates for zero-order optimization: the power of two function evaluations." IEEE Transactions on Information Theory, to appear.
2. Harlé, K M, Zhang, S, Schiff, M, Mackey, S, Paulus, M P, and Yu, A J., "Altered belief formation and reward-maximization strategy in amphetamine dependence: evidence from a Bayesian modeling approach," Frontiers in Psychology, in review.
3. Kristjan Greenewald, Edmund Zelnio, and Alfred O. Hero III, "Kronecker PCA Based Robust SAR STAP," arxiv 1502:07481, Jan 2015. Submitted.
4. Yoann Altmann, Steve McLaughlin, and Alfred Hero, "Robust Linear Spectral Unmixing using Anomaly Detection," arxiv 1501.03731, Jan 2015. To appear in IEEE Trans on Computational Imaging.
5. Visar Berisha, Alan Wisler, Alfred O. Hero, and Andreas Spanias, "Empirically Estimable Classification Bounds Based on a New Divergence Measure," arxiv 1412.6534, Jan 2015. In review.
6. N. Sugavanam and E. Ertin, "Recovery guarantees for multifrequency chirp waveforms in compressed radar sensing," submitted to IEEE JSTSP, (in review) July 2015.

e. Books (2)

1. A. O. Hero, "Sparsity regularized image reconstruction," in *Review of Progress in Quantitative Nondestructive Evaluation*, Vol. 34, edited by Dale E. Chimenti and L. J. Bond, published by American Institute of Physics, Melville, NY, 2015. This paper supports Hero' keynote address given at QNDE 2014 in Boise ID.
2. S. Kumar, M. Al'Absi, G. Beck, E. Ertin, and M. Scott, "Behavioral Monitoring and Assessment via Mobile Sensing Technologies," Book Chapter in *Behavioral Health Care and Technology Using Science-Based Innovations to Transform Practice*, edited by Lisa Marsch, Sarah Lord and Jesse Dallery, Oxford University Press, Dec 2014.

f. Honors and Awards

i. Alfred Hero, Keynote speaker, IEEE International Telecommunications Symposium, Sao Paolo Brazil, Aug 2014.
ii. Alfred Hero, Plenary speaker, IEEE Intl. Conference on Image Processing, Paris France. Oct 2014.
iii. Alfred Hero, Keynote speaker, Scale Space and Variational Methods in Computer Vision Conference, Lege Cap Ferrat, France June 2015.
iv. Michael Jordan, John von Neumann Lecture, Brown University.
v. Michael Jordan, Coxeter Lecture Series, Fields Institute for Research in Mathematical Sciences.

vi. Michael Jordan, Bahadur Memorial Lecture, University of Chicago.
vii. Michael Jordan, Keynote speaker, 34th ACM Symposium on Principles of Database Systems (PODS).
viii. Michael Jordan, Keynote speaker, International Conference on Computing, Networking and Communications (ICNC 2015).
ix. Michael Jordan, Keynote speaker, Hadoop/Strata Conference.
x. Michael Jordan, Keynote speaker, Stanford/Berkeley Robotics Symposium,
xi. Michael Jordan, Keynote speaker, Artificial Intelligence and Statistics (AISTATS).
xii. Michael Jordan, Keynote speaker, Statistical Society of Canada Annual Meeting.
xiii. Michael Jordan, Keynote speaker, Computational Learning Theory Annual Conference (COLT).
xiv. Michael Jordan, Keynote speaker, International Conference on Machine Learning (ICML).
xv. R. Moses, Sparse Methods in Radar Signal Processing, Plenary Keynote Lecture, Sensor Signal Processing for Defence Conference, Edinburgh, Scotland, Sep 2014.
xvi. Jon How was appointed Editor in Chief of the IEEE Control Systems Magazine
xvii. Stefano Soatto, Best Conference Paper Award, Intl. Conf. on Robotics and Automation (ICRA), 2015
xviii. Raj Rao Nadakuditi, DARPA Young Investigator Award 2014.
xix. Doug Cochran, Fulbright Distinguished Chair in Science and Technology, 2015-2016.

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 (21)

(a) ASU student Lauren Crider supported at 33% annualized FTE (since beginning PhD study in May 2015)
(b) ASU student Shih-Ling Phuong supported at 5% annualized FTE (100
(c) ASU student Kaitlyn Beaudet supported at 25% FTE
(d) MIT student Beipeng Mu supported at 50% annualized FTE
(e) MIT student Georgios Paperchristoudis at 50% annualized FTE
(f) MIT student Christopher Dean at 50% annualized FTE
(g) MIT student Randi Cabezas at 50% annualized FTE
(h) MIT student Julian Straub at 50% annualized FTE
(i) OSU student Nithin Sugavanam supported at 50% annualized FTE
(j) OSU student Diyan Teng supported at 50% annualized FTE
(k) 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)

   (l) UC Berkeley student Yuchen Zhang supported at 50% annualized FTE
   (m) UC Berkeley student Lihua Lei supported at 50% annualized FTE
   (n) UCLA student Georgios Georgiadis supported at 10% annualized FTE
   (o) UCLA student Vasiliy Karasev supported at 5% annualized FTE
   (p) UCSD student Sheeraz Ahmad supported at 50% annualized FTE
   (q) UM student Zhaoshi Meng supported at 25% GSRA Aug 2014
   (r) UM student Tianpei Xie: 50% Fall 2014 + Winter 2015 + SS 2015
   (s) UM student Pin-Yu Chen: 50% Fall 2014
   (t) UM student Brandon Oselio 50% GSRA May - June 12, 2015
   (u) UM student Himanshu Nayyar 50% GSRA SS 2015

2. Masters Students (2)
   (a.) ASU student Kaitlyn Beaudet supported at 10% annualized FTE
   (b.) UM student Alfredo Bravo Iniguez: 25% annualized FTE

b. Post Doctorates (10)
   1. Virginia Estellers, UCLA, 50% annualized FTE
   2. Shunan Zhang, UCSD, 50% annualized FTE
   3. Goran Marjonovic, UM, 4% annualized FTE
   4. Jie Chen 50%, UM, 8% annualized FTE
   5. Hye Won Chung, UM, 100% annualized FTE
   6. Taposh Banerjee, UM, 33% annualized FTE
   7. Yasin Yilmaz, UM, 33% annualized FTE
   8. Oren Freifeld, MIT, 25% annualized FTE
   9. Qiang Liu, MIT, 40% annualized FTE
   10. Guy Rosman, MIT, 25% annualized FTE

c. Faculty (10)
   1. Stefano Soatto, UCLA (10% FTE)
   2. Douglas Cochran, ASU (5% FTE)
   3. Emre Ertin, OSU (20% FTE)
   4. Randy Moses, OSU (0% FTE)
   5. John Fisher, MIT (20% FTE)
   6. Alfred Hero, UM (0% FTE)
   7. Jonathan How, MIT (0%)
   8. Michael Jordan, UC Berkeley (0%)
   9. Raj Nadakaduti, UM (8% FTE)
   10. Angela Yu, UCSD (8% FTE)

d. Undergraduate Students (2)
   1. ASU undergraduate Lauren Crider supported at 25% FTE
   2. Derek Allman 20% FTE

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

f. Masters Degrees Awarded (3)
   1. Alfredo Bravo Iniguez, M.S., UM
   2. Kaitlyn Beaudet, M.S., ASU
   3. Christoper Dean, S.M., MIT

g. Ph.D.s Awarded (3)
   1. Zhaoshi Meng, UM
   2. Nick Asendorf, UM
   3. Giorgos Papachristoudis, MIT

h. Other Research Staff (3)
   1. Taco Cohen, UCLA
   2. Gottfried Graber, UCLA
   3. Massimo Cairo, UCLA

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) UM student Brandon Oselio did an internship at ARL on dynamic social media analysis under Lance Kaplan in summer 2014 and 2015.
      (b) UM student Brandon Oselio did an internship on anomaly detection in social media at MIT Lincoln Laboratory under mentorship of Kevin Carter in summer 2015.
      (c) UM student Tianpei Xie did an internship on robust fusion at ARL in summers of 2013, 2014 and 2015 under mentorship of Nasser Nasrabadi.
      (d) UM student Himanshu Nayar did an internship at Army Research Lab with Dr. Ananthram Swami and Dr. Terence Moore in summer 2015.
      (e) ASU student Davis Gilton did an internship at AFRL in summer 2015 with a project on SAR.
      (f) ASU student Theresa Scarnati did an internship at AFRL in summer 2015 with a project on geolocation.
   3. Co-PI interactions with ARL and other Federal research labs
      (a) co-PI Ertin has participated in a recent data collection effort by ARL at the Aberdeen Proving Grounds in July 2015 for characterization of multimodal signatures of personnel and civilian vehicles.
      (b) PI Hero and co-PIs Cochran, Ertin, and Fisher visited ARL in July 2015 where they briefed ARL researchers on research under this grant and engaged in technical discussions. ARL researchers engaged in this visit included Brian Sadler, and Nasser Nasrabadi.

   (c) PI Hero and co-PIs Cochran, Ertin and Fisher interacted with ARL researchers Lance Kaplan and Tien Pham at the IEEE Fusion meeting in Washington DC in July 2015.
   (d) PI Hero and co-PI Fisher interacted with LANL researchers during their visit to Los Alamos National Laboratory in Jan 2015.
   (e) PI Hero was one of 5 invited panelists for the AFOSR Workshop Challenges in Fusion, Dayton OH Aug. 2014.
   (f) co-PI Ertin has applied performance prediction theory developed under this MURI, for algorithm development to assess information value of Radar/Ladar signatures in ATR tasks under the AFRL/OPERA program under a subcontract to Leidos, Inc.

4. Other relevant co-PI activities on national committees
   (a) PI Hero is a member of the US National Academies of Sciences, Engineering, and Medicine Intelligence Science and Technology Group (ISTEG) (2015-).
   (b) PI Hero serves on the National Academy of Sciences Committee on Applied and Theoretical Statistics (2011-present)
   (c) PI Hero serves on the DARPA Biosynchronicity Grand Challenge committee (2015-)
   (d) Co-PI Moses serves on the National Academy of Sciences Panel on Information Science at the Army Research Laboratory (2011-present).
   (e) R. Moses serves on the Board of Directors for the American Society of Engineering Education (ASEE) and as Chair of the ASEE Engineering Research Council, 2014-present.
   (f) co-PIs Hero and Fisher co-organized a special session on Value centered fusion at the IEEE Fusion conference that took place in Washington DC in July 2015.
   (g) Co-PI Soatto is Program Co-Chair, SIAM Conference on Imaging Science, 2016 (with Rebecca Willett)

5. Co-PI Cochran organized a workshop on passive radar at AFRL in September 2014.