

REPORT DOCUMENTATION PAGE					Form Approved OMB No. 0704-0188	
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Service Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.						
1. REPORT DATE (DD-MM-YYYY) 01/23/2012		2. REPORT TYPE Final		3. DATES COVERED (From - To) 06/01/2009-11/30/2011		
4. TITLE AND SUBTITLE Performance-driven multimodality sensor fusion				5a. CONTRACT NUMBER		
				5b. GRANT NUMBER FA9550-09-1-0471		
				5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) Professor Alfred O.Hero				5d. PROJECT NUMBER		
				5e. TASK NUMBER		
				5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Regents of the University of Michigan, 3003 S. State Street, Ann Arbor, MI 48109-1274				8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research 875 North Randolph St. Room 3112 Arlington, VA 22203				10. SPONSOR/MONITOR'S ACRONYM(S) USAF AFOSR/NL		
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION/AVAILABILITY STATEMENT Distribution Unlimited						
13. SUPPLEMENTARY NOTES						
14. ABSTRACT The broad objective of this grant was to develop a generally applicable theory of performance of information-level fusion that * provides accurate prediction of post-fusion algorithm accuracy in uncertain environments. * determines factors affecting fundamental performance tradeoffs, e.g., sample size, resolution, specificity, and sensitivity of sensors. * specifies performance benchmarks allowing quantitative comparison of different fusion algorithms. * provides guidelines for algorithm design and optimization. The effort focused on information theoretic fusion methods and our analysis was based on geometric properties of information. Our research has impacted application domains where information theoretic fusion is applied. These included georegistration, remote sensing, multimodality anomaly detection, visualization, and dimensionality reduction.						
15. SUBJECT TERMS georegistration, remote sensing, multimodality anomaly detection, visualization, and dimensionality reduction.						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT		18. NUMBER OF PAGES	
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified	UU		1	
					19a. NAME OF RESPONSIBLE PERSON A. O. Hero	
					19b. TELEPHONE NUMBER (Include area code) 734-763-0564	

AFOSR Grant: Performance-driven multimodality sensor fusion

AFOSR Grant Number: FA9550-09-1-0471
Grant period: June 1, 2009 to Nov 31 2011

Final Report

Co-PIs:

Alfred O Hero, III (PI)
University of Michigan
Ann Arbor, MI 48109-2122
hero@umich.edu

Raviv Raich (co-PI)
Oregon State University
Corvallis, OR 97331
raich@eecs.oregonstate.edu

1 Project overview

The broad objective of this grant was to develop a generally applicable theory of performance of information-level fusion that

- provides accurate prediction of post-fusion algorithm accuracy in uncertain environments.
- determines factors affecting fundamental performance tradeoffs, e.g., sample size, resolution, specificity, and sensitivity of sensors.
- specifies performance benchmarks allowing quantitative comparison of different fusion algorithms.
- provides guidelines for algorithm design and optimization.

The effort focused on information theoretic fusion methods and our analysis was based on geometric properties of information. Our research has impacted application domains where information theoretic fusion is applied. These included georegistration, remote sensing, multimodality anomaly detection, visualization, and dimensionality reduction.

2 Technical accomplishments

Our technical accomplishments are reported in several papers, submitted or appeared, cited in the citation references below. These papers make the following contributions:

1. We have obtained the (to-date) sharpest asymptotic expressions for estimator bias, variance and a CLT for a wide class of information divergence estimators [6], [7]. The utility of these expressions is that they can be used to optimize over tuning parameters of the fusion criterion, thereby circumventing the need for manual parameter tuning. This theory has been applied to non-parametric estimation of the mutual information [8] and estimation of intrinsic dimension [9].
2. A new framework was introduced, involving boundary compensation and weighted estimation, to significantly enhance information divergence estimator performance [10], [11]. These estimators have provably better performance than state-of-the-art divergence estimators proposed by Leonenko, Barishnikov and others.
3. A new framework for entropy estimation using maximum entropy principles was introduced leading to an upper bound on the true fusion criterion (entropy and relative entropy) [12]. When the boundary is known, this maximum entropy method is competitive with the non-parametric methods of entropy estimation discussed above.
4. A new approach was proposed for estimating parameters of non-parametric topic models [13, 3, 1]. Topic models are useful for text data and other discrete "soft information" sources. The new method merges soft and hard information. The confidence constraint approach was also considered in the general context of multiple observation setup [2].

Each of these accomplishments is briefly described in the following paragraphs. A summary of the research accomplishments in non-parametric entropic fusion methods (first two bullets above) was published as part of the proceedings of the 2011 Defense Applications of Signal Processing Workshop held in Coolumb Australia [15].

2.1 Expressions for divergence estimator bias, variance and a CLT

The key to performance-driven fusion is an accurate theory of performance that can be used to identify the important underlying factors. Our focus has been on deriving asymptotic bias and mean-squared error (MSE) in the regime of large sample size for estimators of information divergences between feature distributions. Information divergences are used as objective functions that are minimized or maximized during the process of image registration, blind deconvolution, source separation, model selection and other algorithms relevant to fusion. For a broad class of density-plug-in estimators, that includes the common kernel density and k nearest neighbor (kNN) plug-in estimators, we have developed a generally applicable theory that gives analytical closed-form expressions for asymptotic bias and MSE in terms of the sample size, the dimension

of the feature space, and the underlying feature probability distribution. These results appear in the technical report [6] co-authored by co-PI's Hero and Raich and the supported University of Michigan graduate student Kumar Sricharan. This report incorporates comparisons to state of the art divergence and entropy estimation algorithms and is an extension of the report cited in last year's progress report. A paper that covers the basic convergence results was submitted to the IEEE Transactions on Information Theory and is currently under revision. The asymptotic theory developed in [6] was applied to anomaly detection in [16]. This has led to the fastest and most reliable anomaly detection method to date and can be applicable to large datasets with millions of samples. The asymptotic theory was applied to intrinsic dimension estimation in [9], which was implemented for fusion and segmentation of hyperspectral imagery in [15]. The theory was also applied to high dimension correlation screening [17].

2.2 Weighted and boundary compensated divergence estimators

The fusion criteria that are studied in this project are all derived from entropy functions of the underlying probability density of features of the data. These features are arbitrary and in any fusion application the probability densities are unknown. To be implemented for fusion these criteria need to be determined accurately from data and we developed estimators and confidence intervals in [6] based on k-nearest neighbor density estimates. The theory developed in [6] has motivated two types of improvements to these k-NN estimators that translate into reduction in bias and significantly enhanced performance. The first improvement is boundary compensation. If the range of amplitudes of the data are bounded then there can be severe bias in the k-NN density estimator. This bias does not decrease as the number of samples increases. We derived a new boundary compensated divergence estimator that forces the bias at the boundary to decrease without affecting the bias at interior points. This estimator does not require knowledge of the boundary of the underlying density. The boundary compensated k-NN estimator was published in a paper [10], co-authored by UM GSRA Sricharan and co-PI's Hero and Raich, in the Proceedings of the IEEE Workshop on Machine Learning in Signal Processing, (MLSP), Aug 2010. The second improvement is a weighted version of the k-NN plug-in entropy estimator of divergence. In [6] we show that the bias of the standard k-NN plug-in estimator appears as a series of terms that decay with rates $(k/M)^{-j/d}$, $j = 2, 4, 6, \dots$, where M is the total number of samples and d is the dimension of the feature space. We define a new estimator by forming the weighted average of k-NN plug-in entropy estimators implemented with different values of k ($k = 1, 2, 3, \dots$). By judicious choice of the weighting coefficients we show that this new estimator achieves a d -independent convergence rate of much faster order $M^{-1/2}$. The weighted k-NN entropy estimator was published in a paper [11], co-authored by UM GSRA Sricharan and co-PI's Hero and Raich, in the Proceedings of the IEEE 2011 Workshop on Statistical Signal Processing (SSP).

2.3 Feature selection for entropy estimation using the maximum entropy principle

To simultaneously address the problem of feature selection and entropy estimation we considered a parametric approach to density estimation. The logarithm of the probability density function is estimated as an m -term approximation over a large dictionary. Using a greedy approach for the m -term approximation, we show in [12] that entropy can be estimated with accuracy $\mathcal{O}(\sqrt{\frac{\log n}{n}})$. We considered two estimators. The first considers brute-force estimation of entropy based on m component density approximation. Although the estimator is not practical (i.e., the computational complexity associated with it is prohibitive) its accuracy is analyzed and used as a baseline to compare to. The second estimator uses a greedy approach for the m -term approximation reducing the optimization to one component at a time and thus enabling a significant reduction in computation complexity relative to the first algorithm. For each algorithm, we were able to show (under specific conditions) that the entropy estimation error is $\mathcal{O}(\sqrt{\frac{\log n}{n}})$. The paper [12] is co-authored by OSU GRA Behrouz Behmardi and the Co-PI's Raich and Hero and was published in the IEEE Proc. of 2011 Intl Conf. on Acoustics, Speech, and Signal Processing.

2.4 Dimension estimation in topic models

In the past few years, probabilistic topic models have been developed and applied to problems in text document classification and computer vision. Such models provide a probabilistic framework for characterizing a corpus of documents (or images) in the bag-of-words representation. These results are directly applicable to fusion of soft (textual or contextual) information and hard (sensor) information. A key feature of such models is that a low dimensional representation is facilitated through latent topic variables. Most inference algorithms in topic models assume a fixed number of topics and determine the number of topics empirically. We developed a new approach for identifying the number of topics in topic models through rank estimation. In [13] we present a rank minimization framework and provided sufficient conditions, which guarantee exact recovery of the number of topics. Moreover, we proposed a heuristic convex relaxation to the rank minimization. Using simulations, we showed that the proposed convex relaxation provides exact rank recovery under the sufficient conditions proposed for the rank minimization problem. The core principle that allows for a tuning-parameter free optimization is the statistical error analysis. A similar principle which we developed here was afterwards utilized in [2] for solving the multiple system of equations setup in which solutions share a similar sparsity pattern. In [3] we presented a convex optimization frame for solving the problem in [13] efficiently. Our approach allowed us to consider problems of real-world dimensions (e.g., thousands of documents consisting of thousands of vocabulary words). A more comprehensive journal version of this work [1] is currently under review. The paper [13] is co-authored by OSU GRA Behrouz Behmardi and was published in the IEEE Proc. of the 2011 Statistical Signal Processing Workshop. The paper [3] is co-authored by OSU GRA Behrouz Behmardi and was published in the IEEE Proc. of the 2011 Machine Learning for Sig-

nal Processing Workshop. The paper [1] co-authored by OSU GRA Behrouz Behmardi is currently under review for the IEEE Trans. on Signal Processing. The paper [2] is co-authored by OSU GRA Evgenia Chunikhina and is accepted for publication in the IEEE Proc. of the Intl. Conf. on Acoustics Speech and Signal Processing, 2012.

2.5 Multiclass-preserving dimension reduction

Classical methods for dimension reduction (e.g., PCA, LDA) are often motivated by a pre-specified class conditional data distribution. In our recent submission [14], we explore a framework for dimension reduction that is based on a multi-class generalization of Chernoff bound applicable both to parametric and non-parametric models. We were able to show through numerical analysis of classification error rates across multiple datasets that the framework yields comparable (and sometimes superior) performance to other state-of-the-art methods in dimension reduction. The objective we explored can be regarded as an estimator of functional of probability density functions measuring aggregating pairwise dissimilarities of such densities. As this objective fits our framework of estimators of functionals of densities, we suspect that further development of the theory towards a MSE analysis of m-estimates (associated with functionals of densities) would be applicable. Our results are documented in our IEEE Trans. on Pattern Analysis and Machine Intelligence submission (March 2010) by Raich and the graduate student Madan Thangavelu from Oregon State University [14].

3 Personnel supported

The project supported the two co-PI's: Alfred Hero at Michigan and Raviv Raich at Oregon State. It also supported several graduate students: Kumar Sricharan at Michigan and Evgenia Chunikhina and Behrouz Behmardi at Oregon State.

4 Technology Assists and Transitions

Our efforts have been focussed on developing a new theoretical framework for performance-driven sensing and fusion. The theory was not sufficiently mature to be transitioned through technology assists or transitions during the short (2.5 year) duration of this grant.

References

- [1] B. Behmardi and R. Raich, "On confidence-constrained rank recovery in topic models," *IEEE Trans. Signal Processing*, 2011, In Review.
- [2] Evgenia Chunikhina, Gregory Gutshall, Raviv Raich, and Thanh Nguyen, "Tuning-free joint sparse recovery via optimization transfer," in *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Processing*, Kyoto, Japan, March 2012, accepted.

- [3] B. Behmardi and R. Raich, "Convex optimization for exact rank recovery in topic models," in *proc. of IEEE Workshop on Machine Learning for Signal Processing*, Beijing, China, 2011, IEEE, pp. 1–6.
- [4] B. Behmardi and R. Raich, "On provable exact low-rank recovery in topic models," in *proc. of IEEE Workshop on Statistical Signal Processing*, Nice, France, 2011, IEEE, pp. 265–268.
- [5] B. Behmardi, R. Raich, and A. O. Hero, "Entropy estimation using the principle of maximum entropy," in *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Processing*, Prague, Czech Republic, 2011, IEEE, pp. 2008–2011.
- [6] K. Sricharan, R. Raich, and A. Hero, "Empirical estimation of entropy functionals with confidence," tech. rep., CSPL Technical Report 398, Dept. of EECS, University of Michigan, Ann Arbor, Dec 2010. arXiv:1012.4188v1, Dec 2010.
- [7] K. Sricharan, R. Raich, and A. Hero, "Empirical estimation of entropy functionals with confidence," *IEEE Information Theory Transactions*, Submitted.
- [8] K. Sricharan, R. Raich, and A. Hero III, "Global performance prediction for divergence-based image registration criteria," in *IEEE Statistical Signal Processing Workshop*, (Cardiff), Sept. 2009.
- [9] K. Sricharan, R. Raich, and A. Hero III, "Optimized intrinsic dimension estimator using nearest neighbor graphs," in *IEEE Intl Conf. on Acoust., Speech, Signal Processing*, (Dallas), Mar. 2010.
- [10] K. Sricharan, R. Raich, and A. Hero III, "Boundary compensated knn graphs," in *IEEE Workshop on Machine Learning in Signal Processing*, (Kittila, Finland), Aug 2010.
- [11] K. Sricharan, R. Raich, and A. Hero, "Weighted k-nn graphs for renyi entropy estimation in high dimensions," in *IEEE Workshop on Statistical Signal Processing*, June 2011.
- [12] B. Behmardi, R. Raich, and A. Hero, "Entropy estimation using the principle of maximum entropy," in *IEEE Proc of Acoust, Speech and Sig. Proc.*, (Prague, Czech Republic), 2011.
- [13] B. Behmardi and R. Raich, "On provable exact low-rank recovery in topic models," in *IEEE Workshop on Statistical Signal Processing*, June 2011.
- [14] M. Thangavelu and R. Raich, "Error bounds for multiclass linear dimension reduction," *IEEE Trans. on Pattern Anal. and Machine Intell.*, vol. submitted, 2010.
- [15] K. Sricharan, R. Raich, and A. Hero III, "Performance-driven information fusion," in *Workshop on Defense Applications of Signal Processing (DASP)*, (Coolum, Australia), July 2011.

- [16] K. Sricharan, R. Raich, and A. Hero III, "Efficient anomaly detection using bi-partite k-nn graphs," in *Proc of Conf on Neural Information Processing Systems (NIPS)*, (Granada), Dec. 2011.
- [17] K. Sricharan, A. Hero, and B. Rajaratnam, "A local dependence measure and its application to screening for high correlations in large data sets," in *Intl Conference on Information Fusion*, (Chicago), July 2011.