DISTRIBUTED VECTOR DECORRELATION AND ANOMALY DETECTION USING THE VECTOR SPARSE MATRIX TRANSFORM

Leonardo R. Bachega
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052, USA
lebacheg@microsoft.com

Charles A. Bouman
School of Electrical and Computer Engineering
Purdue University
West Lafayette, IN 47907, USA
bouman@purdue.edu

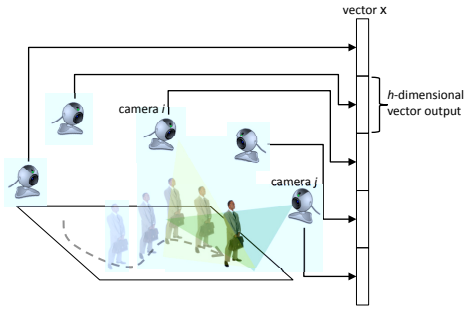


Fig. 1: A camera network where each camera captures an image of the environment from one viewpoint and encodes the image into a vector output. The aggregated outputs from all cameras form the high-dimensional vector, x. Cameras i and j have overlapping views. Since outputs from cameras with overlapping views tend to be correlated, so does the aggregated vector x.

ABSTRACT

Here, we propose the *vector Sparse Matrix Transform (SMT)*, a novel decorrelating transform suitable for performing distributed processing of high dimensional signals in sensor networks. We assume that each sensor in the network encodes its measurements into vector outputs instead of scalar ones. The proposed transform decorrelates a sequence of pairs of vector sensor outputs, until these vectors are decorrelated. In our experiments, we simulate distributed anomaly detection by a camera network monitoring a spatial region. Each camera records an image of the monitored environment from its particular viewpoint and outputs a vector encoding the image. Results show that the vector SMT effectively decorrelates images from the multiple cameras in the network and significantly improves anomaly detection accuracy while requiring low overall communication energy.

1. INTRODUCTION

Event detection and more specifically anomaly detection are important applications for many sensor networks [1]. In general, the vector outputs from all sensors in a network can be concatenated to form a single p-dimensional vector x, and then the goal of anomaly detection is to determine if x corresponds to a typical or anomalous event. Fig. 1 illustrates this scenario for a camera network. The vector outputs from different cameras are likely to be correlated, particularly when the cameras capture overlapping portions of the scene; so for best detection accuracy, vector x should be decorrelated as part of the detection process.

Several methods for distributed decorrelation and detection have been proposed since the 1980s. Traditionally, distributed detection methods rely on a centralized fusion center for data processing [2]. Volume anomaly detection in networks has been studied for scalar measurements [3, 4], and multi-view images [5], with offline, centralized processing methods for detection decision. Ortega et. al. proposed methods for distributed decorrelation of scalar sensor outputs using wavelet transforms with lifting for efficient routing in networks of various topologies [6, 7, 8]. Methods to compute distributed PCA for scalar sensor outputs were proposed in [9, 10]. A distributed KLT algorithm [11, 12] is used to encode vector sensor outputs to resconstruct them at a central location with minimum mean-square error.

In order to decorrelate x, we need an accurate estimate of its covariance matrix. Several methods to estimate covariances of high-dimensional signals have been proposed recently [13, 14, 15, 16, 17, 18]. Among these methods, the Sparse Matrix Transform (SMT) [17], here referred to as the scalar SMT, has been shown to be effective, providing full-rank covariance estimates of high-dimensional signals even when the number n of training samples used to compute the estimates is much smaller than the dimension p of a data sample, i.e, $n \ll p$. Furthermore, the decorrelating transform designed by the SMT algorithm consists of the product of planar rotations known as Givens rotations,

$$E = \prod_{k=1}^{K} E_k = E_1 \cdots E_K \ . \tag{1}$$

Each E_k rotates the coordinates (i_k, j_k) of x by an angle θ_k. Because normally $K = O(p)$, this transform is computationally inexpensive to apply. The scalar SMT has been used in detection and classification of high-dimensional signals [19, 20] and Givens rotations have been used in ICA [21]. Since it involves only pairwise operations between coordinate pairs, it is well-suited to distributed decorrelation [22]. However, this existing method is only well suited for decorrelation of scalar sensor outputs.

In this paper, we propose the vector sparse matrix transform (vector SMT), a novel algorithm suited for distributed signal decorrelation in sensor networks where each sensor outputs a vector. It generalizes the concept of the scalar sparse matrix transform in [17] to decorrelation of vectors. This novel algorithm operates on pairs of sensor outputs, and it has the interpretation of maximizing the constrained log likelihood of x. In particular, the vector SMT decorrelating transform is defined as an orthonormal transformation constrained to be formed by a product of pairwise transforms between pairs of vector sensor outputs. We design this transform using a greedy optimization of the likelihood function of x. Once this transform is designed, the

This material is based upon PhD work done while at Purdue University supported by, or in part by, the U.S. Army Research Laboratory and the U.S. Army Research Office under contract/grant number 56541-CI.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

associated pairwise transforms are applied to sensor outputs distributed over the network, without the need of a powerful central sink node. The total number of pairwise transforms is a model order parameter, which can be adjusted to control the total amount of communication energy used during the decorrelation.

Simulation results using multi-camera image data show that the proposed vector SMT transform effectively decorrelates image measurements from the multiple cameras in the network while maintaining low overall communication energy consumption. Since it enables joint processing of the multiple vector outputs, our method provides significant improvements to anomaly detection accuracy when compared to the baseline case when the images are processed independently.

2. DISTRIBUTED DECORRELATION WITH THE VECTOR SPARSE MATRIX TRANSFORM

Our goal is to decorrelate the p-dimensional vector x aggregated from outputs of all sensors, where each of the L sensors outputs an h-dimensional sub-vector of x.

2.1. The Vector SMT Model

Let the p-dimensional vector x be partitioned into L sub-vectors,

$$x = \begin{bmatrix} x^{(1)} \\ \hline \vdots \\ \hline x^{(L)} \end{bmatrix} ,$$

where each sub-vector, $x^{(i)}$ is an h-dimensional output from a sensor $i = 1, \cdots, L$ in a network. The decorrelating vector SMT transform is an orthonormal $p \times p$ transform, T, written as the product of M orthonormal, sparse matrices,

$$T = \prod_{m=1}^{M} T_m , \tag{2}$$

where each pairwise transform, $T_m \in \mathbb{R}^{p \times p}$, is a block-wise sparse, orthonormal matrix that operates exclusively on the $2h$-dimensional subspace of the sub-vector pair $x^{(i_m)}$, $x^{(j_m)}$, and M is a model order parameter.

The vector SMT in (2) generalizes the concept of the scalar SMT [17] to the decorrelation of pairs of vectors. Figs. 2 compares the vector and the scalar SMTs approaches graphically, together with the FFT. In the vector SMT, each orthonormal matrix T_m corresponds to series of decorrelating butterflies that operate exclusively on coordinates of a single pair of sub-vectors of x.

In a sensor network, we compute the distributed decorrelation of x by distributing the application of transforms T_m from the product (2) across multiple sensors. Before the decorrelation, each sub-vector $x^{(i)}$ of x is the output of a sensor i and is stored locally in that sensor. Applying each T_m to sub-vectors $x^{(i_m)}$, $x^{(j_m)}$ requires energy for a point-to-point communication of one h-dimensional sub-vector between sensors i_m and j_m. The resulting sub-vectors are cached at the sensor that performed the decorrelation to avoid extra communication.

2.2. The Design of the Vector SMT

We design the transform in (2) from training data, using the maximum likelihood estimation of its covariance matrix. Let $X = [x_1, \cdots, x_n] \in \mathbb{R}^{p \times n}$, be a $p \times n$ matrix where each column, x_i is a p-dimensional zero mean Gaussian random vector with covariance R. In general, $R = T\Lambda T^t$, where Λ and T are the diagonal eigenvalue and orthonormal matrices, respectively. In this case, the log likelihood of X given T and Λ is

$$\log p_{(T,\Lambda)}(X) = -\frac{n}{2}\mathrm{tr}[\mathrm{diag}(T^t S T)\Lambda^{-1}] - \frac{n}{2}\log(2\pi)^p|\Lambda| , \tag{3}$$

where $S = \frac{1}{n}XX^t$. The functions $\mathrm{diag}(\cdot)$ and $|\cdot|$ are the diagonal and determinant, respectively, of a matrix argument. When constraining T to be of the product form of (2), the joint maximum likelihood estimates $\widehat{\Lambda}$ and $\widehat{T}$ are given by

$$\hat{T} = \arg \min_{T = \prod_{m=1}^{M} T_m} \left\{ \left| \mathrm{diag}(T^t S T) \right| \right\} \tag{4}$$

$$\hat{\Lambda} = \mathrm{diag}(\hat{T}^t S \hat{T}) . \tag{5}$$

Since the minimization in (4) has a non-convex constraint, its global minimizer is difficult to find. Therefore, we use a greedy procedure that designs each T_m, $m = 1, \cdots, M$, independently while keeping the others fixed. We start with $S_1 = S$ and $X_1 = X$, and iterate over the following steps:

$$\hat{T}_m = \arg \min_{T_m \in \Omega} \left\{ \left| \mathrm{diag}(T_m^t S_m T_m) \right| \right\} \tag{6}$$

$$S_{m+1} = \hat{T}_m^t S_m \hat{T}_m \tag{7}$$

$$X_{m+1} = \hat{T}_m^t X_m , \tag{8}$$

where Ω is the set of all allowed pairwise transforms. Since T_m operates exclusively on $x^{(i_m)}$ and $x^{(j_m)}$, once the pair (i_m, j_m) is selected, the design of T_m involves only the components of X_m associated with these sub-vectors and their associated $2h \times 2h$ sample covariance, $S_m^{(i_m, j_m)}$. The minimization in (6) for a fixed subvector pair (i_m, j_m) can be recast in terms of $S_m^{(i_m, j_m)}$, and the $2h \times 2h$ orthonormal matrix E,

$$E_m = \arg \min_{E \in \Omega_{2h \times 2h}} \left\{ \left| \mathrm{diag}(E^t S_m^{(i_m, j_m)} E) \right| \right\} , \tag{9}$$

where $\Omega_{2h \times 2h}$ is the set of all valid $2h \times 2h$ orthonormal transforms. In practice, the minimization in (9) is precisely the scalar SMT design in [17]. Once E_m is selected, it is directly mapped into the $p \times p$ block sparse matrix T_m. Finally, the overall change in the log likelihood in (3) due to applying T_m to X_m and maximized with respect to $\hat{\Lambda}(T_m)$ is given by

$$\begin{aligned} \Delta \log p_{(T_m, \hat{\Lambda}(T_m))}(X_m) &= -\frac{n}{2}\log \frac{|\mathrm{diag}(T_m^t S_m T_m)|}{|\mathrm{diag}(S_m)|} \\ &= -\frac{n}{2}\log \frac{|\mathrm{diag}(E_m^t S_m^{(i_m, j_m)} E_m)|}{|\mathrm{diag}(S_m^{(i_m, j_m)})|} . \end{aligned} \tag{10}$$

Therefore, we use the maximum value of (10) as the criterion for selecting the pair (i_m, j_m) during the design of $\hat{T}_m$ in (6).

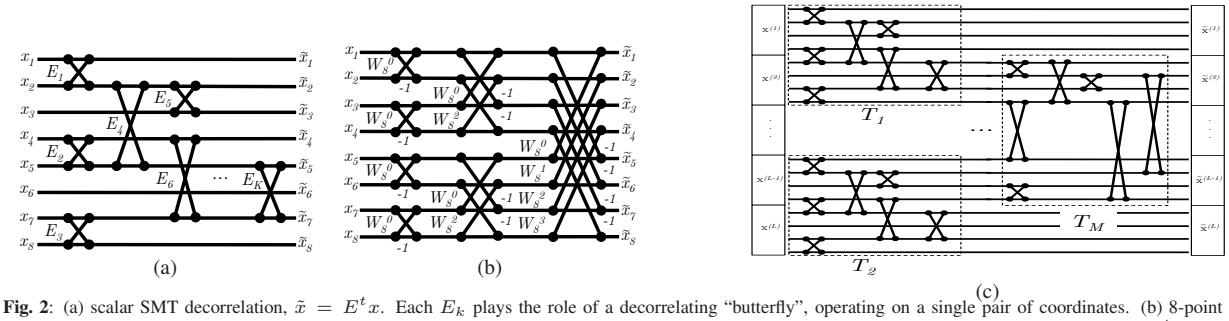


Fig. 2: (a) scalar SMT decorrelation, $\tilde{x} = E^t x$. Each E_k plays the role of a decorrelating "butterfly", operating on a single pair of coordinates. (b) 8-point FFT, seen as a particular case of the scalar SMT where the butterflies are constrained in their ordering and rotation angles. (c) Vector SMT decorrelation, $\tilde{x} = T^t x$, with each T_m decorrelating a sub-vector pair of x instead of a single coordinate pair. T_m is an instance of the scalar SMT with decorrelating butterflies operating only on coordinates of a single pair of sub-vectors.

3. ANOMALY DETECTION

Let $\tilde{x} = \hat{T}^t x$ be a p-dimensional vector decorrelated using the orthonormal transform $\hat{T}$ in (4). As discussed in [23], for a given probability of false alarm, the optimal anomaly detection test is

$$\widetilde{D}_{\hat{\Lambda}}(\tilde{x}) = ||\tilde{x}||^2_{\hat{\Lambda}} = \sum_{i=1}^{p} \frac{\tilde{x}_i^2}{\hat{\lambda}_i} > \eta^2 \,, \qquad (11)$$

where η controls the probability of false alarm and $\hat{\Lambda}$ is the eigenvalue matrix in (5). In addition to using ROC curves, we evaluate detection accuracy of the test in (11) by computing the volume of the ellipsoid within the region $||\tilde{x}||^2_{\hat{\Lambda}} \leq \eta^2$. This ellipsoid volume [24] is evaluated by

$$V(\hat{\Lambda}, \eta) = \frac{\pi^{p/2}}{\Gamma(1 + p/2)} \eta^p \sqrt{|\hat{\Lambda}|} \,, \qquad (12)$$

and serves as a direct measure of the probability of missed detection of (11) for a fixed probability of false alarm.

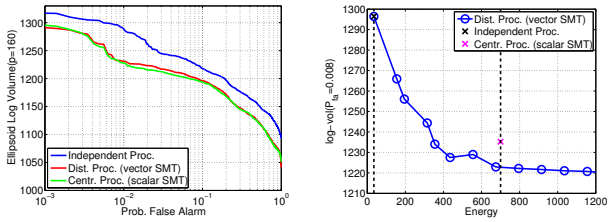
4. SIMULATIONS

We provide anomaly detection simulation results with multi-view camera data to quantify the effectiveness of our method. We assume communications occur between sensors connected in a network with binary tree topology, and that communication of one scalar value between adjacent sensors uses one unit of energy. Fig. 3 compares the vector SMT decorrelation with two other approaches for processing the sensor outputs before making a detection decision, a centralized and an independent one. In the centralized approach, all sensors send their h-dimensional vector outputs to the root of the tree. We then use the scalar SMT [17] to decorrelate x at the root. In the independent approach no decorrelation of outputs is performed, each sensor computes a partial likelihood of its output independently and communicates it to the root of the tree.

Fig. 4 shows $L = 8$ simultaneous camera views of a courtyard, constructed from a 4.2 min video sequence from [25]. We subsample 1 in 3 frames from this sequence, and use 800 of the selected samples to compute the encoding PCA transforms for each camera view. The final courtyard dataset has 1734 samples of $p = 160$ dimensions, with each view encoded in a vector of $h = 20$ dimensions.

Method	Algorithm	Communication	Decorrelation
Vector SMT (distributed)	Vector SMT	Between pairs of nodes / caching	sub-vector pairs in network
Centralized	Scalar SMT	Vector outputs to centralized node	coordinate pairs at single node
Independent	None	Partial likelihoods to centralized node	–

Fig. 3: Summary of the several approaches to sensor output decorrelation compared and their main properties.



(a) (b)

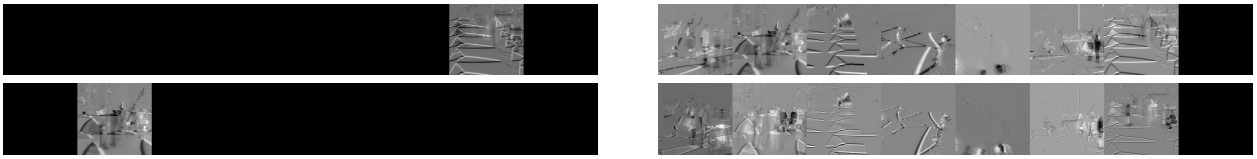
Fig. 6: Detection accuracy measured by the ellipsoid log-volume for the courtyard data set. Coverage plots showing the log-volume *vs.* probability of false alarm: (a) model order, $M = 4$, matching 50% of the energy consumed for the centralized processing; (b) log-volume *vs.* communication energy for fixed probability of false alarm, $P_{FA} = 0.008$. The vector SMT has similar accuracy to the centralized approach while requiring 50% of its communication energy.

Fig. 5 shows two eigen-images associated with the two largest eigenvalues for both the independent and vector SMT approaches. In the independent processing case (Fig. 5(a)), each eigen-image corresponds to a single camera view, containing no information regarding the relationship between different views. On the other hand, the vector SMT eigen-images (Fig. 5(b)) contain joint information of the correlated views. Since camera view 8 is not correlated with any other view, it does not appear together with others in the same eigen-image.

Fig. 6 compares the accuracy of all approaches using the volume of the ellipsoid covering test samples. We split the samples into a training set, with 300 samples, and a test set, with 1434 samples. Fig. 6(a) shows the ellipsoid log-volume computed for all false alarm rates. We select the vector SMT model order so that the distributed decorrelation consumes only 50% of the energy required for the centralized approach. Fig. 6(b) shows the ellipsoid log-volume for a fixed false alarm rate (0.8%) *vs.* communication energy. The independent approach has low accuracy while requiring low communication energy. The centralized decorrelation is highly accurate, but it requires large amounts of communication energy. The vector SMT increases the detection accuracy after each pairwise transform.

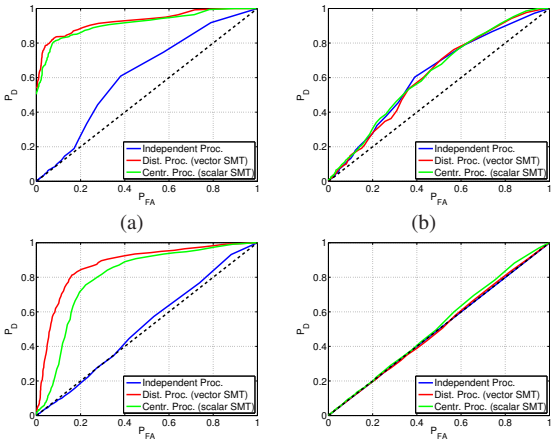
Fig. 7 shows ROC curves for detection of anomalous

Fig. 4: The courtyard dataset [25]: eight cameras, with ids 1 to 8 from left to right, monitor a courtyard from different viewpoints. Several activities in the courtyard are captured simultaneously by several cameras.



(a) (b)

Fig. 5: Two eigen-images from the eight camera views of the courtyard dataset. Each eigen-image has eight views (columns) associated to it. (a) independent processing of camera views: each eigen-image corresponds to a single view and does not contain correlation information among multiple views; (b) joint processing modeled by the vector SMT: each eigen-image contains joint information of all correlated views.



(a) (b) (c) (d)

Fig. 7: ROC analysis of detection accuracy: (a) and (b) artificially generated anomalies by a 4-fold increase in the largest eigenvalue of a single view for views 2 and 8, respectively. (c) and (d) Ocean's Eleven anomalies, generated by swapping images of a single camera view between samples for views 2 and 8, respectively. Decorrelation improves detection accuracy when anomalies appear in correlated camera view.. When the anomaly is inserted in a uncorrelative view, decorrelation methods do not improve the detection accuracy.

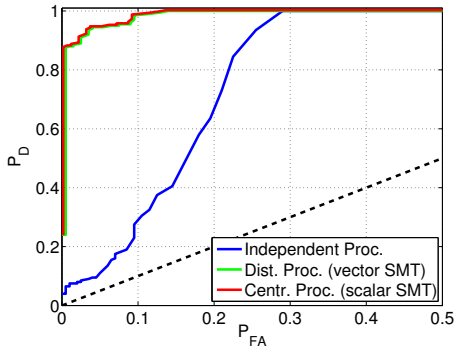


Fig. 8: People coalescing in the middle of a courtyard: scalar and vector SMTs are highly accurate for small probabilities of false alarm with vector SMT consuming approximately 60% of communication energy required for the scalar SMT.

samples. We use 200 typical samples to learn the decorrelating transform and the remaining samples for testing. Figs. 7(a) and (b) show the results for anomalies generated by an artificial 4-fold increase in the largest component of the vector output of a single camera view, and injected in views 2 and 8, respectively. Figs. 7(c) and (d) show the results for what we call the "Ocean's Eleven" anomaly, injected into the camera views 2 and 8, respectively. This anomaly is generated by swapping images of a single view between two samples captured at different instants.[1] Since view 2 is correlated with other views, detection of anomalies in these views is accurate when we decorrelate the views using the vector and scalar SMT approaches, and very inaccurate when we process the views independently. Because view 8 is uncorrelated with other views, decorrelation does not improve accuracy in this case.

Fig. 8 shows the ROC curves for detection of a suspicious (anomalous) activity where people coalesce in at the center of the courtyard. We select 200 samples where a group of people coalesces at the center of the courtyard and label them as anomalous, while selecting another 200 samples where the group does not coalesce and label them as typical. We use another 300 typical samples to train the vector SMT. The vector SMT decorrelation in this experiment consumes 60% of the communication energy required for the scalar SMT. Detection is very accurate when using vector and scalar SMTs for view decorrelation, and inaccurate when processing the views independently, specially for low probabilities of false alarm.

[1]We refer to it as the Ocean's Eleven anomaly because of the resemblance with the anomaly created to trick the surveillance cameras during the casino robbery in the Ocean's Eleven film [26].

5. CONCLUSIONS

We have proposed a novel method for decorrelation of vector measurements distributed across sensor networks. The new method is based on the constrained maximum likelihood estimation of the joint covariance of the measurements. It generalizes the concept of the previously proposed sparse matrix transform to the decorrelation of vectors. We have demonstrated its effectiveness in anomaly detection experiments using multi-view image data. In future work, we will provide a distributed algorithm for the design of the vector SMT decorrelating transform.

6. REFERENCES

[1] J.-F. Chamberland and V. V. Veeravalli, "Wireless sensors in distributed detection applications," *IEEE Signal Processing*

Magazine, vol. 25, no. 5, pp. 16–25, May 2007.

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

[3] A. Lakhina, M. Crovella, and C. Diot, "Diagnosing network-wide traffic anomalies," in *Proc. 2004 Conf. Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM'04)*, vol. 34, Oct. 2004, pp. 219–230.

[4] ——, "Mining anomalies using traffic feature distributions," in *Proc. 2005 Conf. Applications, Technologies, Architectures, and Protocols for Computer Communications (SIGCOMM'05)*, vol. 34, Oct. 2005, pp. 217–228.

[5] T. Ahmed, B. Oreshkin, and M. Coates, "Machine learning approaches to network anomaly detection," in *Proc. Second Workshop on Tackling Computer Systems Problems with Machine Learning (SysML)*, Cambridge, MA, Apr. 2007.

[6] A. Ciancio and A. Ortega, "A distributed wavelet compression algorithm for wireless sensor networks using lifting," in *Proc. IEEE Int. Conf. Acoustics Speech and Signal Processing (ICASSP)*, Montreal, Quebec, Canada, May 2004.

[7] A. Ciancio, S. Pattem, A. Ortega, and B. Krishnamachari, "Energy-efficient data representation and routing for wireless sensor networks based on a distributed wavelet compression algorithm," in *Proc. 5th Int. Conf. Information Processing in Sensor Networks (IPSN)*, Nashville, TN, Apr. 2006.

[8] G. Shen, S. Pattem, and A. Ortega, "Energy-efficient graph-based wavelets for distributed coding in wireless sensor networks," in *Proc. IEEE Int. Conf. Acoustics Speech and Signal Processing (ICASSP)*, Taipei, Taiwan, Apr. 2009.

[9] Y.-A. Le Borgne, S. Raybaud, and G. Bontempi, "Distributed principal component analysis for wireless sensor networks," *Sensors*, vol. 8, no. 8, pp. 4821–4850, 2008.

[10] A. Wiesel and A. O. Hero, "Decomposable principal component analysis," *IEEE Trans. Signal Processing*, vol. 57, no. 11, pp. 4369–4377, Nov. 2009.

[11] M. Gastpar, P. Dragotti, and M. Vetterli, "The distributed Karhunen-Loeve transform," *IEEE Trans. Information Theory*, vol. 52, no. 12, pp. 5177–5196, Dec. 2006.

[12] A. Amar, A. Leshem, and M. Gastpar, "A greedy approach to the distributed Karhunen-Loeve transform," in *IEEE Int. Conf. Acoustics Speech and Signal Processing (ICASSP)*, Dallas, TX, Mar. 2010, pp. 2970–2973.

[13] J. P. Hoffbeck and D. A. Landgrebe, "Covariance matrix estimation and classification with limited training data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 7, pp. 763–767, 1996.

[14] M. J. Daniels and R. E. Kass, "Shrinkage estimators for covariance matrices," *Biometrics*, vol. 57, no. 4, pp. 1173–1184, 2001.

[15] O. Ledoit and M. Wolf, "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection," *Journal of Empirical Finance*, vol. 10, no. 5, pp. 603–621, 2003.

[16] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, Jul. 2008.

[17] G. Cao, L. Bachega, and C. Bouman, "The sparse matrix transform for covariance estimation and analysis of high dimensional signals," *IEEE Trans. Image Processing*, vol. 20, no. 3, pp. 625–640, Mar. 2011.

[18] J. Theiler, "The incredible shrinking covariance estimator," *Proc. SPIE*, vol. 8391, p. 83910P, 2012.

[19] L. R. Bachega, C. A. Bouman, and J. Theiler, "Hypothesis testing in high-dimensional space with the sparse matrix transform," in *The 6th IEEE Sensor Array and Multichannel Signal Processing Workshop*. Israel: IEEE, Oct. 2010.

[20] J. Theiler, G. Cao, L. Bachega, and C. Bouman, "Sparse matrix transform for hyperspectral image processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 1, pp. 424–437, Jun. 2011.

[21] J.-F. Cardoso, "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11, no. 1, pp. 157–192, Jan 1999.

[22] S. Hariharan, L. R. Bachega, N. Shroff, and C. A. Bouman, "Communication efficient signal detection in correlated clutter for wireless sensor networks," in *Asilomar*, Pacific Grove, CA, Nov. 2010.

[23] L. R. Bachega, J. Theiler, and C. A. Bouman, "Evaluating and improving local hyperspectral anomaly detectors," in *Proc. Applied Imagery Pattern Recognition Workshop (AIPR), 2011 IEEE*, Washington, DC, Oct. 2011, pp. 1–8.

[24] J. Theiler and D. R. Hush, "Statistics for characterizing data on the periphery," *Proc. IEEE Int. Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 4764–4767, Jul. 2010.

[25] G. Deninan, B. Bhanu, H. Nguyen, C. Ding, A. Kamal, C. Ravishankar, A. Roy-Chowdhury, A. Ivers, and B. Varda, "Videoweb dataset for multicamera activities and non-verbal communication," in *Distributed Video Sensor Networks*, B. Bhanu, C. Ravishankar, A. Row-Chowdhury, H. Aghajan, and D. Terzopoulos, Eds. Springer, 2010.

[26] S. Soderbergh, "Ocean's Eleven (film)," *Warner Bros.*, Dec. 2001.