Distributed Signal Decorrelation in Wireless Sensor Networks Using the Sparse Matrix Transform

Leonardo R. Bachega *, Srikanth Hariharan, Charles A. Bouman, Ness Shroff

ABSTRACT

In this paper, we propose the vector SMT, a new decorrelating transform suitable for performing distributed anomaly detection in wireless sensor networks (WSN). Here, we assume that each sensor in the network performs vector measurements, instead of a scalar ones. The proposed transform decorrelates a sequence of pairs of vector sensor measurements, until the vectors from all sensors are completely decorrelated. We perform simulations with a network of cameras, where each camera records an image of the monitored environment from its particular viewpoint. Results show that the proposed transform effectively decorrelates image measurements from the multiple cameras in the network. Because it enables joint processing of the multiple images, our method provides significant improvements to anomaly detection accuracy when compared to the baseline case when we process the images independently.

1. INTRODUCTION

In this paper, we address the problem of high-dimensional signal detection in a wireless sensor network (WSN) based on the combined vector measurements from all sensors in the network. The detection of vector signals is a central problem in signal processing. When the p coordinates of the signal vector x are correlated, the ability to apply a decorrelating transform to x is paramount to good detection accuracy. Sensor networks typically function in an ad-hoc manner without the requirement of a powerful centralized node. Therefore, we are interested in developing an algorithm that decorrelates these vector measurements, and can be both computed and communicated in a distributed manner in these networks.

One of the important applications of signal detection is detecting anomalies. Consider a scenario where each sensor is equipped with a camera, and multiple cameras collectively monitor the same environment (Figure 1). Each camera records images of the same environment, each from its specific viewpoint, and outputs a vector encoding its image. In this scenario, the p-dimensional vector x corresponds to the joint output over all cameras in the network. Since the images from different cameras are likely to be correlated, we need to decorrelate them before using the combined information to make a decision of whether an event/anomaly is present in the environment or not.

Our approach for a distributed *in network* decorrelation is a modified version of the Sparse Matrix Transform (SMT).[1] The SMT provides full-rank covariance estimates of high-dimensional signals even when the number of training samples used to compute the estimates, $N \ll p$. The associated decorrelating transform designed by the SMT algorithm consists of a product of $O(p)$ Givens rotations, and therefore, it is computationally inexpensive to apply. Previously, we showed[2] how to use the SMT for detection and classification of high-dimensional signals, and used the SMT for classification of faces,[3] as well as for detection of anomalous as well as deterministic signals in hyperspectral data sets.[4] Most recently, we showed how to apply the SMT in a wireless sensor network in a distributed manner, and analyze the communication cost of this method.[5] Note that in these approaches we assume that each sensor has a scalar measurement.

A number of other transforms have also been studied in the literature for decorrelation and signal detection. Decorrelation using a wavelet transform with lifting in a sensor network has been studied for both a linear topology,[6] and for two-dimensional networks.[7] Finally, a unidirectional wavelet transform with lifting in a network with a tree topology[8] has also been proposed. Here, communication is only towards the root of the tree, i.e, the transform is such that a parent node need not communicate with a child node. Also, for

*This material is based upon work 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.

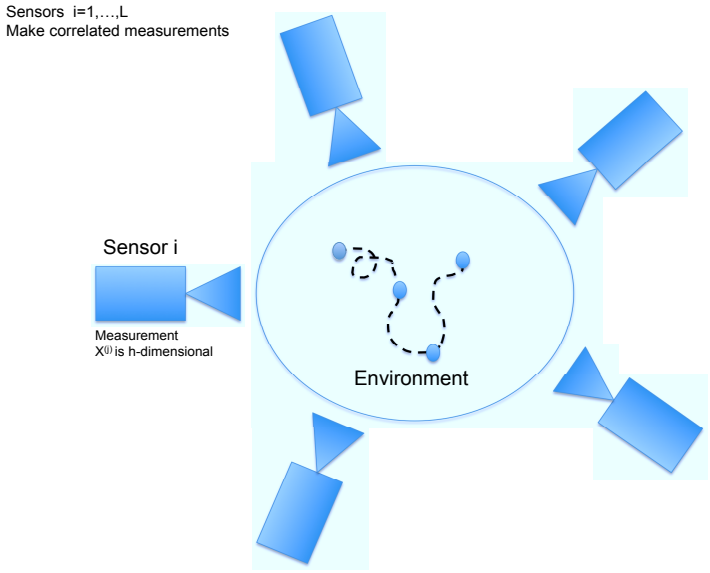


Figure 1. A wireless network of cameras. Each camera captures one image, i.e, a vector measurement, from one viewpoint of the same environment being monitored. The aggregated measurements from all the cameras form a high-dimensional vector.

specific types of signals such as piecewise constant signals, and Gaussian signals, a modified Haar transform, and a DPCM, respectively, have been studied for decorrelating these signals.[9]

Here, we propose a modified version of the SMT that is suitable to work in a distributed manner in environments such as in a wireless sensor network, where each sensor contains a contiguous piece of the high-dimensional signal x. In this **vector SMT**, the decorrelating transform is approximated by a sequence of SMTs, each operating over the dimensions of x contained in a pair of sensors. We apply these techniques to study an anomaly detection problem in a sensor network where each sensor is equipped with a camera. We compare our approach to the baseline case of assuming that the captured images are uncorrelated. Our experimental results show that the vector SMT based method provides significant improvements in detection accuracy.

The rest of this paper is organized as follows: Section 2 briefly describes the main concepts of the SMT.[1] Section 3 lays out the distributed SMT algorithm between vector observations. Section 4 shows how to use the SMT between vector observations to enable distributed detection in a WSN. Section 5 shows simulation results of detection using multi-camera views of objects. Finally, the main conclusions and future work are discussed in Section 6.

2. THE SPARSE MATRIX TRANSFORM (SMT)

The SMT design consists of estimating the full set of eigenvectors and associated eigenvalues for a general p-dimensional signal. More specifically, the objective is to estimate the orthonormal matrix E and diagonal matrix Λ such that the signal covariance can be decomposed as $R = E\Lambda E^t$, and to compute this estimate from N independent training vectors, $Y = [y_1, \cdots, y_N]$. This is done by assuming the samples are i.i.d. Gaussian random vectors and computing the constrained maximum log-likelihood (ML) estimates of E and Λ. We have shown[1] that these constrained ML estimates are given by

$$\hat{E} = \arg \min_{E \in \Omega_K} \left\{ \left| \text{diag}(E^t S E) \right| \right\} \quad (1)$$

$$\hat{\Lambda} = \text{diag}(\hat{E}^t S \hat{E}) \, , \quad (2)$$

where $S = \frac{1}{N} YY^t$ is the sample covariance matrix, and Ω_K is the set of allowed orthonormal transforms.

If $N > p$ and Ω_K is the set of all orthonormal transforms, then the solution to (1) and (2) is the diagonalization of the sample covariance, i.e, $\hat{E}\hat{\Lambda}\hat{E}^t = S$. However, the sample covariance is a poor estimate of the covariance when $N < p$.

In order to improve the accuracy of the covariance estimate, we will impose the constraint that Ω_K be the set of sparse matrix transforms (SMT) of order K. More specifically, we will assume that the eigen-transformation has the form

$$E = \prod_{k=1}^{K} E_k = E_1 \cdots E_K \ , \tag{3}$$

where each E_k is a Givens rotation over the coordinate pair (i_k, j_k) by an angle θ_k,

$$E_k = I + \Theta(i_k, j_k, \theta_k) \ , \text{ where}$$

$$[\Theta]_{ij} = \begin{cases} \cos(\theta_k) - 1 & \text{if } i = j = i_k \text{ or } i = j = j_k \\ \sin(\theta_k) & \text{if } i = i_k \text{ and } j = j_k \\ -\sin(\theta_k) & \text{if } i = j_k \text{ and } j = i_k \\ 0 & \text{otherwise} \end{cases} \ . \tag{4}$$

Intuitively, each Givens rotation, E_k, plays the same role as the butterflies of a fast Fourier transform (FFT). In fact, the SMT is a generalization of both the FFT and the orthonormal wavelet transform. However, since both the ordering of the coordinate pairs, (i_k, j_k), and the values of the rotation angles, θ_k, are unconstrained, the SMT can model a much wider range of transformations. It is often useful to express the order of the SMT as $K = rp$, where r is the average number of rotations per coordinate, being typically very small: $r < 5$. The optimization of (1) is non-convex, so we use a greedy optimization approach in which we select each rotation, E_k, in sequence to minimize the cost. The greedy optimization can be done fast if a graphical constraint can be imposed to the data.[10] The parameter r can be estimated using cross-validation over the training set[10,11] or using the minimum description length (MDL).[4]

Typically, r is small (< 5), so that the computation to apply the SMT to a vector of data is very low, i.e, $2r + 1$ floating-point operations per coordinate. Therefore, we can apply the SMT decorrelating transform to p-dimensional random vectors in only $(2r + 1)p$ steps.

3. DISTRIBUTED SMT BETWEEN VECTOR OBSERVATIONS

In this section, we introduce the concept of a vector SMT. The vector SMT is designed to decorrelate observations from a network of sensors, each of which produces an h-dimensional vector measurement. As with the SMT of Section 2, the vector SMT works by sequentially decorrelating pairs of vector observations until the entire set of observations across the network is decorrelated.

Figure 3 illustrates the concept graphically. Each of L sensors makes an h-dimensional measurement denoted by the column vector $x^{(i)} \in \mathbb{R}^h$ for the i^{th} sensor. For design of the vector SMT, we will also require training data, so for training we will assume that each sensor has made N independent observations of the h dimensional measure. In this case, we will denote the N independent vector observations by the matrix $X^{(i)} \in \mathbb{R}^h \times \mathbb{R}^N$. Using this notation, the full set of observations on the network is given by the $p = hL$ column vector

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

and the corresponding training data is given by the $p \times N$ matrix

$$X = \begin{bmatrix} X^{(1)} \\ \vdots \\ X^{(L)} \end{bmatrix} \ .$$

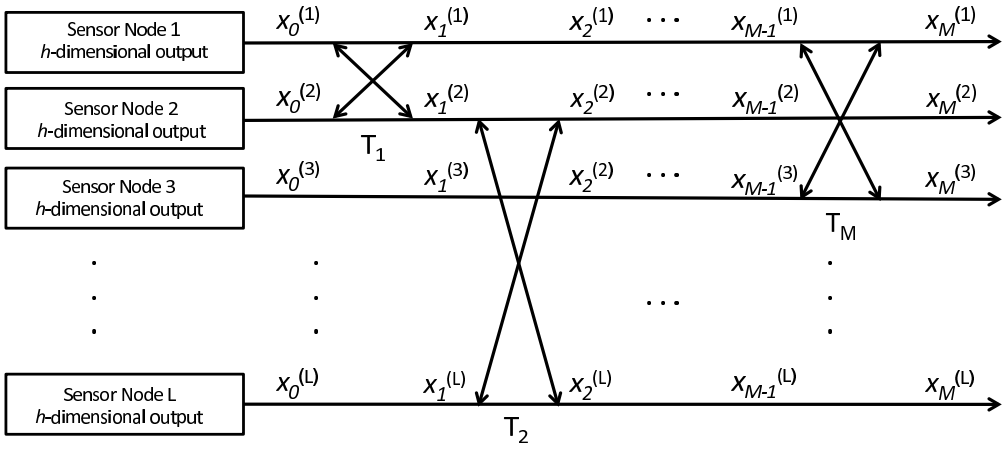


Figure 2. Decorrelation across multiple sensor vector measurements

The sample covariance is then given by

$$S = \frac{1}{N} X X^t \ .$$

Since we will be operating on pairs of vector observations, we will define some additional useful notation. The $2h \times N$ matrix $X^{(i,j)}$ is given by

$$X^{(i,j)} = \left[\begin{array}{c} X^{(i)} \\ X^{(j)} \end{array} \right] \ ,$$

and the associated $2h \times 2h$ sample covariance matrix is given by

$$S^{(i,j)} = \frac{1}{N} X^{(i,j)} \left[X^{(i,j)} \right]^t \ .$$

Each decorrelating butterfly of the vector SMT operates on a pair of h-dimensional vector observations. Let T_m represent the m^{th} orthonormal decorrelating transform. So if we let $X_0 = X$ and $S_0 = S$, then each successive transformation results in the recursion

$$\begin{aligned} X_m &= T_m X_{m-1} \\ S_m &= T_m S_{m-1} T_m^t \ , \end{aligned}$$

But in fact, only the sub-vectors $X_{m-1}^{(i_m)}$ and $X_{m-1}^{(j_m)}$ are modified by the transform T_m, and the remaining sub-vectors are unchanged. Consequently, it is useful to denote this transform on the two lower dimensional sub-vectors as

$$\left[\begin{array}{c} X_m^{(i_m)} \\ X_m^{(j_m)} \end{array} \right] = \left[\begin{array}{cc} E_m^{(1,1)} & E_m^{(1,2)} \\ E_m^{(2,1)} & E_m^{(1,2)} \end{array} \right] \left[\begin{array}{c} X_{m-1}^{(i_m)} \\ X_{m-1}^{(j_m)} \end{array} \right] \ ,$$

or more compactly as

$$X_m^{(i_m, j_m)} = E_m X_{m-1}^{(i_m, j_m)} \ ,$$

where E_m is a $2h \times 2h$ matrix. Figure 3 illustrates the relationship between the $2h \times 2h$ transform E_m, and the larger, but sparse, transform T_m. Notice that the four blocks of E_m are inserted in the appropriate block locations to form the larger matrix T_m.

As in the scalar SMT, we use greedy optimization of the log likelihood to design the transformation T_m. So we assume that the columns of X_m are zero-mean multivariate Gaussian i.i.d. vectors with covariance

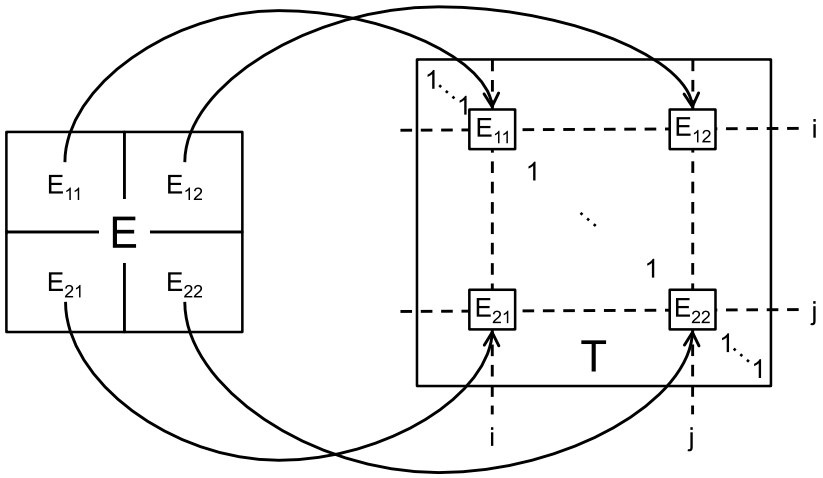


Figure 3. Mapping from the $2h \times 2h$ matrix E to the $p \times p$ matrix T associated to the node pair (i, j).

R, and that the covariance can be represented with the eigen-decomposition of $R = T\Lambda T^t$ where T is orthonormal and Λ is diagonal. Under this assumption, the log likelihood is given by

$$\log p_{(T,\Lambda)}(X_{m-1}) = -\frac{N}{2} tr[diag(T^t S_{m-1} T)\Lambda^{-1}] - \frac{Np}{2}\log(2\pi) - \frac{N}{2}\log|\Lambda| \ . \tag{5}$$

If we plug in the maximum likelihood estimate of $\hat{\Lambda} = diag(T^t S_{m-1} T)$, then this results in the likelihood function

$$\begin{aligned} L_m(T) &= \log p_{(T,\hat{\Lambda})}(X_{m-1}) \\ &= -\frac{Np}{2} - \frac{Np}{2}\log(2\pi) - \frac{N}{2}\log|diag(T^t S_{m-1} T)| \ . \end{aligned}$$

Therefore, the change in the log likelihood resulting from the decorrelating transform T is given by

$$\Delta L_m(T) = L_m(T) - L_m(I) = -\frac{N}{2} log \frac{|diag(T^t S_{m-1} T)|}{|diag(S_{m-1})|}$$

Now since the transform T is assumed to only operate between the sub-vectors i and j, only correlations between the vectors $X_{m-1}^{(i)}$ and $X_{m-1}^{(j)}$ are effected, and we have that

$$\frac{|diag(T^t S_{m-1} T)|}{|diag(S_{m-1})|} = \frac{|diag(E^t S_{m-1}^{(i,j)} E)|}{|diag(S_{m-1}^{(i,j)})|}$$

So we can find the optimum transform T_m by minimizing over the quantities E_m, i_m, and j_m as

$$(E_m, i_m, j_m) = \arg \min_{E \in \Omega_k, i, j} \left\{ \frac{|diag(E^t S_{m-1}^{(i,j)} E)|}{|diag(S_{m-1}^{(i,j)})|} \right\} \ ,$$

where Ω_k is the valid set of orthonormal transforms. Once the quantities (E_m, i_m, j_m) are obtained, the transform T_m can be formally evaluated using Kronecker product $\otimes$ as

$$\begin{aligned} T_m = & \quad J^{(i_m,i_m)} \otimes E_m^{(1,1)} + J^{(i_m,j_m)} \otimes E_m^{(1,2)} \\ & + \quad J^{(j_m,i_m)} \otimes E_m^{(2,2)} + J^{(j_m,j_m)} \otimes E_m^{(2,1)} \\ & + \quad I_{p\times p} - (J^{(i_m,i_m)} + J^{(j_m,j_m)}) \otimes I_{h\times h} \end{aligned} \tag{6}$$

where

$$\left[J^{(i,j)}\right]_{i'j'} = \begin{cases} 1 & \text{if } i' = i \text{ and } j' = j \\ 0 & \text{otherwise} \end{cases} . \tag{7}$$

In practice, the optimization of E is precisely the same problem as SMT design presented in Section 2; so we use the greedy optimization procedure for this part. Then optimization over (i, j) is done by exhaustive search over these variables.

4. DISTRIBUTED SIGNAL DETECTION OVER A WSN

Once the vector SMT is designed, we know the sequence in which rotations need to be performed. In order to decorrelate x, we need to compute $\tilde{x} = T^t x = T_M^t T_{M-1}^t ... T_1^t x$. Recall that each rotation T_m only depends on the outputs of the two sensors that have the lowest cost. Further, at each rotation, only the outputs corresponding to the two sensors having the lowest cost are updated. Therefore, $\tilde{x}$ is computed by communicating the outputs between pairs of sensors sequentially corresponding to the sequence of rotations. Once $\tilde{x}$ is computed, we can study the detection accuracy as explained below.

4.1 Anomaly Detection

The objective of anomaly detection is to distinguish *typical* measurements from *anomalous* ones (i.e., outliers). Let the vector

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

be an aggregated measurement from all L sensors in the network. We presume that x is *typical* if it is sampled from a multivariate Gaussian parent distribution with covariance R, i.e., $x \sim \mathcal{N}(0, R)$.

Let $p(x; R)$ be the probability density function of the multivariate Gaussian distribution with covariance R. We measure the "anomalousness" of x by performing a significance test of x against the assumed parent distribution. Therefore, for small values of $p(x; R)$, we classify x as *anomalous*.

Using these assumptions, the significance test for an anomalous event is given by

$$D_R(x) = x^t R^{-1} x > \eta^2 , \tag{8}$$

where η is chosen to control the probability of false alarm. Assuming R can be decomposed as $R = T\Lambda T^t$, the significance test is then given by

$$\widetilde{D}_\Lambda(\tilde{x}) = \sum_{i=1}^{p} \frac{\tilde{x}_i^2}{\lambda_i} > \eta^2 \tag{9}$$

where $\tilde{x} = T^t x$ are the decorrelated values at each node after the vector SMT is applied, and λ_i $(1 \le i \le p)$ are the estimated variance for each component $\tilde{x}_i$. In practice, this test statistic can be computed by summing together the weighted squared values at each sensor node after application of the vector SMT.

4.2 Detection Accuracy

Different versions of the anomaly detector use different techniques to estimate the actual covariance matrix R used in (8) . In order to compare the accuracy of two different anomaly detectors, the receiver operating characteristics (ROC) curves are widely used;[12] however, one disadvantage of the ROC is that it requires the choice of a distribution for anomalous events. An alternative approach is to evaluate the volume of the hyper-ellipsoid contained within $D_R(x) \le \eta^2$ and to use this as a proxy for the missed detection rate.[4] Such a volume is computed by

$$V(R, \eta) = \frac{\pi^{p/2}}{\Gamma(1 + p/2)} \eta^p \sqrt{|R|} . \tag{10}$$

For a fixed η (i.e., fixed false alarm rate), smaller values of $V(R, \eta)$ corresponds to fewer missed anomalies by the detector.[4] Alternatively, one can vary η to produce different probabilities of false alarm and, for each η, plot the log-volume of the ellipsoid *versus* the probability of false alarm.

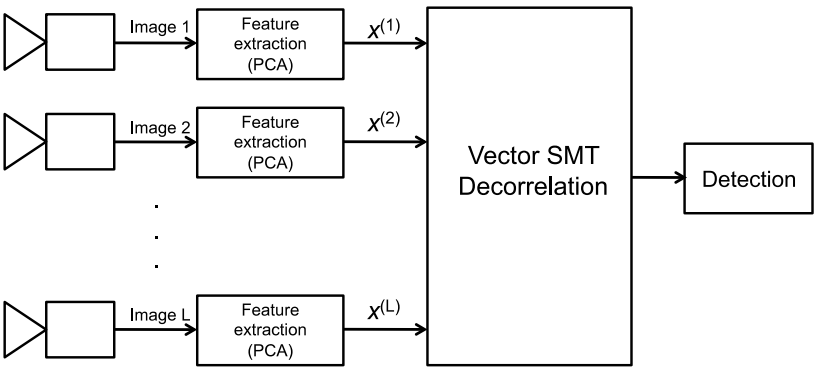


Figure 4. Block diagram for the decorrelation and detection of multiple camera images using the vector SMT.

4.3 Relative Accuracy of Two detectors

Here, we analyze the relative sensitivity of two detectors. We consider two detectors, one that uses the covariance matrix R_2 and another that uses covariance matrix R_1. Our goal is to characterize the subspace of events to which the detector using the covariance R_2 is relatively more sensitive than the one using R_1.

We start with the relative measure of sensitivity given by the ratio,

$$\frac{D_{R_2}(x)}{D_{R_1}(x)} = \frac{x^t R_2^{-1} x}{x^t R_1^{-1} x} \ .$$

Let

$$R_1 = E_1 \Lambda_1 E_1^t$$

be the eigen-decomposition of R_1 and consider the following eigen-decomposition

$$\Lambda_1^{-1/2} E_1^t R_2 E_1 \Lambda_1^{-1/2} = \widetilde{E}_2 \widetilde{\Lambda}_2 \widetilde{E}_2^t. \tag{11}$$

We define

$$H = E_1 \Lambda_1^{-1/2} \widetilde{E}_2$$

such that $HH^t = R_1$ and $H\Lambda_2 H^t = R_2$, i.e., H whitens a vector $x \sim \mathcal{N}(0, R_1)$ and decorrelates a vector $x \sim \mathcal{N}(0, R_2)$. The transform in (11) is the generalized eigendecomposition of R_1 and R_2. The diagonal elements of $\widetilde{\Lambda}_2$ are the generalized eigenvalues, and the columns of H are the generalized eigenvectors of R_1 and R_2. Therefore, rewriting the ratio above in terms of $\widetilde{x} = H^t x$, we obtain the ratio

$$\frac{\widetilde{x}^t \widetilde{\Lambda}_2^{-1} \widetilde{x}}{\widetilde{x}^t \widetilde{x}},$$

where $\frac{1}{[\widetilde{\Lambda}_2]_{kk}}$ is a measure of relative sensitivity of the detector using covariance R_2 to the dimension k of $\widetilde{x}$. We can select the set of coordinates with largest relative sensitivity and form a base for a subspace of events to which the detector with R_2 is more sensitive than the detector with R_1 using the corresponding columns of the matrix H.

5. SIMULATION EXPERIMENTS

Here, we show simulation results of anomaly detection by a network of cameras using the vector SMT, described in Section 3 , to decorrelate vector signals from multiple cameras.

Figure 5 illustrates the main steps of our simulation experiments. We begin by generating simple 3D objects in a region of the 3D space and set virtual cameras in several positions of this 3D region. Each

virtual camera renders 2D images of the same region from its own viewpoint, and uses PCA to extract the principal components of the sensed image. The final encoded output from each camera is a 10-dimensional feature vector ($h = 10$), containing the 10 principal components of largest variance.

We apply the vector SMT to decorrelate the vector outputs from the multiple cameras in the network. Finally, we use the decorrelated signal, combined from all cameras, to perform anomaly detection. The goal is to decide on whether the 3D configuration of objects observed by all the cameras is a *typical* one, or an *anomaly*.

We perform two experiments. In the first experiment, we monitor a sphere taking positions in a linear trajectory using two cameras ($L = 2$) and, in the second experiment, we monitor a cloud of spheres using 14 cameras ($L = 14$). In both cases, we search for anomalies in the configurations. We compare two different scenarios:

1. *independent processing*, when the camera outputs are considered independently, without any joint processing of the outputs from multiple cameras;

2. *joint processing*, when the camera outputs are considered jointly and *in network* decorrelation of the joint signal is applied using the vector SMT developed in Section 3.

In both experiments, we used training sets containing $N_{train} = 100$ *typical* samples to train the detectors. During testing, we used testing sets, disjoint from the training sets, containing a total of $N_{test} = 200$ samples, with 100 *typical*, and another 100 *anomalous* samples.

5.1 Moving Sphere

In this experiment, we generated a 3D sphere that is placed at random locations along the straight diagonal line indicated by the double arrow in Figure 5(a). We refer to each of these locations as a *typical* one, and to the corresponding set of captured camera images as a *typical* sample. Furthermore, we generate image samples of the sphere at locations across the mirrored line indicated by the double arrow in Figure 5(b), to be used as *anomalous* samples during the detection experiments. Two cameras monitor the sphere locations. Camera 1 has the X-Y (top) view of the scene, as illustrated in Figure 5(c), while Camera 2 captures the X-Z (side) view, shown in Figure 5(d). Notice that by looking at views shown in Figures 5(c) and (d) separately, it is impossible to distinguish an *anomalous* sample from a *typical* one.

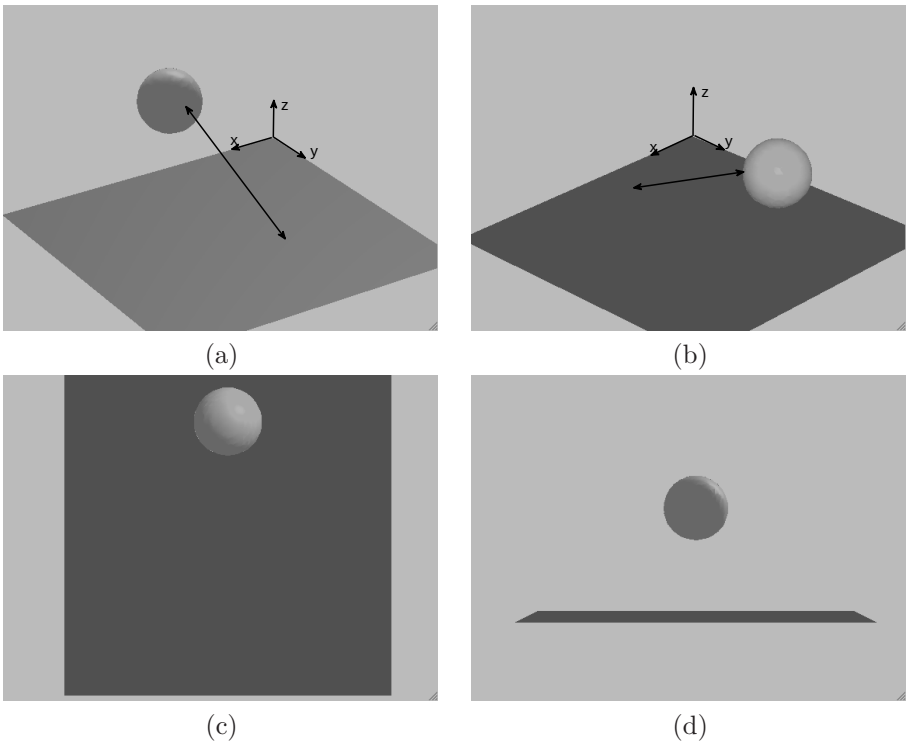
Figure 6 compares the detection accuracy of both the independent and joint ways of processing the camera outputs. Both the ROC and log-volume plots suggest that when the two images are processed independently, the system is not capable of distinguishing *anomalous* samples from the *typical* ones. However, when the images are processed jointly using the vector SMT framework, the detection is very accurate.

5.1.1 Relative Sensitivities of Detectors

As suggested by the detection accuracy experiments in Figure 6, the different ways of processing data from multiple sensors yield to anomaly detectors with different sensitivities. We use the framework in Section 4.3 to analyze the relative sensitivity of the anomaly detector that uses the vector SMT to jointly process the camera images, when compared the detector that processes the images independently (i.e., assumes the sensor outputs to be uncorrelated).

Figure 7 shows the generalized eigenvalues for both processing methods. The eigenvalues with smallest magnitude (indices 14-20) correspond to the dimensions to which the joint processing is most sensitive when compared to the independent processing.

Figures 9 and 10 show the sets of eigen-images for the independent processing, and the joint processing cases respectively, sorted in descending order according to their corresponding eigenvalues. In the independent processing case (Figure 9), the eigen-images originating from each view are independent. On the other hand, the eigen-images from the joint processing case (Figure 10) are given in pairs associated with both camera views.



(a) (b)

(c) (d)

Figure 5. Simulated 3D space region where a sphere takes random positions along the line indicated by the double arrow. (a) *typical* behavior; (b) *anomalous* behavior; (c) top view captured by the camera monitoring the X-Y dimension; (d) side view captured by the camera monitoring the X-Z dimension.

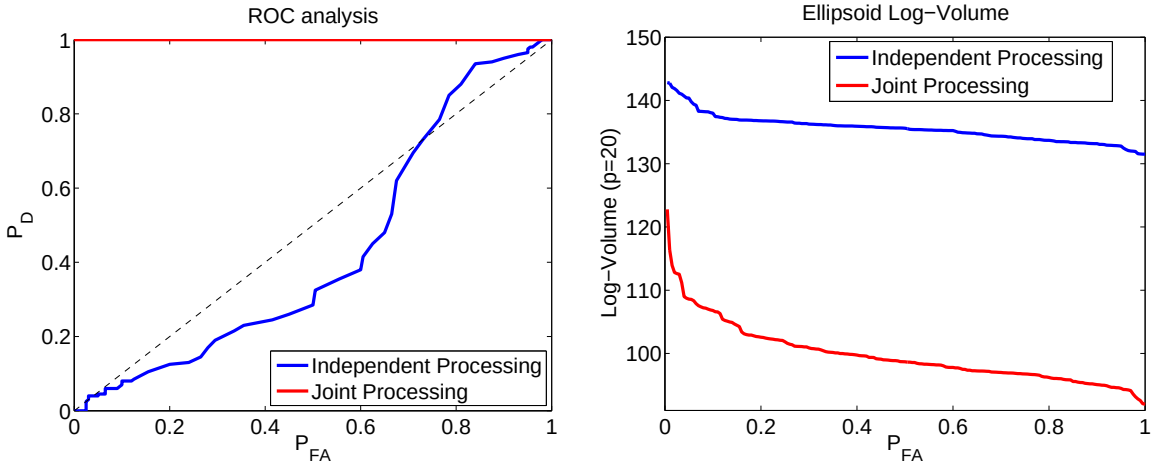
Figure 8 shows the generalized eigen-images, sorted in descending order according to their corresponding generalized eigenvalues in Figure 7. The generalized eigen-images at the bottom, associated with the smallest generalized eigenvalues, correspond to the modes which are likely to be undetectable when the independent processing is used but are likely to be detected when joint processing is used.

5.2 3D Cloud of Spheres

In this experiment, we monitor clouds of spheres that are generated so that each cloud contains 30 spheres randomly positioned in the 3D space. There are two types of clouds according to the distribution of the sphere positions: (i) *typical*: the sphere positions are generated from the $\mathcal{N}(0, I_{3\times3})$ distribution, but only the ones with distance from the origin exceeding a fixed threshold are selected, so that the resulting clouds are hollow; and (ii) *anomalous*: the random positions for the spheres are drawn from the $\mathcal{N}(0, I_{3\times3})$ distribution without further selection so that the resulting cloud is dense.

We use 14 different cameras to monitor the same 3D cloud from different viewpoints. Figure 11 shows 10 of these 14 different views of two cloud samples. The views in Figure 11(a) are from a *typical* (hollow) cloud sample, while the views in Figure 11(b) are from an *anomalous* (dense) one. By inspecting the images in Figures 11(a) and (b), we notice intuitively that it is difficult to distinguish between *typical* and *anomalous* samples by looking at the images from each viewpoint independently. Instead, the information that helps one distinguish an *anomalous* cloud from the *typical* ones is contained in the joint view of the camera images.

Figure 12 shows the anomaly detection accuracy results using both joint and independent processing of the images from different camera sensors, when these are monitoring sphere clouds. The anomaly detection with joint processing of camera images is significantly more accurate than the detection with independent processing. This is consistent with our intuition that one needs to consider the multiple views jointly in order to accurately distinguish a dense cloud from a hollow one.



(a) (b)

Figure 6. Accuracy of the independent processing *versus* the joint processing of camera images in the moving sphere experiment: (a) ROC curve; (b) log-volume of ellipsoid *versus* probability of false alarm.

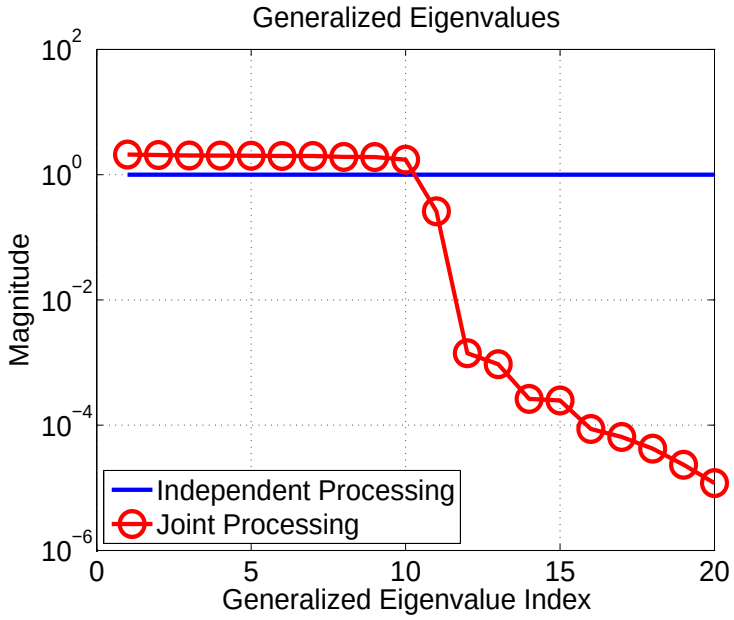


Figure 7. Generalized eigevalues of the independent and joint processing cases in the moving sphere experiment.

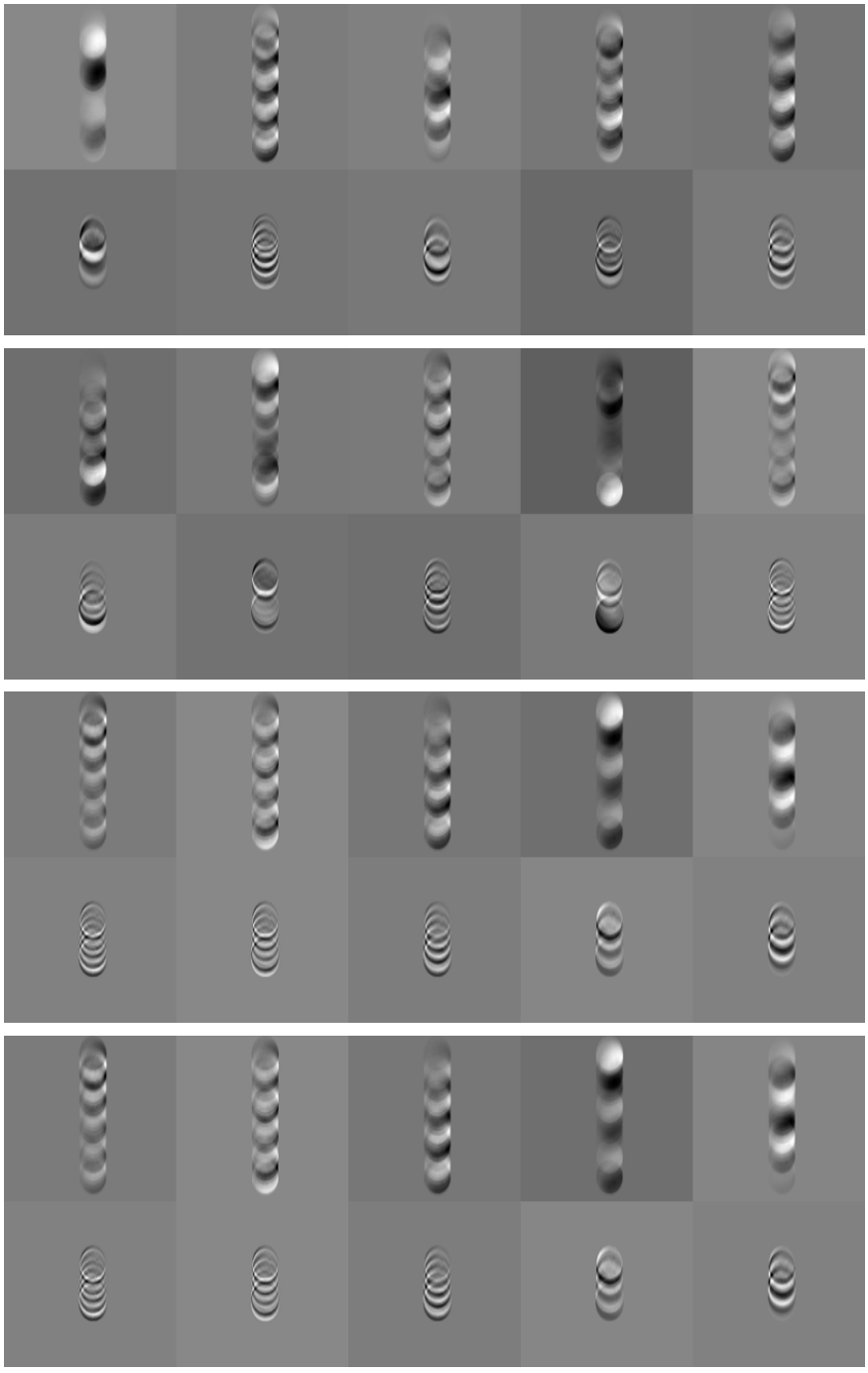


Figure 8. Pairs of generalized eigen-images associated with the joint processing and the independent processing cases in the moving sphere experiment.

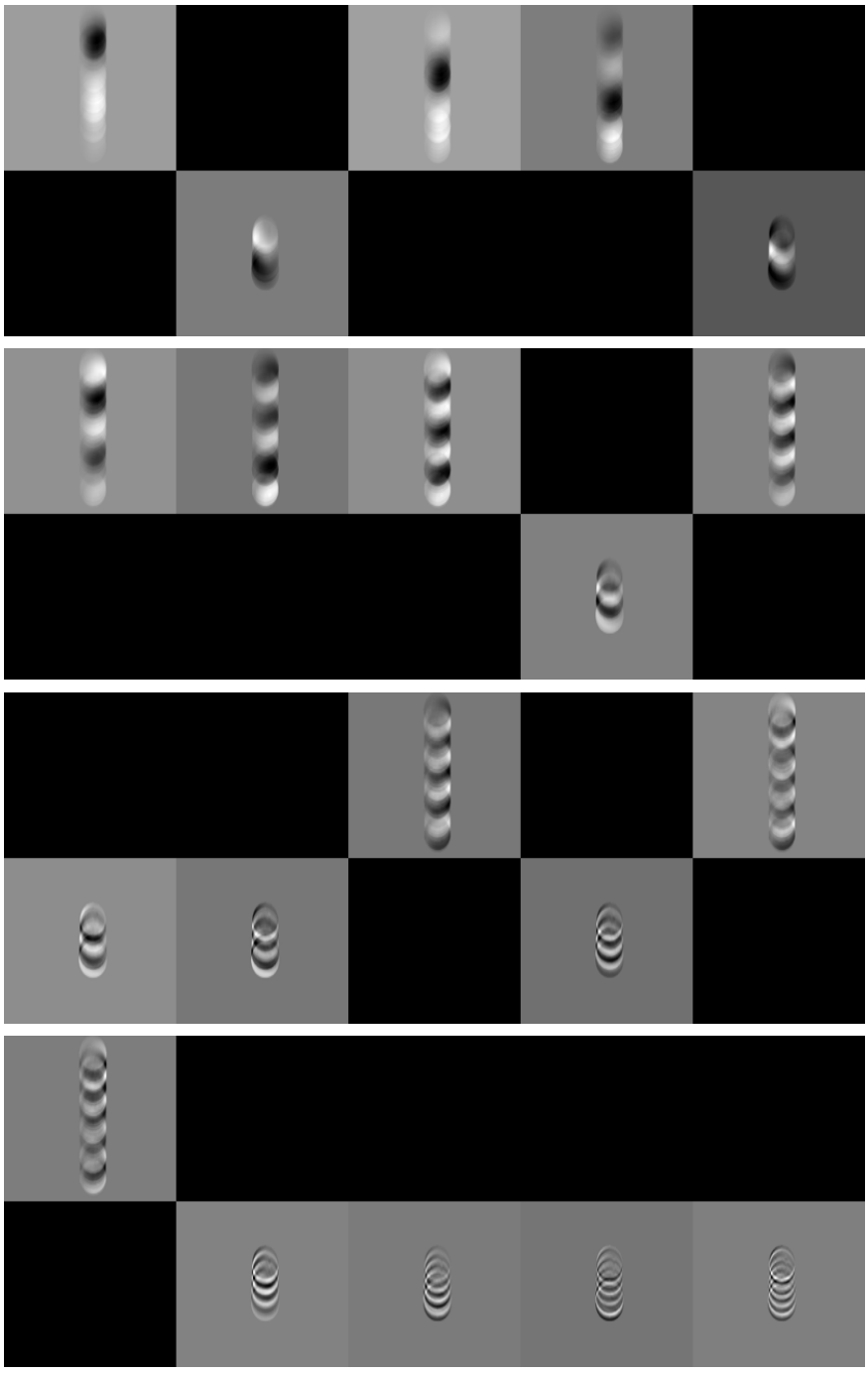


Figure 9. Eigen-images from each sensor (processed independently) in the moving sphere experiment sorted according to their corresponding eigenvalue in descending order (left-to-right and top-down).

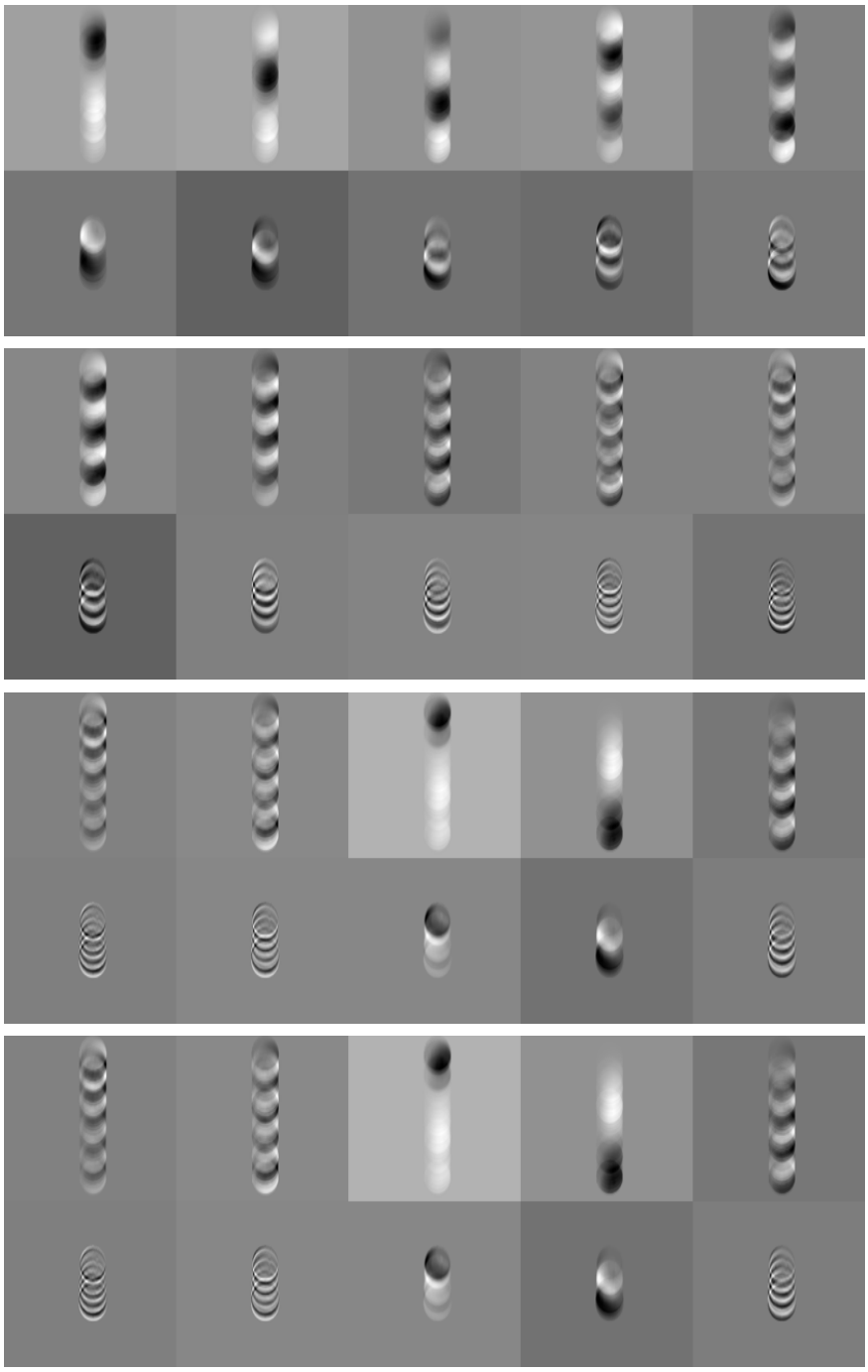


Figure 10. Pairs of eigen-images produced when the camera images are processed jointly in the moving sphere experiment. The eigen-images are sorted according to their corresponding eigenvalue in descending order (left-to-right and top-down).

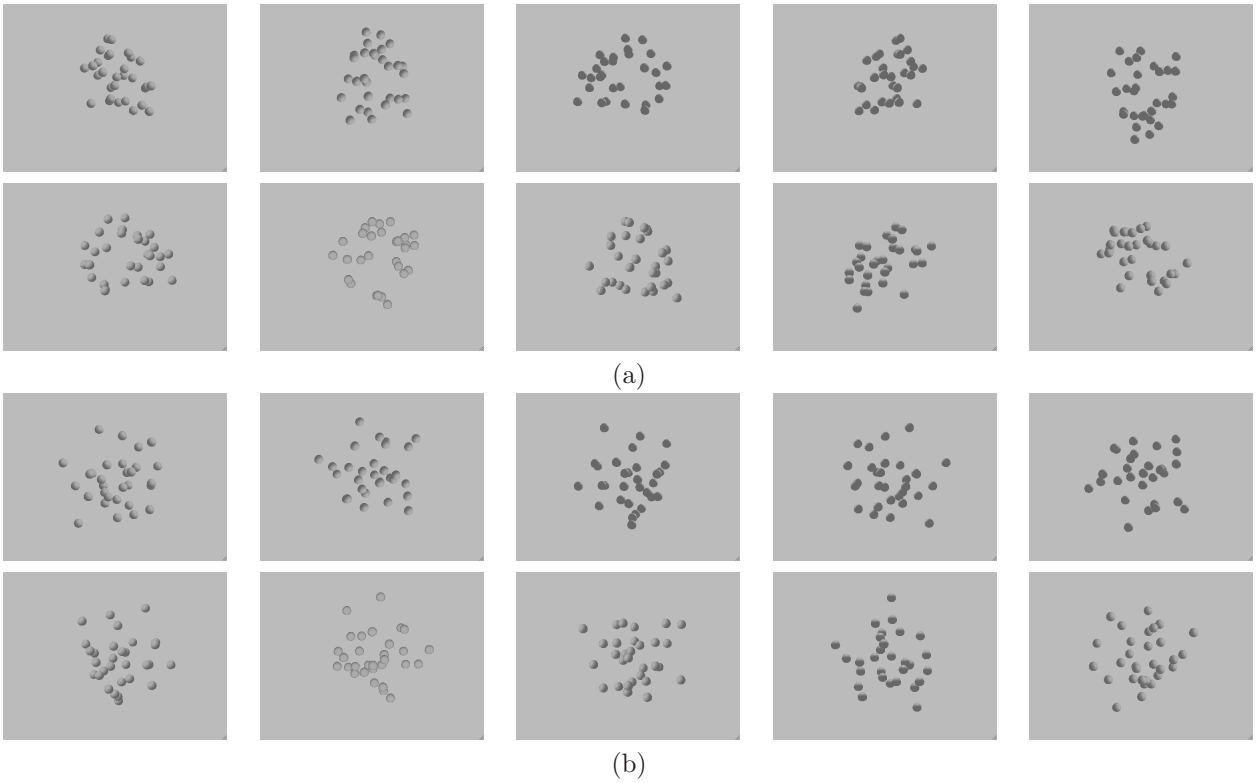


Figure 11. The different views of sphere cloud samples: (a) a hollow cloud (*typical*); (b) a dense cloud (*anomalous*).

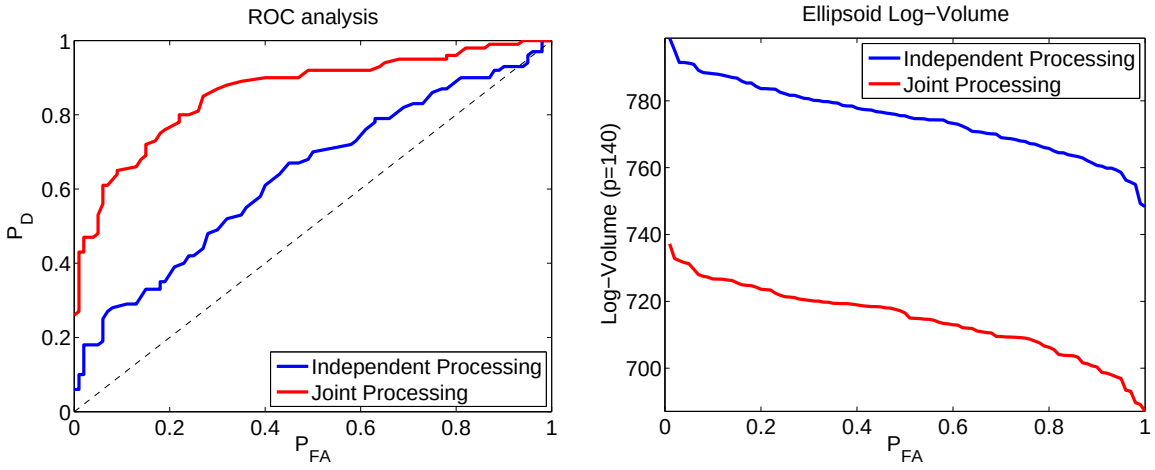


Figure 12. Results of anomaly detection in clouds of 30 spheres: (a) ROC analysis; (b) log-volume of the ellipsoid *versus* probability of false alarm.

6. CONCLUSIONS

In this paper, we presented the vector SMT framework, based on the previously proposed SMT,[1] suited for distributed analysis of high-dimensional vector measurements. These measurements are distributed across multiple sensors in a wireless sensor network. Simulation results suggest that our method has great potential to be used in distributed monitoring applications, particularly in multi-view detection of visual anomalies. In the future, we plan to analyze the communication costs associated with the vector SMT as well as to perform experiments with real application data.

REFERENCES

[1] Cao, G., Bachega, L., and Bouman, C., "The sparse matrix transform for covariance estimation and analysis of high dimensional signals," *Image Processing, IEEE Transactions on* **20**(3), 625 –640 (2011).

[2] Bachega, L. R., Bouman, C. A., and Theiler, J., "Hypothesis testing in high-dimensional spase with the sparse matrix transform," in [*The 6th IEEE Sensor Array and Multichannel Signal Processing Workshop*], IEEE (October 2010).

[3] Bachega, L. R., , and Bouman, C. A., "Classification of high-dimensional data using the sparse matrix transform," in [*Proceedings of the ICIP*], (2010).

[4] Theiler, J., Cao, G., Bachega, L., and Bouman, C., "Sparse matrix transform for hyperspectral image processing," *Selected Topics in Signal Processing, IEEE Journal of* (2011).

[5] Hariharan, S., Bachega, L., Shroff, N. B., and Bouman, C. A., "Communication efficient signal detection in correlated clutter for wireless sensor networks," in [*Proceedings of the IEEE Asilomar Conference on Signals Systems and Computers*], (2010).

[6] Ciancio, A. and Ortega, A., "A distributed wavelet compression algorithm for wireless sensor networks using lifting," in [*Proceedings of IEEE ICASSP*], (2004).

[7] Ciancio, A., Pattem, S., Ortega, A., and Krishnamachari, B., "Energy-efficient data representation and routing for wireless sensor networks based on a distributed wavelet compression algorithm," in [*Proceedings of IEEE IPSN*], (2006).

[8] Shen, G., Pattem, S., and Ortega, A., "Energy-efficient graph-based wavelets for distributed coding in wireless sensor networks," in [*Proceedings of IEEE ICASSP*], (2009).

[9] Asimovic, J., Beferull-Lozano, B., and Critescu, R., "Energy efficient decorrelating processing for data gathering in sensor networks," in [*Proceedings of IEEE IPSN*], (2005).

[10] Bachega, L. R., Cao, G., and Bouman, C. A., "Fast signal analysis and decomposition on graphs using the sparse matrix transform," in [*Proceedings of the ICASSP*], (2010).

[11] Cao, G. and Bouman, C. A., "Covariance estimation for high dimensional data vectors using the sparse matrix transform," in [*Advances in Neural Information Processing Systems*], MIT Press (2008).

[12] Kay, S. M., [*Fundamentals of Statistical Signal Processing, Vol.2: Detection Theory*], Prentice-Hall, Inc., Upper Saddle River, NJ, USA (1998).