

Hyperspectral detection algorithms: Operational, next generation, on the horizon

A. Schaum, *Member, IEEE*

Abstract—The multi-band target detection algorithms implemented in hyperspectral imaging systems represent perhaps the most successful example of image fusion. A core suite of such signal processing methods that fuse spectral channels has been implemented in an operational system; more systems are planned. Stricter performance requirements for future remote sensing applications will be met by evolutionary improvements on these techniques. Here we first describe the operational methods and then the related next generation nonlinear methods, whose performance is currently being evaluated. Next we show how a “dual” representation of these algorithms can serve as a springboard to a radically new direction in algorithm research. Using nonlinear mathematics borrowed from Machine Learning concepts, we show how hyperspectral data from a high-dimensional spectral space can be transformed onto a manifold of even higher dimension, in which robust decision surfaces can be more easily generated. Such surfaces, when projected back into spectral space, appear as enveloping blankets that circumscribe clutter distributions in a way that the standard, covariance-based methods cannot. This property may permit the design of extremely low false-alarm rate solutions to remote detection problems.

Index Terms—Hyperspectral Imaging, Kernels, Covariance Equalization, Elliptically contoured distributions

TABLE OF CONTENTS

I. INTRODUCTION	1
II. REAL TIME HYPERSPECTRAL DETECTION ALGORITHMS.	1
III. ADVANCED DETECTION ALGORITHMS	3
IV. SUMMARY.....	7
REFERENCES.....	7

I. INTRODUCTION

The US Civil Air Patrol (CAP) is in the process of deploying a fleet of VNIR (Visible/Near Infrared) airborne hyperspectral sensor systems called “ARCHER” [1] for Search and Rescue missions. The hardware designs and signal processing architecture were derived from experimental systems operated by the Naval Research Laboratory (NRL). The algorithms

used for detection were developed jointly by NRL and Space Computer Corporation (Los Angeles, CA).

The operational detection algorithms used in ARCHER are relatively simple heuristic modifications of theoretically ideal methods. Meanwhile, research at NRL has developed more sophisticated algorithms that outperform the deployed ones. These methods can be generated by incorporating into an hypothesis test generalizations of the simplistic target and background models [2] that are associated with the simpler algorithms.

One indication that superior performance might be expected of these newer techniques is the shapes of the decision surfaces that they produce in hyperspectral space. They are more selective than those of the standard algorithms, and arise principally as a consequence of a single assumption: That the second-order statistics of background- and target-distributions differ.

This paper first describes the standard techniques. It then proposes a more general detection approach using elliptically contoured distributions (ECDs). These generalize multivariate Gaussians in a simple way, but they can produce selective surfaces without the need to define separate target and background statistics.

Finally, the paper describes how an alternative “dual” representation of the vector spaces used to describe these algorithms defines a convenient theoretical platform for launching into a much larger class of selective decision surfaces. These are generally associated with “kernel” methods of discrimination in the literature on *Machine Learning*.

II. REAL TIME HYPERSPECTRAL DETECTION ALGORITHMS

In ARCHER, an onboard signal processor accepts blocks of hyperspectral data that are used to estimate radiance statistics of the physical background (usually located beneath an airborne collection platform). These statistics are required for constructing all the onboard real time detection algorithms. Estimates

$$C = \frac{1}{N} \sum_{i=1}^N x_i x_i^T \quad (1)$$

of the mean and covariance matrix describing the background distribution of the data are constructed from a block of N background training vectors, with x_i the sensor-derived estimate of spectral radiance at pixel i . The matrix transpose operation is denoted by \dagger .

Each component of x_i in this vector notation corresponds to a different discretized wavelength. The operational dimensionality is typically at least several dozen, although this value usually results from a local spectral integration of the several hundred bands collect by a hyperspectral sensor.

The approximations in Equation (1) are the maximum likelihood estimates of the first- and second-order statistics of a Gaussian model of the background probability density function (pdf)

$$p_B(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{\|C\|^{N_c/2}} \exp\left\{-\frac{1}{2}(x - \mu)^T C^{-1}(x - \mu)\right\}, \quad (2)$$

with $\|C\|$ the determinant of C and N_c the number of colors.

Using these statistical estimates, two algorithms are currently implemented, called ‘‘RX’’ and ‘‘LMF.’’ These are representative of two general classes of algorithm: anomaly detection and signature-based detection. We refer to RX and LMF as the ‘‘standard’’ algorithms.

The RX Anomaly Detector

The standard (RX) anomaly detector is a test for rare pixels, as determined by the model pdf of Equation (2). Unusual pixel values—corresponding to a pdf value less than some threshold value—are labeled *targets*. This decision is logically equivalent to a test of the form

$$x^T C^{-1} x > k \text{ threshold} \quad (3)$$

for deciding that a pixel with radiance vector x contains a target. The values of μ and C are defined in Equation (1); no distinction is made between their estimated and ‘‘true’’ values. The spectral region corresponding to a target declaration is the exterior of a hyperellipsoid (Figure 1) defined by Equation (3).

The Linear Matched Filter

The standard signature-based detector is the liner matched filter (LMF). A mean value t of the target signature is as-

sumed to be available for use in a detection algorithm. Knowing an approximate value of the target signature allows the construction of a more selective decision surface than defined by Equation (3), to wit, the hyperplane of Figure 1.

For the LMF, a target is declared at any pixel x satisfying

$$t^T C^{-1} x > k. \quad (4)$$

The corresponding decision boundary hyperplane is perpendicular to the vector $C^{-1} t$. To see why this choice of hyperplane is a good one, it is convenient to view the data in a transformed representation of radiance space.

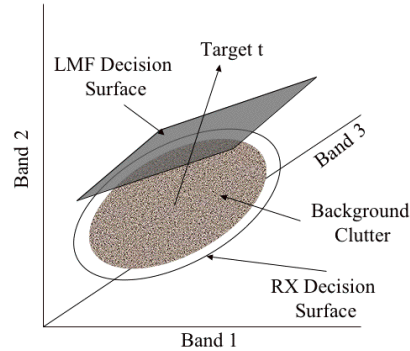


Figure 1. Gaussian clutter model with decision surfaces for anomaly detection and for the linear matched filter.

Whitened Formulation

The RX and LMF detectors admit simple geometrical interpretations. In spectral space their decision surfaces account for the non-sphericity of the background distribution. The RX decision boundaries (Figure 1) are nothing but the surfaces of constant background pdf (see Equation (2)). And instead of using the hyperplane orthogonal to the line between background and target means, the LMF decision surface has a normal that is rotated (by the action of the matrix C^{-1}) to account for a possible asymmetry in the shape of the background distribution relative to $(t - \mu)$, the vector connecting target and background means.

An intuitive explanation of these algorithms is easier if the asymmetry is removed. This can be accomplished by a ‘‘whitening’’ transformation of the radiance variable x

$$w = C^{-1/2} x. \quad (5)$$

The square root matrix $C^{-1/2}$ of the covariance matrix is defined by

$$C^{-\frac{1}{2}} D^{-1} t, \quad (6)$$

where

$$C = D^2 t^\dagger \quad (7)$$

is a singular value decomposition (SVD) of the sample background matrix C , t^\dagger is an orthonormal matrix (of eigenvectors), t^\dagger is its adjoint, and D is the non-negative diagonal matrix of (root) eigenvalues of C . Viewed in the whitened variable w , instead of the radiance variable x , the collection of background pixels has mean value zero and, more importantly, unit variance in any direction. This latter characteristic makes the particular forms of RX and LMF algorithms obvious choices, as discussed below.

The RX criterion (Equation (3)) for choosing “target” as the true hypothesis is equivalent to:

$$w^2 > k, \quad (8)$$

in the whitened space, that is, the Euclidean length of the vector w is compared to a threshold.

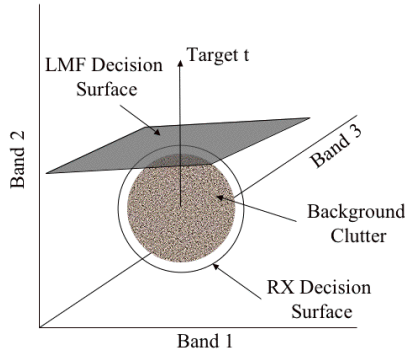


Figure 2. Simplified geometry in whitened variables. The coordinate origin is actually at the center of the sphere.

Similarly, in whitened variables the LMF statistic (Equation (4)) becomes a Euclidean inner product

$$w_t = w \cdot k, \quad (9)$$

with the whitened target signature given by

$$w_t = C^{-\frac{1}{2}} t. \quad (10)$$

These algebraic simplifications arise because the zero-mean variable w has a covariance matrix equal to the identity:

$$I = \frac{1}{N} \sum_{i=1}^N w_i w_i^\dagger \quad (11)$$

When viewed in terms of the variable w , Figure 1 transforms into Figure 2, in which the background distribution appears more spherically symmetric. According to Equation (8), anomalies are simply any pixels situated a sufficient distance from the mean of the background distribution.

Because W is a linear transformation, it maps the hyperplane of Figure 1 into another hyperplane in Figure 2. But in the whitened space, the target mean vector w_t defines the only preferred data-driven direction, to which the hyperplane decision surface must therefore be orthogonal. Equation (9) is the algebraic expression of this geometrical interpretation of the LMF decision surface. It represents a (scaled) projection of the test vector onto the target direction, which defines a one-dimensional preferred subspace of the complete whitened space.

III. ADVANCED DETECTION ALGORITHMS

The LMF arises as the optimal solution to a detection problem. However, the simplicity of its planar decision surface depends strongly on the assumption that background and target second-order statistics are identical. That is, although the mean values of these two distributions can be different, they are assumed to have identical covariance matrices.

Joint Subspace Detection

A family of techniques that produce more discriminating surfaces than a hyperplane have been studied for several years. Joint Subspace Detection (JSD) is a detection concept [3] that is motivated by the frequently found experimental fact that subspace versions of RX or LMF often work better than the baseline versions described above. For example, the real time version of RX that is most commonly implemented operationally is SSRX (Subspace RX), which is RX applied to all but the first several higher-variance dimensions (as determined by an SVD analysis) of the spectral representation of the background.

This modification improves performance because targets typically appear (in a spectral space representation of the data—as, for example, in Figure 1) embedded deeply within the higher-variance dimensions of the background clutter. RX applied in only these dimensions would produce worse performance than a coin flip. (Anti-RX—gotten by adding a minus sign to the left-hand side of Equation (8)—would, on the other hand, produce a nontrivial detector.)

One branch of the JSD family of detectors is represented by “Confined Target Models” (CTMs). These can generate quadratic decision surfaces (Figure 3) in place of hyperplanes, thereby greatly enhancing the selectivity of hyperspectral de-

tection algorithms. The general form of the CTM detector statistic in whitened space is

$$D x = w^\dagger w - w^\dagger M_{eff}^{-1} w, \quad (12)$$

with w given by Equation (5).

The first term is the standard RX anomaly detector statistic (see Equation (8)). The second depends on the model details through M_{eff} . The negative sign reflects the inclusion of an anti-RX term. It reflects generally the idea that high-variance background dimensions which, by themselves, are not competitive with low variance ones in enabling target detection, are nonetheless useful.

Dimensions that degrade RX performance, which are deleted by SSRX, are kept and exploited by JSD. A principal lesson taught by the JSD algorithms is: Generally the only situation when sensed dimensions should be ignored by a detection algorithm is if the distributions of target and background pixels in those dimensions are identical. Large differences can be exploited, even when clutter dominates.

One drawback to the current incarnations of JSD is the lack of a consistent method for estimating the target parameters needed to compute the optimal M_{eff} . This could indicate that it is much more important to use the kind of selective surfaces (such as the quadric) for detection that JSD produces, than to divine the correct target distribution model used to generate them.

One final note on JSD detectors. Generally when different variabilities are associated with background and target distributions, the decision surfaces can include “ghost surfaces” associated with the mirror branch of a hyperboloid. They occur in dimensions for which the background distribution is less variable than the target distribution. For example, in the Gaussian case, for any given threshold setting, there can be pixels on the far side (opposite the target mean t) of the background distribution (in such dimensions) that a likelihood ratio test will assign the label *target*. These regions are correct theoretical predictions, so long as the distribution models are strictly correct, but one is loath in practice to associate real targets with such distant, trans-background regions of spectral space.

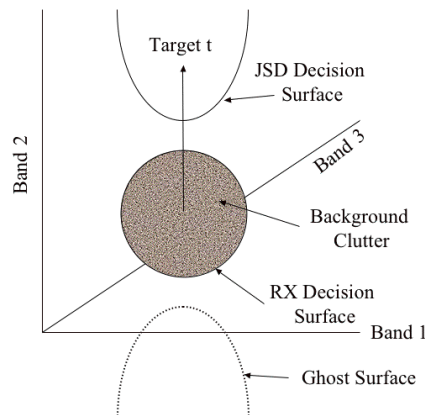


Figure 3. Advanced detection algorithms produce more selective, quadric decision surfaces. JSD surfaces can include ghost surfaces, corresponding to the mirror branch of a hyperboloid.

ECD Detectors

If the radiance variable x were truly Gaussian distributed, then the whitened form w , being derived from x by a linear transformation, would also be Gaussian. And with identity covariance matrix, the stochastic variable w would also be distributed spherically symmetrically. The general form of Gaussian distribution (Equation (2)) would then reduce to the product of N_C one-dimensional Gaussians in the variables w_i . It is then easy to show that the distribution of the RX statistic, $w^\dagger w$, should be chi-squared with N_C degrees of freedom. This has been observed for typical hyperspectral data sets to give a poor fit [4].

Non-Gaussianity is an important concern in devising CFAR (constant false alarm rate) detection algorithms. Better modeling, especially of the tails of the measured distributions is vital for creating a truly autonomous detection system. One method of improved modeling allows a generalization of the Gaussians form to one called elliptically contoured distributions (ECDs).

Here we exploit ECDs for the purpose of target detection.

The most general form of ECD has pdf

$$p(x) = f(x^\dagger C^{-1} x), \quad (13)$$

with C some arbitrary invertible matrix. It can be shown that any such matrix appearing as in Equation (13) must be proportional to the covariance matrix (whose sample value is given in Equation (1)). Therefore, we assume without loss of generality that the C of Equation (13) is the covariance matrix of the background data. Consequently, we may take the distribution to be

$$p(x) = f(w^\dagger w), \quad (14)$$

with w given by (5).

The whitened Gaussian is an example of an ECD:

$$p_B(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}w^\dagger w\right). \quad (15)$$

This suggests a natural generalization:

$$p_B(x) = k_p \exp\left(-\frac{1}{2}\|w\|^p\right), \quad (16)$$

with the choice of parameter $p = 2$ corresponding to the standard Gaussian. The normalization constant k_p is chosen to make the pdf integral unity.

Assuming the standard “additive target” model, the pdfs for background and target are related:

$$p_T(x) = k_p \exp\left(-\frac{1}{2}\|w - w_t\|^p\right). \quad (17)$$

When the target and background distributions are both known, so also is the optimal detector, which is the (log) likelihood ratio test:

$$\ln p_T(x) - \ln p_B(x) > k, \quad (18)$$

which condition dictates the declaration “target” at pixel x .

The algorithms described in (16) through (18) comprise a one-parameter family of ECD detectors. For example, for $p = 1$, the detector is equivalent to:

$$\|w\| \geq \|w - w_t\| + k. \quad (19)$$

The bounding surface between decision regions is defined by (19) (with equality replacing the inequality), which is a classic definition of a hyperbola in two dimensions, and a hyperboloid of revolution in higher dimensions. The foci are located at $w = 0$ and $w = w_t$. The detector therefore produces a decision surface similar to the type shown in Figure 3 for JSD detectors. There is, however, no ghost.

Furthermore, while this type of algorithm can produce the kind of target/background decision surface usually associated with JSD methods, the selectivity of that shape for ECDs is not dependent on any differences between the modeled target and background distributions. It can be expected that the simplifying assumption of equal target and background covariance matrices—which also underlies the generation of the classic RX and LMF algorithms—will suffice for producing JSD-type performance in ECD detectors. This would obviate the need to specify the large number of degrees of freedom that can be associated with JSD methods and pave the way for the operational implementation of a new class of highly selec-

tive detectors. Tests will be conducted in the next year to ascertain the relative performance of ECD and JSD detectors on data collected by NRL.

Detection in Feature Space

There is an alternative formulation of the multivariate detection concepts discussed above that is usually of no practical value. It can, however, serve as a springboard to a new class of detection algorithms.

If we define the sample data matrix as

$$X = [x_1 \ x_2 \ \dots \ x_N], \quad (20)$$

with N the number of sample points, then the covariance matrix (Equation (1)) can be written (apart from an unimportant factor $1/N$)

$$C = XX^\dagger. \quad (21)$$

Let

$$X = D\Lambda^\dagger \quad (22)$$

be an SVD of the matrix X . The diagonal matrix D has dimension $N_d \times N_d$, with N_d equal to the rank of X , also equal to the number of nonzero eigenvalues of the matrix C . The maximum value that N_d can assume is the smaller of the row- and column-dimensions of the matrix X .

In hyperspectral applications, the number of samples N is usually much greater than the number of colors N_C . In this case usually N_d equals N_C . (For this not to be the case, the N data samples would all have to lie in a subspace of dimension $N_d < N_C \ll N$.) The columns of the matrices A and F are the left and right unit-length eigenvectors of X , and so these matrices always have left inverses equal to their adjoints. If furthermore, $N_d = N_C$, then Λ is square, and so is invertible, which we assume to be the case. (Otherwise, from (21) and (22), $C = AD^2A^\dagger$ would not be invertible, and RX and LMF would not be computable.)

We can solve (22) for Λ :

$$X = D\Lambda^\dagger. \quad (23)$$

With the coordinate origin chosen to coincide with the mean of the background, and using Equations (21)–(23), the RX statistic (Equation (3)) becomes

$$x^\dagger D^{-2} x = X^\dagger x^\dagger D^{-4} X^\dagger x. \quad (24)$$

Like the covariance matrix, the Gram matrix is defined in terms of the sample data:

$$G = X^\dagger X. \quad (25)$$

Note from Equation (22) that

$$G = D^2 \Gamma. \quad (26)$$

Therefore, if we have diagonalized G (i.e., we have computed its SVD form, Equation (26)), then both D and Γ are available for use in Equation (24).

Usually, in order to compute the RX statistic, it is much easier to invert the smaller $N_C \times N_C$ matrix C (and use Equation (3)), than the $N \times N$ matrix G (and use Equation (24)). The dimension N , which is the number of data samples x_i used to estimate the covariance matrix, numbers typically in the thousands, while N_C , the number of processed colors, is typically a few dozen. Therefore, conventional RX or LMF hyperspectral processing has no need to compute (or invert) the large Gram matrix.

Notice, however, that G depends on the sample vectors only through inner products. From Equations (25) and (20), its i, j component is

$$G_{ij} = x_i \mathbf{g}_j, \quad (27)$$

with \cdot denoting the ordinary Euclidean inner product. Finding the covariance matrix C in (21), on the other hand, requires calculating the *outer* products of sample vectors, which requires knowledge of each vector component.

The so-called *dual* representation of the test vector, $X^\dagger x$, which appears in Equation (24), also requires knowledge of only inner products, of x with the sample vectors x_i :

$$X^\dagger x = x_i \mathbf{g}_i. \quad (28)$$

The inner-product dependence of the dual representations of the test pixel and data matrices permits a radically new interpretation of the vectors x .

Let r (instead of x) represent the radiance vector of the data measured by a sensor, and consider a general transformation from radiance to a so-called *feature vector* x

$$r = x. \quad (29)$$

Suppose further that inner products in the feature space can be computed from the corresponding radiance vectors, using a *kernel* [5] function

$$k(r, r) = r \mathbf{g} r = x \mathbf{g} x. \quad (30)$$

One can then apply RX or LMF in this feature space, instead of in radiance space. The advantage of the dual representation is that it allows this without requiring explicit knowledge of the transformation in Equation (29). One need only define a kernel $k(r, r)$ for which the usual properties of inner products

hold, such as Schwartz's inequality. This means that large classes of transformations, and therefore many feature space transformations ϕ , can be investigated within a single analysis.

As an example of why one might want to pursue this approach, consider the transformation

$$r = [r_1, r_2, \dots, r_{N_C}, \sqrt{2}r_1r_2, \sqrt{2}r_1r_3, \dots, \sqrt{2}r_{N_C-1}r_{N_C}]^\dagger x, \quad (31)$$

which corresponds to the choice of kernel

$$k(r, r) = r^\dagger r^2. \quad (32)$$

The covariance matrix C of x now encodes (cf. Equations (20), (21), and (31)) not only the second-order statistics, but also third- and fourth-order statistics of the radiance distribution.

The application of conventional detectors like RX and LMF in the feature space produces the usual kinds of ellipsoidal or planar decision surfaces in that $N_C(N_C+1)/2$ -dimensional space. But when transformed back into the N_C -dimensional radiance space, the surfaces assume more complex shapes that can better segregate non-Gaussian sample data from target spectra. These surfaces are generated using moments of radiance up to fourth-order, rather than just first- and second-order ones.

As a second example of these "feature space" methods, a popular kernel choice ("radial basis functions") uses the Gaussian:

$$k(r, r) = e^{-\frac{\|r - r\|^2}{2\sigma^2}}, \quad (33)$$

with σ a free parameter. This corresponds to a mapping ϕ into an infinite-dimensional feature space. It thus has the potential for exploiting background distribution moments of *all* orders.

It is instructive to examine the form the matched filter takes for this choice of kernel function. Analogously to the RX form in Equation (24), the (no longer linear) Matched Filter becomes

$$\begin{aligned} X^\dagger t^\dagger D^{-4} X^\dagger x &= X^\dagger t^\dagger D^{-4} x_i \mathbf{g}_i \\ &= X^\dagger t^\dagger D^{-4} k(r_i, r) \\ &= X^\dagger t^\dagger D^{-4} e^{-\frac{\|r_i - r\|^2}{2\sigma^2}}, \end{aligned} \quad (34)$$

where t is the target mean radiance, r is the radiance of the test pixel, and r_i is the radiance of the i^{th} training sample.

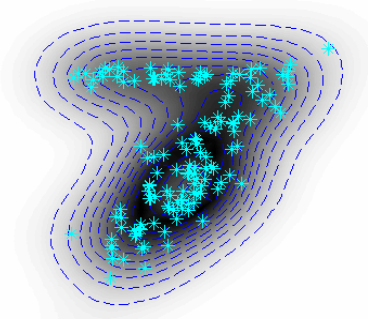


Figure 4. Decision surfaces of kernelized RX are better able to track contours of the data than conventional RX. [Plot courtesy Nasser Nasrabadi, Army Research Lab, Adelphi, MD.]

This feature space version of the matched filter is thus a sum of weighted Gaussian functions, each centered on a test pixel. As such, it has much in common with nonparametric methods of discrimination. A standard nonparametric method would convolve the Gaussian function in (33) over the entire background sample space, creating contoured decision surfaces that can track the boundary of an arbitrary confined background spectral distribution. The kernel matched filter does this also, with weight values influenced by the target signature. Similar comments apply to the kernel version of RX. An example of the contours it can produce is shown in Figure 4.

The principal drawback of these kernel methods, in their fundamental forms, is a computational burden comparable to that associated with nonparametric methods. Large ($N \times N$) matrices need to be inverted. Automatic selection of the scale parameter σ can also be problematical. The principal advantage of kernel methods is expected to be robustness of performance, a characteristic inherited from the robustness of the simple forms their decision surfaces assume in feature space.

IV. SUMMARY

We have reviewed the core detection algorithms implemented in the first operational hyperspectral detection systems. Next we described some of next generation nonlinear methods that will be tested with data derived from ongoing experimental sensor system programs. Borrowing from Machine Learning concepts, we have also shown how a “dual” formulation of these algorithms can generate a large class of nonlinear decision surfaces. Viewed in spectral radiance space, these surfaces should circumscribe clutter distributions in a form-fitting way that the standard, covariance-based methods cannot. Quantitative performance and implementation studies are beginning.

REFERENCES

- [1] Brian Stevenson et al., Design And Performance of the Civil Air Patrol ARCHER Hyperspectral Processing System, *Algorithms and Technologies for Multispectral, Hyper-spectral, and Ultraspectral Imagery XI*, Proc. of SPIE Vol. 5806 (SPIE Bellingham, WA), 28 March – 1 April 2005, pp. 731-742.
- [2] A. Schaum, Joint Subspace Detection Applied to Hyperspectral Target Detection, Proc. 2004 IEEE Aerospace Conference, IEEE Catalog Number: 04TH8720C ISBN: 0-7803-8156-4, 10 March.
- [3] A. Schaum, Alan Stocker, Joint Hyperspectral Subspace Detection Derived from a Bayesian Likelihood Ratio Test, Proceedings of SPIE Vol. 4725, *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery VIII*, pp. 225-233, April, 2002.
- [4] D. Manolakis, M. Rossacci, Statistical Characterization of Natural Hyperspectral Backgrounds using t-elliptically Contoured Distributions, Proceedings of SPIE Vol. 5806, *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI*, pp. 56-65, March, 2005.
- [5] B. Scholkopf, A. Smola, K.R. Muller, Nonlinear component analysis as a kernel eigenvalue problem, *Neural Comput.* 10 (5) (1998), pp. 1299-1319.