THE FULLY ADAPTIVE GMRF ANOMALY DETECTOR FOR HYPERSPECTRAL IMAGERY

Susan M. Thornton and José M.F. Moura

Carnegie Mellon University, ECE Department
Pittsburgh, PA 15237
moura@ece.cmu.edu

ABSTRACT

The use of hyperspectral imagery for remote sensing detection applications has received attention recently due to the ability of the hyperspectral sensor to provide registered information in both space and frequency. However, this coupling of spatial and spectral information leads to an immense amount of data for which it has proven difficult to develop an efficient implementation of the Maximum-Likelihood (ML) detector. In this paper we present the Gauss-Markov random field (GMRF) detector which we have developed for detecting man-made anomalies in hyperspectral imagery. The GMRF detector is the first computationally efficient ML-detector for hyperspectral imagery. We compare the detection performance and the computational requirements of our detector implementation to the benchmark RX detection algorithm for hyperspectral imagery.

1. INTRODUCTION

In this paper we address the issue of detection of man-made objects using airborne hyperspectral imagery. Hyperspectral sensors collect hundreds of narrow and contiguously spaced spectral bands of data organized in the so called hyperspectral cube. The hyperspectral imagery provides fully registered high resolution spatial and spectral information that is invaluable in discriminating between man-made objects and natural clutter backgrounds. This comes at a cost. The high volume of data in the hyperspectral cube has precluded the development of computationally practical Maximum-Likelihood (ML) detectors of man-made anomalies in clutter. Our focus is on the "inflight" processing of the data, so computational considerations are of extreme importance.

In this paper, we present the Gauss-Markov random field (GMRF) detector, a computationally efficient ML anomaly detector that fully adapts to the unknown statistics of the clutter, and fully exploits the spatial and spectral correlation of the hyperspectral imagery. We show that the GMRF detector is significantly simpler computationally than the benchmark anomaly detection algorithm - the RX method [1]. The GMRF approach avoids the costly step of inverting the large sample covariance matrix of the clutter which is the limiting factor in other ML-implementations. We present new detection results using real hyperspectral imagery from the SEBASS and HYDICE hyperspectral sensors which highlight the fully adaptive GMRF detector as a promising alternative to the benchmark algorithm.

This work was supported by the ONR and an Amelia Earhart Fellowship from Zonta International. Previous work by the author Susan M. Thornton is under her maiden name, Susan M. Schweizer

2. THE FULLY ADAPTIVE GMRF DETECTOR

2.1. The GMRF Clutter Model

The GMRF clutter model captures the spatially-spectrally correlated nature of the background in hyperspectral sensor data through two important properties: Markovianity and noncausality. Considering the hyperspectral data cube as a 3-D finite lattice, the intensity at each pixel location in the lattice is referenced by the variable x and three subindices i,j, and k, which indicate the spatial location and the particular spectral band in which the pixel lies. Processing is done on small regions of the data set in which it is valid to assume that the clutter is homogeneous. The processing region is further divided into sub-lattices of size $N_i \times N_j \times N_k$ where N_k is equal to the total number of available spectral bands. We refer to these sub-lattices as Markov windows. Each of the Markov windows is lexicographically ordered to form a set of independent data vectors.

Within each Markov window, we let x_{ijk}, $1 \leq i \leq N_i$, $1 \leq j \leq N_j$, $1 \leq k \leq N_k$ represent a 3-D finite lattice field. The intensity, x_{ijk}, of each clutter pixel is described by an extension of the minimum mean square error (MMSE) representation of Woods [2]. For simplicity of the presentation, we adopt a first-order, homogeneous, noncausal GMRF described by

$$\begin{aligned} x_{ijk} &= \beta_h(x_{i(j-1)k} + x_{i(j+1)k}) + \beta_v(x_{(i-1)jk} + x_{(i+1)jk}) \\ &+ \beta_s(x_{ij(k-1)} + x_{ij(k+1)}) + \epsilon_{ijk} \end{aligned} \quad (1)$$

The parameters β_h, β_v, and β_s are the MMSE predictor coefficients for the spatial and spectral dimensions, respectively, and ϵ_{ijk} is the prediction error. At the edges of the Markov window we assume zero Dirichlet boundary conditions [3]. The noise field, ϵ_{ijk}, is correlated, with correlation structure discussed below . Equation (1) corresponds to a first order 3-D Markov model. It is assumed, without loss of generality, that the clutter is zero mean. In practice, the spatially varying mean is locally estimated and removed from the data.

Using equation (1), the data within a Markov window can be compactly represented by the matrix-vector equation $A\underline{X} = \underline{\epsilon}$ where we use the Kronecker product [4] to represent the matrix A in a concise manner,

$$A = I_{N_k} \otimes I_{N_i} \otimes B + I_{N_k} \otimes H^1_{N_i} \otimes C + H^1_{N_k} \otimes I_{N_i} \otimes D \quad (2)$$

The matrices B, C, and D are themselves structured and defined as

$$B = -\beta_h H^1_{N_j} + I_{N_j} \quad (3)$$

$$C = -\beta_v I_{N_j} \quad (4)$$

$$D = -\beta_s I_{N_j} \quad (5)$$

The symbols I_{N_k},I_{N_j}, and I_{N_i} are identity matrices, while $H^1_{N_k}$, $H^1_{N_j}$, and $H^1_{N_i}$ are Toeplitz matrices which have zeros everywhere except for the first upper and first lower diagonals which are composed of all 1's. The subscript denotes the size of the matrices.

The matrix A, referred to as the potential matrix, is a sparse block tridiagonal matrix and contains all the relevant information regarding the GMRF structure [3]. The error vector, $\underline{\epsilon}$ is a sample from a colored noise process with covariance $\Sigma_\epsilon = \sigma^2 A$. This leads to the direct parameterization of the inverse of the clutter covariance matrix, Σ_x^{-1},

$$\Sigma_x^{-1} = \frac{1}{\sigma^2} A \quad (6)$$

Using equations (2)- (5) in equation (6),the inverse covariance matrix of the field $\underline{X}$ is expressed in Kronecker notation as:

$$\Sigma_x^{-1} = \frac{1}{\sigma^2} I_{N_k} \otimes I_{N_i} \otimes I_{N_j} - \frac{\beta_h}{\sigma^2} I_{N_k} \otimes I_{N_i} \otimes H^1_{N_j} - \frac{\beta_v}{\sigma^2} I_{N_k} \otimes H^1_{N_i} \otimes I_{N_j} - \frac{\beta_s}{\sigma^2} H^1_{N_k} \otimes I_{N_i} \otimes I_{N_j} \quad (7)$$

This parameterization is a function of the four scalar parameters, $\sigma^2, \beta_h, \beta_v$, and β_s. When using real hyperspectral imagery, the clutter model must be adapted to the data by estimating these four parameters.

2.2. Clutter Adaptation

In previous work, we have presented and analyzed three estimation approaches for clutter adaptation [5]. The three methods we have considered are the optimal Maximum-Likelihood (ML) approach, a classical Least Squares (LS) approach, and an approximate-ML (AML) approach. In [5], we show that both the AML and LS estimation methods provide similar detection performance when incorporated into our GMRF modeling framework, and represent desirable alternatives to the optimal ML-estimator since they avoid a nonlinear optimization. Since we have also shown in prior work [5] that the AML estimation algorithm is computationally the best of the three procedures considered, it is the estimator that we use for the work in this paper.

The AML estimation method uses a simple mathematical approximation to the Log Likelihood function (LLF) that is associated with the hyperspectral clutter data. For an in depth derivation of the AML estimates see [5]. The resulting AML estimates are,

$$\widehat{\beta_h} = \frac{\xi \chi_h}{|\chi_h| \cos\left(\frac{\pi}{(N_j+1)}\right) + |\chi_v| \cos\left(\frac{\pi}{(N_i+1)}\right) + \alpha |\chi_s| \cos\left(\frac{\pi}{(N_k+1)}\right)}$$

$$\widehat{\beta_v} = \frac{\xi \chi_v}{|\chi_h| \cos\left(\frac{\pi}{(N_j+1)}\right) + |\chi_v| \cos\left(\frac{\pi}{(N_i+1)}\right) + \alpha |\chi_s| \cos\left(\frac{\pi}{(N_k+1)}\right)}$$

$$\widehat{\beta_s} = \frac{\alpha \xi \chi_s}{|\chi_h| \cos\left(\frac{\pi}{(N_j+1)}\right) + |\chi_v| \cos\left(\frac{\pi}{(N_i+1)}\right) + \alpha |\chi_s| \cos\left(\frac{\pi}{(N_k+1)}\right)}$$

where $\alpha = \frac{N_k(N_j-1)}{N_j(N_k-1)}$, $\xi = 0.5 - \delta$, and δ is a small number included to ensure that the estimates are inside the parameter space. The quantities χ_h, χ_v, and, χ_s are defined by,

$$\chi_h = \sum_{m=1}^{n} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j-1} \sum_{k=1}^{N_k} x^m_{ijk} x^m_{i(j+1)k} \quad (8)$$

$$\chi_v = \sum_{m=1}^{n} \sum_{i=1}^{N_i-1} \sum_{j=1}^{N_j} \sum_{k=1}^{N_k} x^m_{ijk} x^m_{(i+1)jk} \quad (9)$$

$$\chi_s = \sum_{m=1}^{n} \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} \sum_{k=1}^{N_k-1} x^m_{ijk} x^m_{ij(k+1)} \quad (10)$$

where x^m_{ijk} represents the intensity of the pixel at spatial location i, j, spectral band k, and within the mth Markov window.

2.3. Single Hypothesis Detection

In an effort to use a 3-D GMRF model for the target, as well as the clutter, we use a single hypothesis test rather than the binary hypothesis test discussed in our previous work [6]. A single hypothesis test is useful for situations in which one class is well-defined and the others are not. Although various types of natural clutter are present in the imagery, we work with the underlying assumption that within the $N_i \times N_j$ processing block, there is only one type of clutter present, and, therefore, the clutter is well-defined by using the AML method to estimate the parameters of the GMRF model.

The processing window is separated into two regions: a perimeter *clutter* region and an interior *target* region. Both the clutter mean and Markov parameters, which completely define the clutter statistics, are estimated from data in the *clutter* region. From the data in the target region, a set of observation vectors, $\{Y_q\}_{q=1}^m$, is formed, and the statistical distance between these vectors and the clutter vectors is measured. Mathematically, the distance measure for the set of m independent observation vectors is defined by:

$$\frac{1}{m} \sum_{q=1}^{m} d_q^2 = \frac{1}{m} \sum_{q=1}^{m} \left(\underline{Y}_q - \underline{\mathrm{m}}_x\right)' \Sigma x^{-1} \left(\underline{Y}_q - \underline{\mathrm{m}}_x\right) \quad (11)$$

where $\underline{\mathrm{m}}_x$ is the clutter mean.

Substituting the spatially-spectrally correlated Kronecker parameterization from equation (7) for Σ_x^{-1} into equation (11),

and simplifying, the distance between the observation vectors from the *target* region and those of the *clutter* region becomes,

$$\frac{1}{m}\sum_{q=1}^{m} d_q^2 = \frac{1}{m\widehat{\sigma^2}}\sum_{q=1}^{m}\left(S_y - \widehat{\beta_h}\mathcal{Y}_h - \widehat{\beta_v}\mathcal{Y}_v - \widehat{\beta_s}\mathcal{Y}_s\right) \quad (12)$$

where $\widehat{\beta_h}, \widehat{\beta_v}$, and $\widehat{\beta_s}$ are defined by the equations in section 2.2 and $\widehat{\sigma^2}$ is a linear combination of $\widehat{\beta_h}, \widehat{\beta_v}$, and $\widehat{\beta_s}$. The quantities $\mathcal{Y}_h, \mathcal{Y}_v$, and $\mathcal{Y}_s$,are defined by equations (8),(9), and (10), with the exception that only data in the central *target* region of the processing window is used in their computation. S_y is the autocorrelation of the observation data.

Since we simply compare the *target* vectors to the *clutter* vectors, it is implied in the single hypothesis approach that the target, in addition to the clutter, is modeled by a 3-D noncausal GMRF. The underlying assumption is that the target and clutter are correlated differently and that these differences will be reflected in the estimates for β_h, β_v, and β_s.

3. PERFORMANCE RESULTS

In this section, we compare the computational and detection performance of our fully adaptive GMRF anomaly detector to the RX algorithm, which is the benchmark anomaly detection algorithm for multispectral imagery. The RX algorithm is also one of the most prevalent anomaly detection algorithms currently being used to analyze hyperspectral imagery, even though to make the analysis feasible, the number of bands from the hyperspectral data set must be significantly reduced.

Since we are interested in the ability of the detector to correctly tag man-made anomalous targets in real-time situations, our first measure of comparison is the computational complexity of the algorithms measured in floating point operations (FLOPs) per pixel. As the second comparison measure, we use detection ROC curves which plot the false alarm rate versus the probability of detection. Throughout this section, we use the acronym SHYP-AML to represent our GMRF detector which uses single hypothesis detection in conjunction with AML estimation.

3.1. Computational Performance

To evaluate the computational effectiveness of the algorithms, we derive from the C-code implementation of each algorithm, an expression relating the number of FLOPS required to compute the detection statistic for one pixel in the image set to the number of spectral bands used for processing. The FLOPs for the GMRF and RX algorithms are also dependent on the sizes of the processing, *target*, and Markov windows. A plot of the number of FLOPS as a function of the number of spectral bands used for processing is shown in figure 1(a). The processing, target, and Markov windows are fixed at appropriate sizes for 1m GSD data. Most noticeably, the SHYP-AML algorithm has the advantage that the total number of FLOPs increases linearly with the number of spectral bands N_b. As shown in figure 1(a), this is a significant improvement over other ML-detection implementations, such as RX, in which the computational complexity increases with N_b^3. This high order polynomial

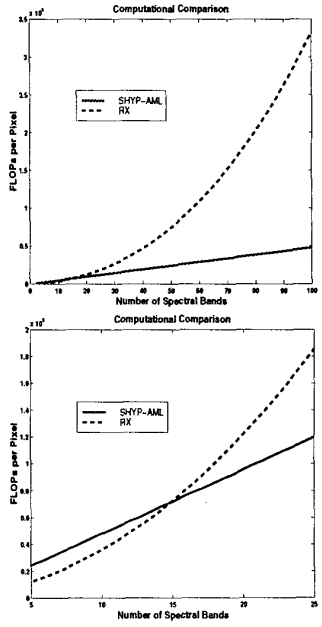

Figure 1: (a) Computational comparison: SHYP-AML vs RX. The GMRF algorithm is linearly rather than exponentially dependent on the number of spectral bands. (b) A magnified view of the cross-over region in (a)

growth in FLOPS as the number of spectral bands grows precludes the use of algorithms, such as RX, on true hyperspectral data. In practice, the application of RX is usually preceded by binning the hyperspectral cubes, i.e., averaging spectral bands. Also, RX decouples the spatial and spectral correlation, and assumes the data is spatially white. In contrast, the linearly increasing complexity of the GMRF algorithm makes it viable and practical, from a computational point of view, even when using a large number of spectral bands and fully exploiting the coupled spatial and spectral correlation of the data.

Figure 1(b) is a zoomed in view of the region in figure 1(a) in which the curves of the SHYP-AML, and RX algorithms cross one another. Due to the overhead involved in estimating the GMRF parameters, the plots indicate that the RX algorithm provides a slight computational advantage when a small number of spectral bands is used for processing, i.e., when using multispectral imagery. However, when using more than approximately 15 spectral bands, the SHYP-AML approach is computationally superior. The GMRF algorithm can potentially perform even better, computationally, than indicated by figure 1(b) because the C-code implementation has not been fully optimized. The code for the RX algorithm, on the other hand, is a mature implementation that is distributed for use in industry.

3.2. Detection Performance

In the last section, we showed that our new fully-adaptive GMRF anomaly detection algorithm provides a significant

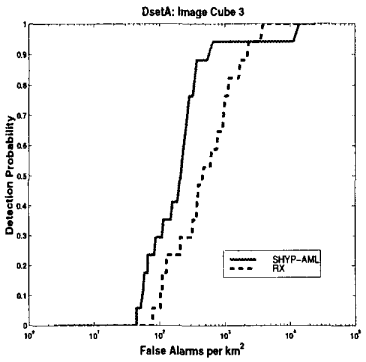

Figure 2: ROC performance of the GMRF and RX algorithms on HYDICE sensor data

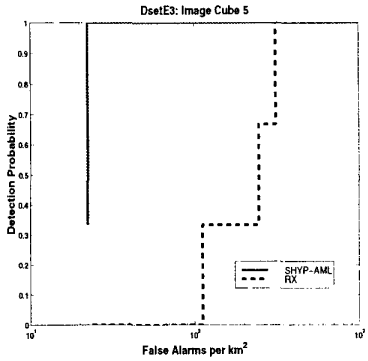

Figure 3: ROC performance of the GMRF and RX algorithms on SEBASS sensor data

computational improvement over the benchmark anomaly detection algorithm. In this section, we show that this significant computational gain is coupled with an improvement in performance, in comparison to the RX algorithm, for targets with significant spatial extent.

The first set of results are shown in figure 2, and are for a HYDICE hyperspectral sensor image. The Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor collects 210 bands of data in the visible to short-wave infrared portion of the electromagnetic spectrum. The ROC curves in figure 2 are a result of averaging the total 210 bands down to 41 bands. This is necessary in order to make comparisons to the RX algorithm which is limited computationally in the number of spectral bands on which it can reasonably operate. The scene, in this case, contains both forest and open dirt field regions. There are a total of 20 targets that are located in the open.

In this example, figure 2 shows that the SHYP-AML algorithm outperforms the RX algorithm. It is significant to note that the GMRF algorithm displays better performance while processing the data in a third of the time of the RX algorithm.

Figure 3 shows the results of the GMRF and RX algorithms on a SEBASS sensor hyperspectral image. The Spatially Enhanced Broadband Array Spectrograph System (SEBASS) sensor collects 128 bands of data in the long-wave region of the electromagnetic spectrum. For this image, the scene is a mixture of forest, fields, and water, but is an easier data set in the sense that there are only 5 targets in the scene, all of which are in the open and surrounded by extremely homogeneous backgrounds. For this example, the 128 bands have been binned into 31 bands.

The results show that the GMRF algorithm significantly outperforms the benchmark RX method. Again, the GMRF algorithm provides this increased performance in a third of the time of the benchmark algorithm.

4. CONCLUSIONS

In this paper, we have presented an overview of the GMRF ML-detector that we have developed for locating spatial-spectral anomalies in hyperspectral sensor imagery. We have shown that our new detector provides a significant computational advantage over the benchmark anomaly detector, the RX algorithm. Specifically, the computational complexity of the GMRF detector increases linearly with the number of spectral bands in contrast to the exponential increase displayed by the RX method. In addition to this computational improvement, we have shown, on both HYDICE and SEBASS real hyperspectral imagery, that the GMRF detector provides an improvement in detection performance. The coupling of reduced processing time with increased detection performance makes the GMRF detector a promising alternative to the current benchmark anomaly detection algorithm for hyperspectral imagery.

5. REFERENCES

[1] Xialo Yu, Lawrence E. Hoff, Irving S. Reed, An Mei Chen, and Larry B. Stotts, "Automatic target detection and recognition in multiband imagery: A unified ML detection and estimation approach," *IEEE Transactions on Image Processing*, vol. 6, no. 1, pp. 143–156, January 1997.

[2] J.W. Woods, "Two-dimensional discrete Markovian fields," *IEEE Transactions on Information Theory*, vol. 18, no. 2, pp. 232–240, 1972.

[3] José M.F. Moura and Nikhil Balram, "Recursive structure of noncausal Gauss-Markov random fields," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 334–354, March 1992.

[4] Anil K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, 1989, Chapter 2:Two Dimensional Systems and Mathematical Preliminaries.

[5] S.M. Schweizer and José M.F. Moura, "GMRF modeling for detection in hyperspectral imagery: Parameter estimation," July 1999, 32 page manuscript submitted for publication.

[6] Susan M. Schweizer and José M.F. Moura, "Modeling and detection in hyperspectral imagery," *ICASSP*, May 1998.