MODELING AND DETECTION IN HYPERSPECTRAL IMAGERY

Susan M. Schweizer and José M.F. Moura

Carnegie Mellon University, ECE Department
Pittsburgh, PA 15213
switz@ece.cmu.edu

ABSTRACT

One aim of using hyperspectral imaging sensors is in discriminating man-made objects from dominant clutter environments. Sensors like Aviris or Hydice simultaneously collect hundreds of contiguous and narrowly spaced spectral band images for the same scene. The challenge lies in processing the corresponding large volume of data that is collected by the sensors. Usual implementations of the Maximum-Likelihood (ML) detector are precluded because they require the inversion of large data covariance matrices. We apply a Gauss-Markov random field (GMRF) model to derive a computationally efficient ML-detector implementation that avoids inversion of the covariance matrix. The paper details the structure of the GMRF model, presents an estimation algorithm to fit the GMRF to the hyperspectral sensor data, and finally, develops the structure of the ML-detector.

1. INTRODUCTION

This paper addresses the problem of detection of man-made objects using image data obtained from hyperspectral sensors. Sensors like Aviris and Hydice simultaneously collect hundreds of contiguous and narrowly spaced spectral bands. The spectral bands typically extend from the visible to the near-infrared portion of the electromagnetic spectrum. The advantage of having such high spectral resolution is that a dense spectral signature is obtained for each pixel in the image set. This spectral signature varies depending on what is present in the pixel, thus providing a means for discriminating between objects of varying composition.

The challenge of using hyperspectral data is in developing algorithms that can efficiently process the massive amounts of data that are collected by the sensors. For instance, with data from the Naval Research Laboratory's Hydice sensor, 210 images, each 300 × 300 pixels in size, are collected. Typically each pixel is represented by 12 bits resulting in over 200 Mbytes of data for just one scene. Data sets of such magnitude make processing and storage difficult, and real-time transmission practically impossible. Therefore, much of the work with hyperspectral imagery focuses on either compression of the data for purposes of storage and transmission or on developing accurate and efficient models of the hyperspectral data in order to decrease the computational complexity of the algorithms used for processing.

This paper develops an efficient model for the hyperspectral data that uses Gauss-Markov random fields (GMRF). We use this model to design the Maximum-Likelihood (ML) detector. Our approach to ML-detection avoids the inversion of the large covariance matrix which is a major obstacle to applying other detection approaches [2, 7] to hyperspectral data. To capture the spatial and unique spectral information provided by hyperspectral sensors, we model the data, after removal of the spatially varying mean, as a 3-D, noncausal Gauss-Markov Random Field (GMRF).

We begin in section 2 by summarizing the relevant prior work that has been done. In section 3 we detail our GMRF model for hyperspectral data. In section 4 we mathematically formulate the detection problem and detail the estimation method by which the detector is able to locally adapt to the varying background statistics of the data. Section 5 outlines the overall detection algorithm and section 6 presents our preliminary research results regarding the statistical nature of the clutter background of hyperspectral images.

2. PRIOR WORK

There is limited work published concerning the processing of hyperspectral images. The work available in the literature mostly focuses on compression. The detection work has generally been intended for geological classification of ground surfaces and not for the detection of man-made objects.

Extensions of the detectors developed for other sensors, *e.g.* the detectors in [2, 7] for multispectral images, are computationally hampered by the required inversion of the data covariance matrix. For hyperspectral images, this covariance matrix is exceedingly large. Also, these previous works usually make additional assumptions, like spatially uncorrelated clutter or negligible spatial-spectral cross-correlations, which make them sub-optimal.

We show in the following sections how our GMRF model not only captures the fully spatially and spectrally correlated nature of the hyperspectral data, but imposes structure on the second order statistics of the clutter leading to significant reductions in the computational complexity of our detector. Specifically, the use of the GMRF model provides us with a known structure for the inverse of the clutter's spatial-spectral covariance matrix. By dealing directly with the inverse, no matrix inversion is necessary.

3. GMRF MODEL

Unlike prior work, see [2, 7], we use a 3-D GMRF model to describe pixel radiance. The 3-D model assumes both spatial and spectral correlation among the image pixels. Each pixel location in the hyperspectral image set is referenced by the variable x and three sub-indices i,j, and k, which indicate the spatial location and the particular spectral band in which the pixel lies, see figure 1. Due to the varying nature of the statistics of the data, processing is done on windows of the image set of size N_i x $N_j = N$ pixel locations. For purposes of simplifying the presentation of our proposed

0-7803-4428-6/98 $10.00

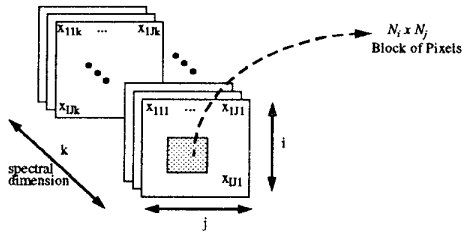

Figure 1: 3-Dimensional Indexing of the Hyperspectral Image Cube

method, we proceed by using a first-order GMRF model. Under the GMRF assumption, each pixel radiance is related to neighboring pixel intensities by:

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

where β_h, β_v, and β_s are the Minimum Mean Square Error (MMSE) predictor coefficients for the horizontal, vertical, and spectral dimensions, respectively, and ϵ_{ijk} is the prediction error [6].

We form a data vector $\underline{X}$ by lexicographically ordering the same N pixel block of data in N_k consecutive spectral bands. This provides us with N_k vectors, $\underline{v}(k)$ $k = 1, \ldots, N_k$, each of length N. By stacking the N_k vectors, we create the data vector

$$\underline{X} = \left[\underline{v}(1)^T \quad \cdots \quad \underline{v}(N_k)^T\right]^T$$

which is size $N_k N \times 1$. An error vector, $\underline{\epsilon}$, is formed in the same manner by beginning with the matrix of prediction errors, ϵ_{ijk}. The error vector is a sample from a zero-mean Gaussian colored noise process which has covariance σ^2A [1, 6]. The resulting data and prediction error vectors are related by $A\underline{X} = \underline{\epsilon}$ where

$$A = \begin{bmatrix} A_1 & A_2 & & \underline{0} \\ A_2 & \ddots & \ddots & \\ & \ddots & \ddots & A_2 \\ \underline{0} & & A_2 & A_1 \end{bmatrix} \quad (2)$$

Using the Kronecker Product as defined in [3], we can compactly represent this matrix by $A = I_{N_k} \otimes A_1 + H^1_{N_k} \otimes A_2$ where $\otimes$ symbolizes the Kronecker Product and the matrices A_1, A_2, B, C, and D are

$$\begin{aligned} A_1 &= I_{N_i} \otimes B + H^1_{N_i} \otimes C \\ A_2 &= I_{N_i} \otimes D \\ B &= -\beta_h H^1_{N_j} + I_{N_j} \\ C &= -\beta_v I_{N_j} \\ D &= -\beta_s I_{N_j} \end{aligned}$$

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 lower diagonals which are composed of all 1's. The resulting matrix A, referred to as the potential matrix, is a sparse matrix and contains all the relevant information regarding the GMRF structure [1].

A main advantage of this model is that we can now define a Kronecker representation for the inverse of the clutter covariance matrix, M^{-1}. GMRF theory tells us that M^{-1} is equal to $\frac{1}{\sigma^2}A$, or, in Kronecker notation:

$$\begin{aligned} M^{-1} &= \tfrac{1}{\sigma^2}\left[-\beta_h\left(I_{N_k} \otimes I_{N_i} \otimes H^1_{N_j}\right) + \left(I_{N_k} \otimes I_{N_i} \otimes I_{N_j}\right)\right] - \\ &\quad \tfrac{1}{\sigma^2}\left[\beta_v\left(I_{N_k} \otimes H^1_{N_i} \otimes I_{N_j}\right) + \beta_s\left(H^1_{N_k} \otimes I_{N_i} \otimes I_{N_j}\right)\right] \end{aligned} \quad (3)$$

M^{-1} is a structured matrix which has unknown parameters σ^2, β_h, β_v, and β_s. Estimating these parameters over different windows of the data allows us to locally adapt to the background clutter statistics.

4. PARAMETER ESTIMATION

4.1. Problem Formulation

Using the GMRF model outlined in the previous section, we now develop a scheme for detection which reveals the advantages of using our model over the models developed in previous work. The goal of the detection stage is to find those pixels which represent possible target locations. We formulate this as a binary hypothesis testing problem where at each pixel the risk is to decide if only clutter is present or clutter plus signal. Formally, the hypothesis testing problem is given by

$$\begin{aligned} H_0 &: \underline{X} = \underline{w} \\ H_1 &: \underline{X} = \underline{w} + \underline{r} \end{aligned}$$

where $\underline{w}$ is the clutter which is a zero-mean GMRF process with unknown second order statistics and $\underline{r}$ is the target signal which is considered to be deterministic and known. The pdfs for the data vector under the two hypotheses are:

$$H_0 : p_{(\underline{x}|\theta,H_0)}(\underline{X} \mid \theta, H_0) = \frac{1}{\sqrt{2\pi^{N N_k}|M|}} \exp(-\tfrac{1}{2}tr(M^{-1}\widehat{M}_0))$$

$$H_1 : p_{(\underline{x}|\theta,H_1)}(\underline{X} \mid \theta, H_1) = \frac{1}{\sqrt{2\pi^{N N_k}|M|}} \exp(-\tfrac{1}{2}tr(M^{-1}\widehat{M}_1))$$

where $\widehat{M}_0 = \underline{X}\underline{X}^T$ and $\widehat{M}_1 = (\underline{X} - \underline{r})(\underline{X} - \underline{r})^T$ are the sample covariance matrices and θ is the vector of the unknown Markov parameters. Examining the pdf's, it is clear that we need an estimate of M^{-1} in order to proceed with the processing. Unlike traditional approaches in which an estimate of M is made and then inverted, we directly estimate the inverse by using equation (3) which results from GMRF theory. The estimate of M^{-1} obtained by using equation (3), is completely determined by estimating the 4 scalar Markov parameters: β_h, β_v, β_s, and σ^2. This is a significant simplification over previous work.

4.2. Optimal Estimation

The optimal detector for a hypothesis testing problem with unknown parameters is the GLRT defined in equation (4).

$$\Delta(\underline{X}) = \frac{p(\underline{X} \mid (H_1, \widehat{\theta}_{ML(1)}))}{p(\underline{X} \mid (H_0, \widehat{\theta}_{ML(0)}))} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \quad (4)$$

The parameters $\widehat{\theta}_{ML(1)}$ and $\widehat{\theta}_{ML(0)}$ are vectors of the Maximum Likelihood (ML) estimates of the unknown Markov parameters. The threshold,η, is determined by the Neyman-Pearson Criterion.

The structured and constrained nature of M^{-1} makes the ML estimates computationally expensive to calculate due to a need to perform a nonlinear optimization. Thus, we present a computationally less expensive LS parameter estimation approach which is intended to approximate the ML estimates. The LS Estimation

algorithm is based on using the constraints: $M^{-1}\widehat{M}_0 = I$ and $M^{-1}\widehat{M}_1 = I$ where $\widehat{M}_0$ and $\widehat{M}_1$ are the sample covariance matrices defined in section 4.1.

4.3. The Least Squares Algorithm

In the first-order case, only four unknown Markov parameters need to be estimated. However, the constraints defined in section (4.2) provide $(N_i N_j N_k)^2$ equations for those four unknowns. This is an overconstrained system, so the unknown parameters are estimated such that the constraints are met in a LS sense. For the sake of brevity, the following estimation procedure will be developed only under hypothesis H_0. The derivation under H_1 is similar and does not warrant a discussion of its own.

Using the Kronecker expression of equation (3) for M^{-1} and rearranging terms we obtain the following equation in terms of the unknown Markov parameters

$$\beta_h\left[\left(I_{N_k}\otimes I_{N_i}\otimes H^1_{N_j}\right)\widehat{M}_0\right]+\beta_v\left[\left(I_{N_k}\otimes H^1_{N_i}\otimes I_{N_j}\right)\widehat{M}_0\right]+ \beta_s\left[\left(H^1_{N_k}\otimes I_{N_i}\otimes I_{N_j}\right)\widehat{M}_0\right]+\sigma^2\left(I_{N_k}\otimes I_{N_i}\otimes I_{N_j}\right)=\widehat{M}_0 \quad (5)$$

By first lexicographically ordering the matrices in equation (5), we can write the equation in matrix notation as $G\underline{\theta} = \underline{\delta}$ where $\underline{\theta}$ is a vector of the unknown parameters that need to be estimated. The matrix G, and vector, $\underline{\delta}$, are:

$$G = [G_1 \quad G_2]$$
$$\underline{\delta} = vec\left(\widehat{M}_0\right)$$

where

$$G_1 = \left[vec\left[\left(I_{N_k}\otimes I_{N_i}\otimes H^1_{N_j}\right)\widehat{M}_0\right] \quad vec\left[\left(I_{N_k}\otimes H^1_{N_i}\otimes I_{N_j}\right)\widehat{M}_0\right]\right]$$
$$G_2 = \left[vec\left[\left(H^1_{N_k}\otimes I_{N_i}\otimes I_{N_j}\right)\widehat{M}_0\right] \quad vec\left(I_{N_k}\otimes I_{N_i}\otimes I_{N_j}\right)\right]$$

The estimates for the Markov parameters are obtained from the pseudo-inverse or LS solution which is $\hat{\underline{\theta}}_{LS} = \left(G^T G\right)^{-1} G^T \underline{\delta}$.

The most computationally demanding operation in the equation is the inverse of $G^T G$. However, for a first order GMRF, this matrix is of dimension 4 x 4 regardless of how many bands and pixel locations are chosen for processing making the computation of the inverse relatively insignificant. For an arbitrary order GMRF the dimensions of $G^T G$ will be equivalent to the number of unknown Markov parameters which, in general, will be small.

5. THE DETECTION ALGORITHM

5.1. The GLRT Detector

Assuming that the LS estimates are "good", meaning that the constraints in section (4.2) are closely met,the GLRT defined in equation (4) reduces to

$$\Delta(\underline{X}) = \frac{|\hat{M}^{-1}_{(1)}|}{|\hat{M}^{-1}_{(0)}|} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \quad (6)$$

$\widehat{M}^{-1}_{(0)}$ is the spatial-spectral covariance matrix which has had the parameter estimates assuming hypothesis H_0 was true substituted in, $\widehat{M}^{-1}_{(1)}$ is the covariance matrix with the estimates obtained assuming hypothesis H_1 is true, and $|\cdot|$ is the determinant operator.

5.2. Determinant Expressions

To make the detection decision it is necessary to find expressions for the determinant of the spatial-spectral covariance matrix under each hypothesis. To do this, we use the property that the determinant of a matrix is equal to the product of all its eigenvalues. Therefore, we need to find an expression for the eigenvalues of M^{-1}.

Beginning with equation (3), and assuming that $N_i = N_j = N_k = n$, we use properties of linear algebra [5] and the Kronecker product [3] to determine that the expression for the eigenvalues is [4]:

$$\lambda_l\left(\hat{M}^{-1}_{(0)}\right) = \frac{1}{\hat{\sigma}^2}\left[1-\left(\hat{\beta}_h+\hat{\beta}_v+\hat{\beta}_s\right)\left(\lambda_i\left(I_n\right)\lambda_j\left(I_n\right)\lambda_k\left(H^1_n\right)\right)\right] \quad (7)$$
$$\text{where } 1\le i\le n, \quad 1\le j\le n, \quad 1\le k\le n, \quad 1\le l\le n^3$$

The eigenvalues of the matrix H^1_n as defined in [1] are $\lambda_k\left(H^1_n\right) = 2\cos\left(\frac{k\pi}{n+1}\right)$ for $1\le k\le n$. Using this expression in equation (7), we determine that there are only n distinct eigenvalues with each of these n eigenvalues repeated n^2 times. The n distinct eigenvalues are the n eigenvalues of the matrix H^1_n. The overall expression for the determinant under assumption H_0 is:

$$|\hat{M}^{-1}_{(0)}| = \prod_l \lambda_l\left(\hat{M}^{-1}_{(0)}\right)$$
$$= \prod_k\left[\frac{1}{\hat{\sigma}^2}\left(1-2\left(\hat{\beta}_h+\hat{\beta}_v+\hat{\beta}_s\right)\cos\left(\frac{k\pi}{n+1}\right)\right)\right]^{n^2} \quad (8)$$

If we replace the parameter estimates in Equation (8) with the estimates obtained assuming hypothesis H_1 is true, we obtain the determinant of $\hat{M}^{-1}_{(1)}$. Computationally, equation (8) requires the multiplication of only k scalar values.

6. DATA RESULTS

The previous sections have explained the theoretical foundation of our proposed GMRF Detection method. The success and significance of this proposed method relies on the accuracy of several assumptions, mainly with regard to the clutter statistics that are being modeled in the hyperspectral imagery. In this section, we use real hyperspectral data obtained from Purdue University to evaluate these underlying assumptions about the statistical nature of the background clutter.

6.1. Spatial-Spectral Correlation

In section 3, we present a fully spatially-spectrally correlated GMRF model. In prior work, *e.g.* [7], it is assumed that the data is spatially independent leading to simplifications in both the theory and complexity of the proposed methods. We assert that this assumption is a rough approximation, and , in general, does not accurately represent the hyperspectral data. To validate our assertion, we begin by empirically analyzing sample spatial-spectral correlation matrices from real data.

Figure 2 shows spatial-spectral correlation matrices that were estimated using 8 x 8 blocks by 8 bands of data for four different sets of consecutive bands. Pixels which appear closer to white indicate higher correlations. We conclude from these images that the spacing of the diagonal lines in the images in figure 2 indicate that

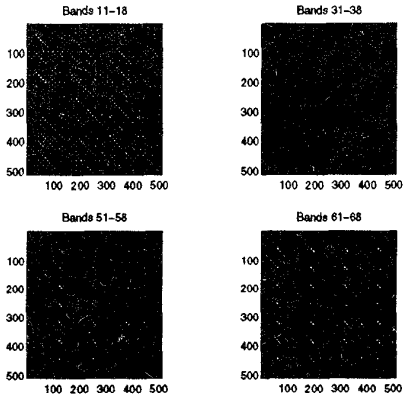

Figure 2: Spatial-Spectral Correlation Matrices for four sets of 8 consecutive spectral bands

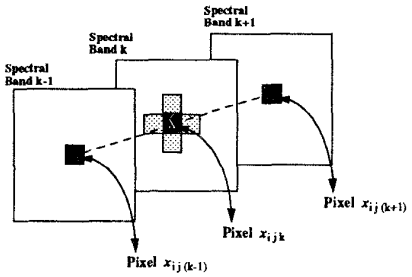

Figure 3: Correlated Neighbors for 1st-order GMRF model. Black squares indicate high correlation while light dotted squares indicate weak correlation.

a pixel is more highly correlated with pixels that are in the corresponding location in other spectral bands than with pixels in the same spectral band that are in neighboring locations, see figure 3. However, this does not indicate that a pixel is spatially uncorrelated. Before drawing any conclusions about spatial independence, we first evaluate estimates of the 3-D Markov parameters.

6.2. Markov Parameter Estimates

If we apply the first-order GMRF LS parameter estimation technique developed in section 4.3, which makes no assumption about spatial independence, to the subsets of data used in figure 2, we obtain estimates for β_h, β_v, and β_s. The resulting estimates are listed in the table in Figure 4. In each case, the β_s estimates are larger than the estimates for β_h and β_v. However, the spatial estimates are significantly large that they should not be neglected. These parameter estimates support the empirical results in that they indicate high degrees of spectral correlation, but, in addition, they suggest non-negligible amounts of spatial correlation. Therefore, the spatially white assumption made in prior work is only a rough approximation to the true nature of the data. A more general and accurate

	β_h	β_v	β_s
Bands 11-18	.1569	.0226	.4020
Bands 31-38	.0425	.0235	.5386
Bands 51-58	.0957	.0235	.4747
Bands 61-68	.1011	.0226	.4715

Figure 4: Table of Estimated Markov Parameters

representation of the data is our fully spatially-spectrally correlated model. Our 3-D Markov model holds regardless of whether the data is fully spatially-spectrally correlated or spatially independent since in the latter case the estimates for β_h and β_v would be zero.

7. CONCLUSION

We have presented a new 3-D noncausal GMRF model for hyperspectral sensor data. We use the model to develop a detection algorithm that is capable of efficiently processing the massive amounts of hyperspectral data that are recorded by the sensors. A main assumption of the model is that the data is both spatially and spectrally correlated. By using our GMRF model, we obtain a highly structured and well-defined expression for the inverse spatial-spectral covariance matrix which contains only 4 unknown parameters. Using this expression, we avoid doing any matrix inverse operations, thus, significantly reducing the computational complexity of the overall algorithm. Also, by estimating the parameters for different windows of the data, the algorithm is able to locally adapt to the background statistics.

8. REFERENCES

[1] N. Balram and J. M. Moura. Noncausal Gauss-Markov random fields and parameter structure and estimation. *IEEE Transactions on Information Theory*, 39(4):1333–1355, July 1993.

[2] C. Ferrara. Adaptive spatial/spectral detection of subpixel targets with unknown spectral characteristics. *SPIE: Signal and Data Processing of Small Targets*, 2235:82–93, 1994.

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

[4] S. Schweizer. Detection in hyperspectral imagery. Technical report, Carnegie Mellon University, September 1997. Doctoral Prospectus.

[5] G. Strang. *Linear Algebra and Its Applications*. Harcourt, Brace, Jovanovich, Publishers, 1988. 3rd Edition.

[6] J. Woods. Two-dimensional discrete Markovian fields. *IEEE Transactions on Information Theory*, 18:232–240, 1972.

[7] X. Yu, L.Hoff, I. S. Reed, A. M. Chen, and L. Stotts. Automatic target detection and recognition in multiband imagery: A unified ML detection and estimation approach. *IEEE Transactions on Image Processing*, 6(1):143–156, January 1997.