Fast hyperspectral data processing methods

Thomas Boggs and Richard B. Gomez
School of Computational Sciences
George Mason University
4400 University Drive, Fairfax, VA 22030 USA

ABSTRACT

The need for fast hyperspectral data processing methods is discussed. Discussion includes the necessity of faster processing techniques in order to realize emerging markets for hyperspectral data. Several standard hyperspectral image processing methods are presented, including maximum likelihood classification, principal components analysis, and canonical analysis. Modifications of those methods are presented that are computationally more efficient than standard techniques. Recent technological developments enabling hardware acceleration of hyperspectral data processing methods are also presented as well as their applicability to various hyperspectral data processing algorithms.

Keywords: Hyperspectral, data processing, image processing, principal components, HSI

1. INTRODUCTION

The development of hyperspectral imaging (HSI) spectrometers has made it possible to extract information not previously available from remotely sensed data. These types of imaging spectrometers collect not only high resolution spatial imagery but also spectral data at a resolution high enough to classify materials in remotely sensed data. For example, the airborne visible/infrared imaging spectrometer (AVIRIS) collects 224 spectral channels from the visible through thermal infrared (IR) regions of the electromagnetic spectrum.

The price paid for the wealth information available from these hyperspectral sensors is the enormous amounts of data that they generate. While standard methods exist for extracting desired information from such huge volumes of data [1], these techniques are often computationally tedious and require lengthy durations to calculate desired quantities. For some applications, the durations required for such lengthy computations are prohibitively long.

Many applications exist for hyperspectral data in which it is desirable to have the desired information calculated from the data in real-time or near real-time. HSI data has been proposed for fusion with synthetic aperture radar (SAR) data for the detection of camouflage, concealment, and decoys in military applications [2]. In such an application where the goal is detection of hostile weaponry, fast processing can be of paramount importance. HSI data has application in the detection and tracking of natural disasters such as forest fires oil spills, and other types of chemical contamination, where timely classification is highly desirable.

Efficient processing of data is the key to HSI realizing its full potential in numerous areas. HSI data has already found utility in the field of agriculture. When the operating costs of collecting HSI data become sufficiently low and access to data becomes widespread (e.g., via the internet), data may be available to farmers and other agriculture professionals as a subscriber service. The availability of such data for precision agriculture would provide valuable data on crop maturity, moisture levels, nutrient levels, etc.

One of the keys to reducing operating costs is to implement real-time or near real-time processing methods. By simply storing data for post-processing, airborne HSI measurement systems are limited by their storage capacity. If the desired information is extracted in-flight, airborne imagers would then be able to fly longer, reducing the overall cost of data collection.

In the following sections, methods are presented which reduce computation time relative to standard methods. The methods discussed include Maximum Likelihood Classification, Principal Components Analysis, Canonical Analysis, and multi-stage methods.

2. MAXIMUM LIKELIHOOD CLASSIFICATION (MLC)

One of the widely used HSI data classification techniques is Gaussian Maximum-Likelihood Classification. The method assumes that each information class, ω_i, is described by l spectral components which are independent gaussian random variables. Under this method, a hyperspectral image pixel is classified as belonging to information class k if

$$d_k(X) = \max\{d_i(X)\}, i = 1,2,...c ,$$

where c is the number of classes and $d_i(X)$, known as the discriminant function, is based on the assumption that the classes have multivariate normal distributions. It is defined by

$$d_i(X) = -\ln|\Sigma_i| - (X - M_i)^T \Sigma_i^{-1}(X - M_i), \qquad \text{Eq. 1}$$

M_i is estimated mean value of class ω_i and Σ_i, is its covariance matrix, estimated by

$$\Sigma_i = \frac{1}{N_i - 1} \sum_{j=1}^{N} (X_j - M_i)(X_j - M_i)^T , \qquad \text{Eq. 2}$$

where N_i is the number of samples for class ω_i.

While the assumption of multivariate normal distributions greatly simplifies the formulation of the classification problem, maximum likelihood classification is nevertheless a computationally intensive process. Before the process of pixel classification begins, estimates of the means and covariance matrices of the information classes can be pre-calculated; however, the second term of Eq. 2 (referred to as the quadratic term, $Q_i(X)$) requires on the order of l^2 multiplications. Thus, for classification among c classes, classification of a single pixel requires on the order of cl^2 multiplications.

Several schemes have been developed to reduce the computational burden of the maximum likelihood classifier. Venkateswarlu and Raju [3] developed a three-stage implementation of MLC, which speeds up the computation of $Q_i(X)$ by up to a factor of two by decomposing the class covariance matrix with a lower triangular matrix. Use of the quadratic form range theorem is then applied to estimate the range of the quadratic term for each class. These ranges are then applied to eliminate classes from consideration for each pixel. The total time of classification using this multi-stage method for the test case used was approximately 1/3 of the time for conventional MLC.

Lee and Landgrebe [4] developed a multi-stage MLC method, which utilizes a reduced feature set for early disqualification of unlikely classes. The first stage of the method calculates the quadratic term for an observation vector using only a small subset of the total number of spectral bands. The discriminant function is then calculated for each class and all classes whose discriminant values fall below a specified threshold are eliminated. After unlikely classes are eliminated, the process is repeated in the next stage using an increased number of features for the reduced set of classes. Efficiency of the method relies on the ability of the equation below to calculate the quadratic term for a given stage based on the values of the reduced feature set from the previous stage:

$$(X_{n+1} - M_{n+1})^T \Sigma_{n+1}^{-1}(X_{n+1} - M_{n+1}) = X_n^T \Sigma_n^{-1} X_n + \alpha \left[(u^T \Sigma_n^{-1} X_n)(u^T \Sigma_n^{-1} X_n) - 2x_{n+1} \right] + x_{n+1}^2 \right]$$

where

$$\Sigma_{n+1} = \begin{bmatrix} \Sigma_n & u \\ u^T & a \end{bmatrix} \quad \text{and} \quad \alpha = \frac{1}{a - u^T \Sigma^{-1} u} .$$

The number of features used increases at each successive stage while the number of potential classes decreases due to the threshold set for the discriminant function. In the final stage, the entire feature set is used for classification from among the remaining classes. The method reduces computation time by factor of 3 to 7, with accuracy comparable to traditional Gaussian MLC.

Hyperspectral data often contains large amounts of redundant information. This is evidenced by the high correlations that often exist between various spectral channels. In many cases, high correlations appear among contiguous bands of spectral channels, with very little correlation appearing between these various sub-bands. In such cases, the covariance matrix will have a block diagonal form. Jia and Richards [5] developed a MLC method that uses the characteristics of a block diagonal covariance matrix to substantially decrease classification time.

It is easily shown that the determinant of a block diagonal matrix is equal to the product of the determinants of the constitutive blocks. Thus, if class ω_i has a block diagonal covariance matrix consisting of M blocks, the determinant of Σ_i is given by

$$\left|\Sigma_i\right| = \prod_{j=1}^{M} \left|\Sigma_{ij}\right|, \qquad \text{Eq. 3}$$

Since the inverse of a block diagonal matrix consists of the inverses of the blocks, it is also true that, for a given vector y_i,

$$y_i^T \Sigma_i^{-1} y_i = \sum_{j=1}^{M} y_{ij}^T \Sigma_{ij}^{-1} y_{ij}$$

So that the ML discriminant can be rewritten as

$$d_i(X) = -\sum_{j=1}^{M} \left\{ \ln\left|\Sigma_{ij}\right| + \left(X - M_i\right)^T \Sigma_{ij}^{-1}\left(X - M_i\right) \right\}$$

The method described can achieve classification times down to less than 10% of conventional MLC. Furthermore, due to the reduced block structure of the problem, fewer training pixels are required relative to the standard method.

3. PRINCIPAL COMPONENTS TRANSFORMATION

Principal Components Transformation (PCT) is the process of applying a linear transformation to an image's covariance matrix to rotate the spectral coordinates into a coordinate space in which the spectral components are uncorrelated. The transformation sought is of the form

$$y_i = A^T x_i$$

where x_i and y_i are the original and transformed coordinates and A is the transformation matrix. The solution of the matrix A is found by solving the characteristic equation defined by

$$\left|\Sigma_x - \lambda I\right| = 0$$

For l frequency components, the equation above amounts to solving the roots of an order l polynomial in λ. Each solution of the eigenvalue λ has a corresponding eigenvector, which is one of the columns of A. The eigenvector represents one of the coordinate axes in the uncorrelated coordinate space and the corresponding eigenvalue (λ) represents the variance of the image data along that axis. We can rank the eigenvectors by order of decreasing eigenvalue.

The utility of PCT lies in the fact that most of the variance in the image can be represented by the first few eigenvectors. We can then project the original image pixel vectors along those few eigenvectors, reducing the image dimensionality while retaining the vast majority of the image information.

Jia and Richards [6] have proposed a method to exploit the block diagonal nature of hyperspectral data to perform more efficient PCT calculations. Utilizing Eq. 3, instead of solving the roots of an order l polynomial, we instead solve M smaller equations, requiring fewer total calculations. In addition to easing the complexity of the eigenvalue solution, the block diagonal formulation makes the actual pixel transformation substantially less complex since the number of multiplications required for transformation is proportional to the total number of pixels. For 3 equal sized blocks within the covariance matrix, the total number of multiplications is reported to be approximately 2/3 less than with the standard PCT implementation.

4. CANONICAL ANALYSIS (CA)

While the Principal Components Transformation (PCT) may be optimal for image compression, it is not necessarily optimal for image classification and class separability. An alternate eigen-analysis method, aimed at maximizing class separability is known as Canonical Analysis (CA). The goal of CA is to perform a linear transformation which minimizes the variance within each class while maximizing the variance between different class means. According to the method, if we have sufficient training data, we can define the within-class and between-class covariance matrices as

$$\Sigma_w = \frac{1}{c} \sum_{i=1}^{c} \Sigma_i$$

$$\Sigma_b = \frac{1}{c} \sum_{i=1}^{c} \frac{N_i}{N} (m_i - m)(m_i - m)^T$$

where c is the number of classes, N_i is the number of samples in class ω_i, m_i is the mean value of class ω_i, and m is the global mean. To maximize class separability, we would like to minimize Σ_w and simultaneously maximize Σ_b. The formulation of the problem reduces to the characteristic equation

$$(\Sigma_b - \Lambda\Sigma_w)D = 0$$

This eigenvalue equation yields at most c-1 non-zero eigenvalues.

Canonical Analysis has been applied to a two-stage classification method to achieve significant speed increase and dimensionality reduction [7]. The first stage of the process applies CA to find the transformed coordinate axes, which maximize class separation. Loading factors are then calculated to rank spectral components is decreasing order of discriminant power. This allows the analyst to then reduce dimensionality by selecting a number of bands, which achieves a desired percentage of the total discriminant power. CA is then applied a second time to remove spectrally redundant features.

The second stage of the method utilizes a fast recursive MLC process. The classification procedure uses a lower diagonal covariance matrix representation similar to [3]. The classification stage also utilizes Winograd's algorithm to further speed up the vector multiplication required for the quadratic term. The two-stage method was applied to experimental data and found to be 30 to 145 times faster than conventional MLC, while using as few as 7 of 56 spectral bands.

Chang et al recently applied a similar two-stage method for HSI classification [8]. Here three different methods are used for band prioritization: Principal Components Analysis, Noise-Adjusted Principal Components Analysis, Canonical Analysis, and Orthogonal Subspace Projection (OSP). Next, a divergence-based band decorrelation procedure is invoked. Classification was performed on a 210 channel HYDICE image. No more than 12 bands were needed for any of the four band selection/prioritization schemes to achieve classification accuracy comparable to using all 210 bands.

5. FASTER PROCESSING VIA HARDWARE

In addition to more efficient numerical algorithms, computing hardware has a profound effect on HSI processing speed. Recent developments in parallel processing architectures have made it possible to achieve supercomputing power using arrays of commercial off-the-shelf (COTS) computing hardware. Parallel architectures such as the Message Passing Interface (MPI) and Parallel Virtual Machine (PVM) allow one to achieve supercomputing speeds with clusters of personal computers (PCs). The matrix orientation of many HSI processing tasks makes such computations suitable for parallel processing environments.

Digital signal processors (DSP's) are also suitable for HSI computations. Such processors are optimized for performing multiply-accumulate operations which are common in signal processing applications. Many DSP's are also designed to be easily implemented in DSP clusters for parallel processing.

Finally, optical processing may also have application in HSI computations. Lui et al [9] make use of a photorefractive correlator to store eigenimages for parallel implementation of the principal components transformation. While their implementation was for spatial pattern recognition, the potential may exist to apply the same technology to HSI.

6. SUMMARY

Several standard techniques for HSI data compression and classification have been presented along with modifications of those methods for efficient processing. The modifications of those standard methods can achieve classification speed increases of up to several orders of magnitude while maintaining comparable accuracy. While some methods perform faster than others, each fills a different need depending on factors such as the character of the HSI data (e.g., block diagonal data), the need for supervised vs. unsupervised classification, separability of information classes, etc.

Several recent developments in data processing hardware have been discussed. While improvements in hardware computing capabilities do increase the performance of HSI processing algorithms, there is always a benefit to be found in more efficient algorithms. As soon as performance of computing hardware improves, there is an associated increase in the processing demands we make on that hardware. So we will always be that much more capable and efficient by making the best use of the processing power available through the use of efficient hyperspectral data processing algorithms.

REFERENCES

1. J. A. Richards, X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*, 2nd ed. Berlin, Germany: Springer-Verlag, 1993.
2. S. M. Hsu, J. P. Kerekes, H. Burke, S. Crooks, "SAR and HSI Fusion for Counter CC& D", MIT Lincoln Laboratory, 1999.
3. N. B. Venkateswarlu, V. S. K. Raju, "Three Stage ML Classifier", *Pattern Recognition*, vol. 24, no. 11, pp.1113-16, 1991.
4. C. Lee, D. A. Landgrebe, "Fast Likelihood Classification", *IEEE Trans. Geoscience Remote Sensing*, vol. 29, no. 4, pp.509-17, 1991.
5. X. Jia, J. A. Richards, "Efficient Maximum Likelihood Classification for Imaging Spectrometer Data Sets", *IEEE Trans. Geoscience Remote Sensing*, vol. 32, no. 2, pp. 274-81, 1994.
6. X. Jia, J. A. Richards, Segmented Principal Components Transformation for Efficient Hyperspectral Remote-Sensing Image Display and Classification", *IEEE Trans. Geoscience Remote Sensing, vol. 37, no. 1, 1999.*
7. T. Tu, C. Chen, J. Wu, C. Chang, "A Fast Two-Stage Classification Method for High-Dimensional Remote Sensing Data", *IEEE Trans. Geoscience Remote Sensing*, vol. 36, no. 1, pp. 182-91, 1998.
8. C. Chang, Q. Du, T. Sun, L. G. Althouse, "A Joint Band Prioritization and Band-Decorrelation Approach to Band Selection for Hyperspectral Image Classification", *IEEE Trans. Geoscience Remote Sensing*, vol. 37, no.6, pp.2631-41, 1999.
9. H. Lui, M. Wu, G. Jin, Q. He, Y. Yan, "Parallel Implementation of the Karhunen-Loeve Procedure for Image Recognition", SPIE Conf. Parallel Dist. Meth. Imag. Proc. III, pp. 158-65, Denver, CO, 1999.