A NEYMAN-PEARSON APPROACH TO ESTIMATING THE NUMBER OF ENDMEMBERS

Joshua Broadwater and Amit Banerjee

The Johns Hopkins University
Applied Physics Laboratory
11100 Johns Hopkins Rd.
Laurel, MD 20723
Joshua.Broadwater@jhuapl.edu
Amit.Banerjee@jhuapl.edu

1. INTRODUCTION

A significant amount of research has gone towards identifying the correct number of endmembers for a scene. Most algorithms have focused on what has been termed "intrinsic" dimensionality [1]. These dimensionality measures such as the Akaike Information Criterion (AIC) [2], the Minimum Description Length (MDL) [3], and the Empirical Indicator Function (EIF) [4] attempt to identify the true number of endmembers in a scene. The problem with these approaches are that different applications require different number of endmembers to achieve desired performance [1][5]. Furthermore, the number of endmembers depends upon the image being analyzed and the specific algorithm to be used [5].

For object detection, the background must be characterized such that the probability of detecting the material is maximized for a given probability of false alarm. In such cases, the number of endmembers required to characterize the background may be significantly more than the intrinsic dimensionality. Although the reasons are varied, this is often due to need for additional endmembers to account for shadowing effects, sensor artifacts, and finer material identification (e.g. coarse sand vs. fine sand). This has been noted in [1] where the best number of endmembers varied for different applications. This measure of dimensionality relative to detection performance has been termed virtual dimensionality [1] and includes a number of metrics by researchers such as Chang and Du [1], Thai and Healey [6], and Broadwater [5].

In this paper, we present a new method for identifying the number of endmembers for material identification using Neyman-Pearson hypothesis testing. Unlike Chang and Du's approach using Neyman-Pearson, this approach directly estimates the conditional distributions of the background and material to be found. From this information, a material can be identified for a given false alarm density. Results on multiple real-world hyperspectral images from various hyperspectral sensors will be provided showing the usefulness of the algorithm.

2. ALGORITHM DESCRIPTION

The proposed algorithm is based on a standard Neyman-Pearson hypothesis test:

$$\hat{m} = \arg\max_{m} P_m(x \geq \tau \mid H_1) \text{ for a given } \alpha_0 = \int_{\tau}^{\infty} P_m(x \mid H_0)dx \tag{1}$$

where m represents the number of endmembers, x is the output of the identification algorithm, H_1 is the alternate hypothesis (material present), H_0 is the null hypothesis (material absent), τ is the threshold separating the two hypotheses, α_0 is the desired false alarm density, and $P_m(x)$ is the probability density function of an algorithm for m endmembers. The idea is to select the number of endmembers that provide the highest probability of detection for a given false alarm rate. The difficulty with this approach is the estimation of the probability density functions.

To address this problem, we propose using Monte Carlo techniques and the theoretical distribution of the specific detection algorithm. This approach allows the number of endmembers to be chosen specifically for a particular image, desired spectral signature, and detection algorithm. A subset of pixels are drawn from the image and the threshold is found using

$$\alpha_0 = \int_{\tau}^{\infty} p(x \mid H_0)dx \approx \frac{1}{N}\sum_{i=1}^{N} I(X_i \geq \tau \mid H_0) \tag{2}$$

where the X_i represent a set of N samples drawn randomly from the image. To estimate the conditional distribution for the alternate hypothesis, the theoretical distribution of the detection algorithm is used. For example, the matched subsapce detector (MSD) results in a non-central F distribution. These distributions are updated with increasing number of endmembers until the maximum probability of detection is found. The point at which the maximum Pd is found dictates the number of endmembers to use. The key is that any endmember extraction algorithm and material detection algorithm can be used with this method.

3. EXPERIMENTAL RESULTS

The detector used for these experiments is the Adaptive Matched Subspace Detector (AMSD) algorithm [7]. This is a standard subpixel detector in the literature that uses the eigenvectors of the image correlation matrix as the background endmembers. Results are shown on multiple hyperspectral images collected with different sensors. Additionally, one set of images is processed in radiance while another is processed in reflectance. Target materials varied from dark tarps to small bright white panels that are placed in both rural and urban environments. In all cases, the proposed algorithm identified the number of endmembers that provided the best material detection scores.

REFERENCES

[1] C-I Chang and Q. Du, "Estimation of Number of Spectrally Distinct Signal Sources in Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 3, pp. 608-619, March 2004.

[2] H. Akaike, "A new look at the statistical identification model," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716-723, December 1974.

[3] J. Rissanen, "A universal prior for integers and estimation by minimum description length," *Annals of Statistics*, vol. 11, pp. 416-431, 1983.

[4] E.R. Malinowski, "Theory of error in factor analysis," *Analytical Chemistry*, vol. 49, no. 4, pp. 606-612, 1977.

[5] J.B. Broadwater, "Effects of Endmember Dimension on Subpixel Detection Performance," in *Proc. of the 2008 IEEE International Geoscience and Remote Sensing Symposium*, July 2008.

[6] B. Thai and G. Healey, "Invariant Subpixel Material Detection in Hyperspectral Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 3, pp. 599-608, March 2002.

[7] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral Subpixel Target Detection Using the Linear Mixing Model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 7, pp. 1392-1409, July 2001.