MODELING NOISE IN A FRAMEWORK TO OPTIMIZE THE DETECTION OF ANOMALIES IN HYPERSPECTRAL IMAGING

Frank M. Mindrup
Department of Operations Research
Air Force Institute of Technology
WPAFB, OH 45434

Trevor J. Bihl
Department of Operations Research
Air Force Institute of Technology
WPAFB, OH 45434

Dr. Kenneth W. Bauer Jr.
Department of Operations Research
Air Force Institute of Technology
WPAFB, OH 45434

ABSTRACT

Hyperspectral imagery (HSI) has emerged as a valuable tool supporting numerous military and commercial missions. Environmental and other effects diminish HSI classification accuracy. Thus there is a desire to create robust classifiers that perform well in all possible environments. Robust parameter design (RPD) techniques have been applied to determine optimal operating settings. Previous RPD efforts considered an HSI image as categorical noise. This paper presents a novel method utilizing discrete and continuous image characteristics as representations of the noise present. Specifically, the number of clusters, fisher ratio and percent of target pixels were used to generate image training and test sets. Replacing categorical noise with the new image characteristics improves RPD results by correctly accounting for significant terms in the regression model that were otherwise considered categorical factors. Further, traditional RPD assumptions of independent noise variables are invalid for the selected HSI images.

Introduction:

Hyperspectral imagery (HSI) has emerged as a valuable tool supporting numerous military and commercial missions including counter concealment, camouflage and deception, combat search and rescue, counter narcotics, cartography and meteorology to name a few (Manolakis (2002); Landgrebe (2003)). A hyperspectral image, also called an image cube, consists of *k* spectral bands of an *m* by *n* spatial pixel representation of a sensed area. Each pixel in the spectral dimension represents an intensity of energy reflected back to the sensor. All spectral dimensions for a given pixel represent a potential target signature. HSI, by its very nature, can provide a method for identifying at most (n - 1) unique spectral signals, where n is the number of independent bands in an HSI image cube. This is (n - 1) rather than n because one band is used to define the background or noise present in an image. Since HSI contains typically hundreds of bands, this number of signals or targets for classification can be large although bands affected by atmospheric absorption contain little useful information and must be removed and bands that are close to each other are typically correlated.

Davis (2009) describes some pitfalls when performing target classification on hyperspectral images. For instance, the spectral library will most often not contain every possible object, manmade or other to be classified. Some objects may be concealed or disguised to make the spectral signature different from what is contained in the library. In addition, environmental effects such as time of day, relative humidity and imaging angle greatly impact the data reflectance values observed by a sensor. Finally, target prior probabilities can be very small in comparison to the number of pixels being considered.

This leads to a desire to create robust classifiers that perform well in all possible environments or to make very specialized systems that are only used on very specific areas to ensure the spectral library containing spectral signatures of the materials within an image is as accurate and separable as possible.

Typical hyperspectral target detection algorithms can be separated into two classes, anomaly detection and signature matching. Signature matching compares the observed intensities for all bands of an individual pixel with a known spectral signature contained in a library. Anomaly detection compares an individual pixel's mean observed intensity with the mean and variance of the background. Pixels which are statistically different from the background are identified as anomalies.

Previous efforts to develop robust HSI classifiers have utilized robust parameter design (RPD) techniques where each image was considered a categorical noise variable. This paper presents a novel method utilizing discrete and continuous image characteristics as representations of the noise present in an image. Specifically, the number of unique clusters within an image, fisher ratio and percent of target pixels were used to identify image training and test sets. Replacing categorical noise with the new image characteristics improves RPD results by correctly accounting for significant terms in the regression model that were otherwise considered categorical factors. In addition, it is simpler to create models when the noise variables are not categorical.

Robust Parameter Design

Genichi Taguchi proposed an innovative parameter design approach for reducing variation in products and processes in the 1980's. Montgomery (2009) describes RPD as an approach to experimental design that focuses on selecting control factor settings that optimize a selected response while minimizing the variance due to noise factors at that optimum. Control factors are those factors that can be modified in practice while noise factors are often unexplained or uncontrollable in practice. These noise factors can typically be controlled at the research and development level allowing RPD to be performed. There are two methods to model RPD, crossed arrays and combined arrays. This paper will focus on the combined array or response surface method.

RSM methods focus on the roles of control variables on mean and variance in order to provide an estimate at any location of interest. Typically, second-order models are developed when using RSM approaches and higher order interactions are ignored due to the sparcity of effects principle; noise by noise interactions are also assumed to be negligible. A general matrix form of the fitted quadratic response surface model is in the following equation (Myers and Montgomery, 2002)

$$\hat{y}(x,z) = \beta_0 + x'\beta + x'\mathrm{B}x + z'\gamma + x'\Delta z + \varepsilon \tag{1}$$

where β_0 is the model intercept, β is a vector of the control variable coefficients, B is a matrix of the quadratic control coefficients, γ is a vector of noise variable coefficients, Δ is a matrix of the control by noise interaction coefficients and ε is the pure error of the model which is assumed to be $NID(0,\sigma^2)$. The mean model for the equation can easily be found since the noise variables, z, are assumed to be random variables with $E(z)=0$ and $\mathrm{var}(z) = \sigma_z^2$; further, the noise variables are considered coded random variables centered at zero with limits $\pm a$ representing high and low settings for a particular noise variable and $\mathrm{cov}(z_i, z_j) = 0, \ \forall\ i \neq j$. Thus the general form of the mean model only includes the control variables and is shown in Montgomery (2009) to be

$$E[\hat{y}(x,z)] = \beta_0 + x'\beta + x'Bx \tag{2}$$

Likewise, the variance model can be found by treating z as a random variable and applying the variance operator to the equation above. The variance model becomes

$$\text{var}[\hat{y}(x,z)] = (\gamma + \Delta' x)' \sigma_z^2 (\gamma + \Delta' x) + \sigma^2 \tag{3}$$

where σ^2 is the Mean Square Error found from performing a regression on the design and σ_z^2 is the variance-covariance matrix of z typically assumed to be 1 since the variables are coded. (Myers and Montgomery, 2002)

Categorical Noise

Brenneman and Myers (2003) developed a methodology for treating noise variables categorically for some situations such as when considering different suppliers or brands of equipment as noise. They assert that fewer assumptions are required when considering noise as a categorical variable. Multiple continuous noise variables can be combined into a single categorical noise variable with $r_z + 1$ categories where $P(\text{category } m) = p_m$. It is assumed that these probabilities are known *a priori*. Further, the distribution of this single categorical noise variable is multinomial. The variance-covariance matrix, σ_z^2, from equation (3) becomes

$$\sigma_z^2 = \begin{bmatrix} p_1(1-p_1) & -p_1p_2 & \cdots & -p_1p_{r_z} \\ -p_2p_1 & p_2(1-p_2) & \cdots & -p_2p_{r_z} \\ \vdots & \vdots & \ddots & \vdots \\ -p_{r_z}p_1 & -p_{r_z}p_2 & \cdots & p_{r_z}(1-p_{r_z}) \end{bmatrix}. \tag{4}$$

The prior probabilities required to characterize the variance-covariance matrix might not be available in all situations. Brenneman and Myers also recognized it is possible for robust control settings to be dependent on the p_m. Moreover, the true noise in a hyperspectral image is better characterized by the observable features within an image. Thus, the proposed noise methodology was developed.

Image Noise Methodology

There are several potential observable noise characteristics within an image and this paper will focus on three: Fisher score, percent of target pixels and number of clusters. These are not the only characteristics but rather a subset that can be easily calculated within a training set with truth information. Fishers ratio is defined by Lohninger (1999) as a measure for the discriminating power of a variable

$$f = \frac{\mu_1^2 - \mu_2^2}{\sigma_1 + \sigma_2} \tag{5}$$

where μ_1 and σ_1 are the mean and variance of the target class and μ_2 and σ_2 are the mean and variance of the background both defined in a truth matrix. The percent of target pixels can be calculated, if there is a truth map, for each image under test defined as

$$t_i = \frac{v_i}{w_i} \tag{6}$$

where v_i and w_i represent the number of target pixels and background pixels in image i respectively. Clustering was performed using Williams (2007) Matlab ® code of X-means as described by Pelleg and Moore (2000).

Unfortunately, these observed noise characteristics do not fit into the traditional experimental designs since the observations are typically correlated and not orthogonal. Figure 1 shows a classical 2^2 factorial design in circles and an example of an observed set of design points in triangles.

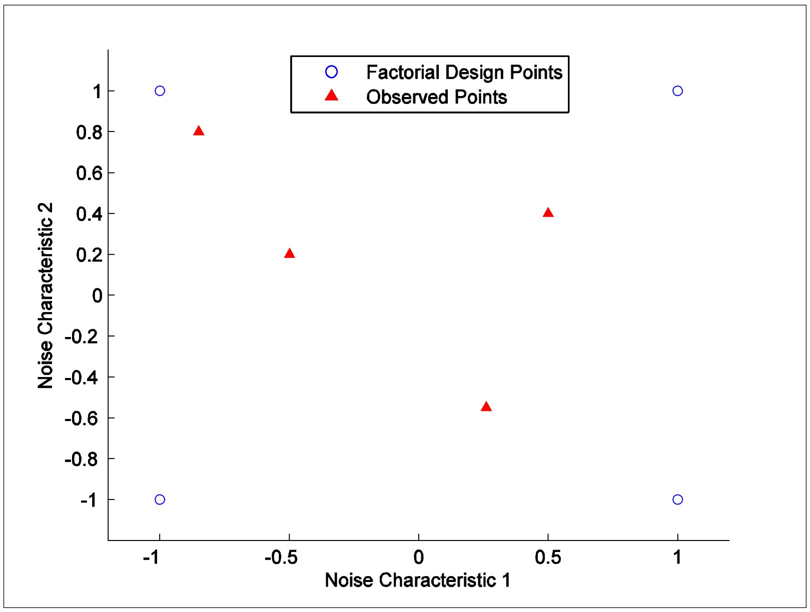

Figure 1: Factorial design (circles) versus observed design (triangles)

The decision for which images to include in the training and test sets is not trivial. Identifying a training and test set of images can be considered a combinatorial optimization problem. Formally, a combinatorial optimization problem is defined as a pair (Ω, f) where Ω is the set of feasible solutions consisting of all possible combinations of images and f is the cost function (Hall, 2009). Now let R_j be the range for noise factor $j = 1, 2, 3, \ldots n_0$ within a given set of images, ω. Further, let the cost function be defined for the previously defined noise variables (1-fisher's score, 2-percent target, 3-number of clusters) as

$$f_\omega = R_1 + R_2 + R_3 \tag{7}$$

summing the total range across all three noise variables for a given set of images, ω. Let d_i be an indicator variable for image *i* such that

$$d_i = \begin{cases} 1 & \text{if } d_i \text{ is in the training set} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

The combinatorial optimization problem can thus be solved as a binary integer program

$$\begin{aligned} &\max_{\omega} \; f_\omega = R_1 + R_2 + R_3 \\ &\text{ST.} \; \sum_{i=1}^{n} d_i = k \end{aligned} \tag{9}$$

where *k* is the number of images to be used in training and *n* is the total number of images available. This formulation can result in multiple alternate optimal training sets since some images have extreme values of all noise characteristics. Thus, another binary integer program can be solved on the set of alternative optimals to choose a test set of images. Let $\overline{\omega}$ be the complement of ω for optimal training sets. Assuming all images are to be used in either the training or test set, $\overline{\omega}$ is the set of test images to go with a selected set of training images ω. Let g_i be the indicator variable for image *i* in the test set and *m* be the number of images to include in the test set. Then equation (9) can be adapted to solve for the optimal set of training and test images.

$$\begin{aligned} &\max_{\overline{\omega}} \; f_{\overline{\omega}} = R_1 + R_2 + R_3 \\ &\text{ST.} \; \sum_{i=1}^{n} g_i = m \end{aligned} \tag{10}$$

Results

Training and test images were selected from toy data sets as well as eight images from the Hyperspectral Digital Imagery Experiment (HYDICE). Each image was halved to double the total number of images to 16; image 1 became image halves 1 and 2, image 2 became image halves 3 and 4 and so on. The observed noise values for all 16 image halves are in Table 1.

Table 1: Observed image noise characteristics

Image	Image Size	Fisher	% Target	# Clusters
1	28855	1.1203	0.0298	3
2	29054	1.1085	0.0177	8
3	11128	2.7357	0.0077	3
4	11232	2.7357	0.0086	3
5	10908	1.097	0.0795	10
6	11016	1.097	0.0803	10
7	12276	1.0443	0.0506	9
8	12276	1.0443	0.0498	9
9	15200	1.123	0.0002	4
10	15360	1.123	0.0004	10
11	23712	2.3458	0.0366	8
12	23712	2.0853	0.0446	8
13	15368	1.1001	0.0637	7
14	15368	1.1001	0.0564	10
15	8160	1.0664	0	9
16	8240	1.0525	0.0015	10

The algorithm selected images {1,2,4,6,7,8,10,15} for training and {3,5,9,11,12,13,14,16} for test. Note this puts both halves of images one and four in training and six and seven in test. Figure 2 shows a pair wise comparison of the training and test image noise characteristics. There is typically a training and test image near each observed extreme of the chart. This provides adequate separation of the noise characteristics to perform RPD.

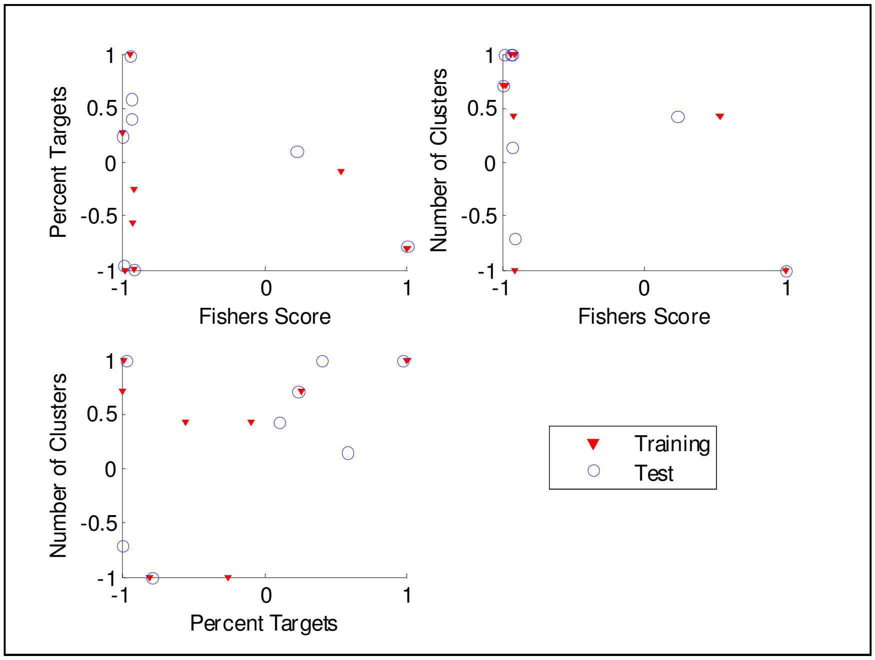

Figure 2: Training and test image pair wise noise characteristics

These images were then used in an RPD of the Autonomous Global Anomaly Detector (AutoGAD) (see Johnson (2008) for the specifics of this algorithm). The experimental design mirrored previous RPD work by Davis (2009) and Miller (2009) using D-optimal designs from Design Expert (see Davis (2009) for specific design information). However, a one-to-one comparison of results will not be presented as Davis and Miller did not halve the images, but rather trained and tested on the complete set of images. Fitting a regression model to the results from training yielded some interesting analysis of variance results. The assumption that no noise by noise interaction exists was not true. R-squared values were as low as 0.5 without noise by noise interactions and were improved by as much as 0.26 when the interactions were included. Table 2 compares the R-squared values with and without noise by noise interactions on four AutoGAD outputs: time, true positive fraction (TPF), false positive fraction (FPF) and target fraction percent (TFP). True positive fraction compares the number of correctly identified pixels with the total number of actual target pixels; false positive fraction compares the total number of falsely labeled (labeled as targets when they were actually noise) pixels with the total number of background pixels. Finally, target fraction percent measures AutoGAD's performance on target clusters. If AutoGAD correctly identifies at least one pixel of a true target cluster, it is counted as a success. TFP is the ratio of these successes to the total number of true target clusters.

Table 2: R-squared values for AutoGAD regression models

Measure	R-squared without interactions	R-squared with interactions
Time	0.5071	0.7719
TPF	0.7145	0.8947
FPF	0.7569	0.8498
TPT	0.5092	0.7527

Future Work

A new mean and variance model with noise by noise interactions will be developed to generate a more adequate regression model for RPD. This model will be compared with a neural network representation. Also, D-optimal designs will be compared with the image noise methodology proposed to assess performance characteristics such as time to generate a set of training and test images as well as the separation of the images within each set.

References

Brenneman, William and William Myers. Robust Parameter Design with Categorical Noise Variables," Journal of Quality Technology, 35 (4):335-341 (2003).

Davis, Matthew. Using multiple robust parameter design techniques to improve hyperspectral anomaly detection algorithm performance. Ms thesis, Air Force Institute of Technology (AU), Wright-Patterson AFB, OH, March 2009. AFIT/GOR/ENS/09-05.

Hall, Shane N. OPER 623 approximation and heuristic search methods. Course notes, September 2009.

Johnson, Robert J. Improved feature extraction, feature selection and identification techniques that create a fast unsupervised hyperspectral target detection algorithm. Ms thesis, Air Force Institute of Technology (AU), Wright-Patterson AFB, OH, March 2008. AFIT/GOR/ENS/08-07.

Landgrebe, David A. Signal Theory Methods in Multispectral Remote Sensing. Wiley. Hoboken, NJ 2003.

Lohninger, H.. Teach/Me data analysis. Springer-Verlag, Berlin-New York-Tokyo, 1999.

Manolakis, D. Detection algorithms for hyperspectral imaging applications. Report ESCTR-2001-044, Lincoln Laboratory: Massachusetts Institute of Technology, Lexington, Massachusetts, Feb 2002.

Miller, Michael K. Exploitation of intra-spectral band correlation for rapid feature selection and target identification in hyperspectral imagery. MS thesis, Air Force Institute of Technology (AU), Wright-Patterson AFB, OH, March 2009. AFIT/GOR/ENS/09-10.

Montgomery, Douglas. Design and analysis of experiments. Wiley, 7th edition, 2009.

Myers, Raymond and D. Montgomery. Response surface methodology. Process and product optimization using designed experiments. Wiley, 2nd edition, 2002.

Pelleg, Dan and A. Moore. X-means: Extending k-means with efficient estimation of the number of clusters. Technical report, Carnegie Mellon University, Pittsburgh, PA, 2000.

Smetek, Timothy E. Hyperspectral imagery target detection using improved anomaly detection and signature matching methods. Dissertation, Air Force Institute of Technology (AU),Wright-Patterson AFB, OH, June 2007. AFIT/DS/ENS/07-07.

Williams, Jason P. Robustness of multiple clustering algorithms on hyperspectral images. MS thesis, Air Force Institute of Technology (AU), Wright-Patterson AFB, OH, March 2007. AFIT/GOR/ENS/07-27.