6-1-2005

Hybridization of Hyperspectral Imaging Target Detection Algorithm Chains

David C. Grimm
Rochester Institute of Technology

David W. Messinger
Rochester Institute of Technology

John P. Kerekes
Rochester Institute of Technology

John R. Schott
Rochester Institute of Technology

Follow this and additional works at: https://scholarworks.rit.edu/other

Hybridization of hyperspectral imaging target detection algorithm chains

David C. Grimm[a][b], David W. Messinger[a], John P. Kerekes[a], and John R. Schott[a]

[a] Digital Imaging and Remote Sensing Lab, Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology, 54 Lomb Memorial Drive, Rochester, NY 14623
[b] Air Force Institute of Technology, USAF

ABSTRACT

Detection of a known target in an image can be accomplished using several different approaches. The complexity and number of steps involved in the target detection process makes a comparison of the different possible algorithm chains desirable. Of the different steps involved, some have a more significant impact than others on the final result - the ability to find a target in an image. These more important steps often include atmospheric compensation, noise and dimensionality reduction, background characterization, and detection (matched filtering for this research). A brief overview of the algorithms to be compared for each step will be presented.

This research seeks to identify the most effective set of algorithms for a particular image or target type. Several different algorithms for each step will be presented, to include ELM, FLAASH, MNF, PPI, MAXD, the structured background matched filters OSP, and ASD. The chains generated by these algorithms will be compared using the Forest Radiance I HYDICE data set. Finally, receiver operating characteristic (ROC) curves will be calculated for each algorithm chain and, as an end result, a comparison of the various algorithm chains will be presented.

Keywords: ELM, FLAASH, MNF, PPI, MAXD, OSP, ASD, target detection

1. INTRODUCTION

Detection of known substances, or targets, is a very common problem in hyperspectral imaging (HSI). Given a target with a known spectral signature, an algorithm that can decide if and where that target is present in an image is needed. In order to accomplish this, matched filters or target detectors have been developed to accentuate pixels in an image that contain the target.

There are several unique steps to any target detection algorithm. Each of these steps plays a role in determining the overall performance of the target detector. Figure 1 is a flow chart detailing each of these steps for a generic target detection algorithm chain. For each individual step along the way, there are several algorithms that will return an acceptable solution. Each step is important in determining the overall quality of the result. For example, the purpose of the atmospheric compensation step is to convert the units of the scene from sensor reaching radiance into reflectance to be consistent with the units of the target spectrum. The noise/dimensionality reduction step aims to not only reduce the effect of the noise in the image, but also to reduce the data set so that it is more computationally manageable without compromising the integrity of the data. Background characterization is important because it is the product of this step that is used to suppress the background of the image so the target pixels can be properly detected. And finally, the matched filter is ultimately used to produce the detection maps. There is an inherent interaction between these various steps and it is this interdependence that this paper addresses and studies.

It is desired to determine the *best* algorithm to accomplish each step. This is where hybridization comes in. A particular combination of algorithms may work well for one image or target but may also perform terribly for a different image or target. Finding an optimal combination of algorithms - or recipes - becomes an intriguing dilemma. In order to best accomplish this, commonly used algorithms for each step must be looked at and compared to each other for many different image/target combinations. This paper will discuss in detail some of the different algorithms used for each step of the target detection process. Background and theoretical explanation of the algorithms will be discussed followed by results, in the form of ROC curves and average false alarm rates (AFAR),[1] that will compare several unique algorithm chains.
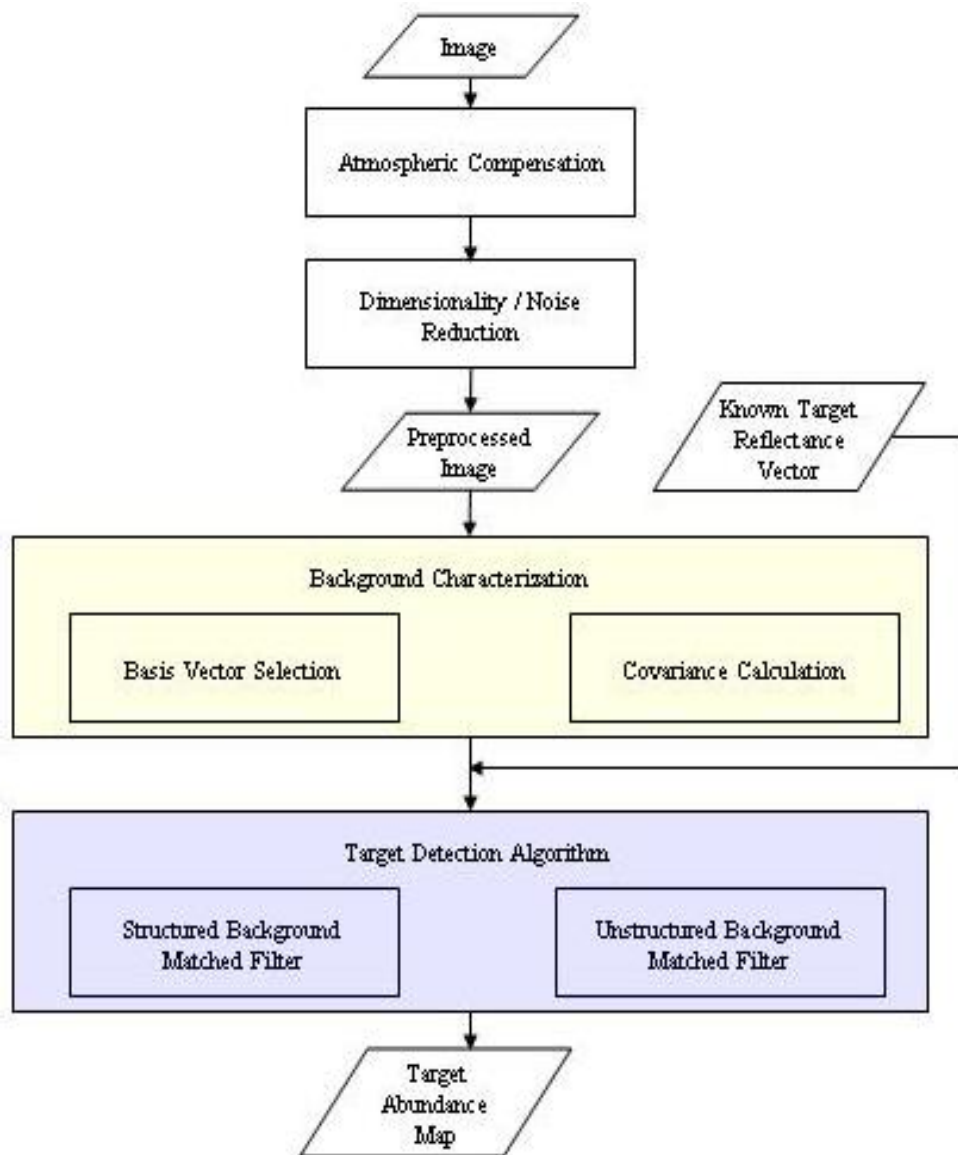
Figure 1. Typical target detection algorithm flow chart.

1.1. Data Set

The data set used for this paper is the "Forest Radiance 1" scene collected by the HYDICE sensor. The image dimensions are 320x1280 with 210 spectral bands covering the spectrum from approximately 350nm to 2500nm. Of these 210 spectral bands, some were excluded, particularly covering the oxygen absorption bands, leaving 145 "good" bands. Two targets will be used to compare the algorithm chains, one of which is "easy" to find and one of which is slightly more difficult to find. There are many man-made targets placed into this scene. In order to determine the overall difficultly of detection, the Spectral Angle Mapper (SAM) detection algorithm was run for each target in the scene. SAM is a very basic target detection algorithm that relies on the spectral angle between the target and pixel in question to determine if the target is present. The test statistic is computed as

$$T_{SAM}(x) = \frac{d^T x}{(d^T d)^{1/2}(x^T x)^{1/2}} \tag{1}$$

where d is the target spectrum and x is the pixel spectrum in question. For the purposes of this paper, the SAM results were used to determine the targets to be used for comparison and to provide a baseline ROC curve for reference.

A second basic matched filter is the Generalized Likelihood Ratio Test (GLRT). The results returned by this matched filter will also be used for comparison purposes. The GLRT relies on an unstructured, or stochastic, representation of the background in the image in the form of a covariance matrix and, for this implementation, a pre-defined target spectrum. The GLRT is written as

$$T_{GLRT}(x) = \frac{(d^T \Sigma^{-1} x)^2}{(d^T \Sigma^{-1} d)(1 + x^T \Sigma^{-1} x)} \tag{2}$$

where d and x are the same as Equation 1 and Σ is the covariance matrix used. The GLRT results shown were generated using the covariance of a spatially selected region consisting of "tree" pixels in the image.

Based on the results of these two matched filters, the "easy" target, Target A, had an AFAR of 0.003 for SAM and 0.009 for the GLRT. The lower the AFAR, the easier the target is to detect. The AFAR is defined as the average number of false alarms encountered for each target pixel detected. The "difficult" target, Target B had AFARs of 0.023 and 0.077 for SAM and GLRT respectively. Figure 2 is an image of the data set and also shows the spatial locations of the targets being used. The SAM and GLRT detectors were both run on an image that was atmospherically compensated using ELM.

By simply looking at the image, it can be seen that, barring the planted targets, the scene is very uncluttered with very little man-made material. There are only about six or seven different classes of pixels present in the image and several of those are rather similar (*i.e.* light and dark grass are considered unique classes). Secondly, there are no "concealed" target pixels, all of them are out in the open and completely uncovered. There are both fully resolved and sub-pixel target pixels for both targets. For the purposes of this paper, all planted targets in the image except for Targets A and B have been masked out for the detection results and background characterization. All pixels with any amount of target present are searched for and counted as part of the detection rate statistic. As a part of the extensive ground truth data associated with this data set, target masks were provided detailing the pixel locations of all target pixels in the image. These target masks were used to calculate the detection and false alarm rates.

2. ATMOSPHERIC COMPENSATION

Atmospheric compensation is the process of "taking the atmosphere out of the image". The known spectral signature of the targets consist of measured reflectance values. The observed signal incorporates atmospheric and illumination variations that dramatically alter the spectral shape of the target reflectance spectrum. The ultimate goal of the atmospheric compensation process is to retrieve reflectance from the radiance recorded by the sensor so that the detectors can operate in the same spectral space in which the target is known. The following sections will briefly outline the theory behind the two atmospheric compensation methods to be compared.
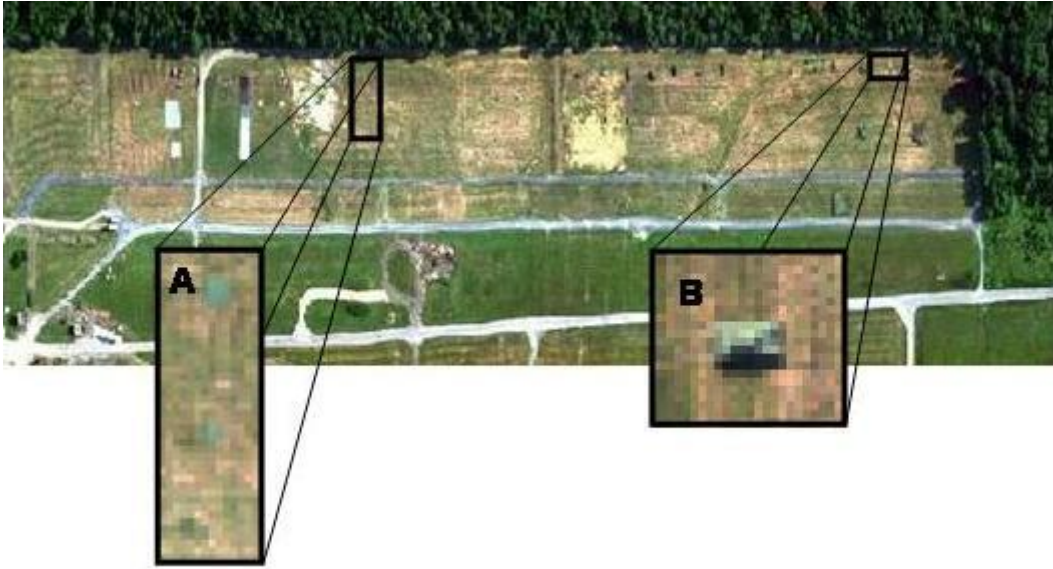
Figure 2. Spatial locations of "easy" target A and "difficult" target B in the Forest Radiance 1 scene.

2.1. Empirical Line Method (ELM)

The ELM relies on ground truth inasmuch as at least two different regions in the image (preferably one dark and one bright across each wavelength) must have a known reflectivity. These regions must be at least one fully resolved pixel large. Ideally, the regions should have corresponding ground truth reflectance spectra taken at the same time as the image. If such regions are not present in the image, an educated guess can be made by the selection of regions for which an approximate reflectance can be estimated. For example, a white cloud may have a reflectivity of about 90% across all wavelengths. Obviously, there is a great opportunity for error introduction if estimates have to be made, but that can be limited if they are made smartly. The Forest Radiance I image has calibration panels present and the values measured from those panels were used for the ELM compensation.

Once the two (or more) regions have been selected, a line is fit through the radiance (or digital counts) vs reflectivity points in each band. ELM assumes a linear relationship between the radiance (or digital counts) and the reflectivity. Mathematically, this relationship is expressed as

$$L = mR(\lambda) + b \tag{3}$$

where L is the observed radiance (or digital count), m is the slope of the line through the ground truth points, $R(\lambda)$ is the wavelength dependent reflectance, and b is the radiance (or digital count) value that represents zero reflectance. Figure 3 illustrates this relationship graphically.

The ELM is calculated and applied separately to each spectral band due to the wavelength dependency of Equation 3. Now, the observed radiance in every pixel in the image can be converted into reflectance using the linear regression coefficients (m and b) obtained via Equation 3. It is important to note that ELM can work with both radiance or digital counts without calibration.

There are two primary circumstances that cause the ELM to fail: the absence of reliable ground truth and a scene that has a very non-linear relationship between reflectance and radiance. Anytime an estimation has to be used in place of actual ground truth, there is opportunity for a great deal of error, and if the underlying assumption of the algorithm does not hold, the outcome will be flawed.

2.2. Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH)

FLAASH is a commercially available atmospheric compensation algorithm developed by the Air Force Research Laboratory, Space Vehicles Directorate (AFRL/VS) to support the analysis of visible to shortwave infrared (0.4
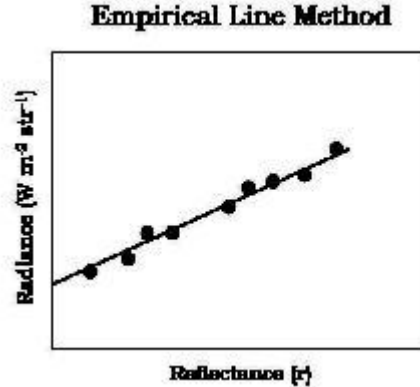
Figure 3. Graphical representation of ELM for a single band. The marked points are the "ground truth" points.

μm to 2.4μm) hyperspectral and multispectral sensors.[2]   Using atmospheric inversion, FLAASH is able to determine the aerosol and water vapor content of the atmosphere in an image as well as the surface pressure.

The sensor reaching radiance (L^*) equation utilized by FLAASH[2] is

$$L^* = \frac{A\rho}{1 - \rho_e S} + \frac{B\rho_e}{1 - \rho_e S} + L_a^* \tag{4}$$

where ρ is the pixel surface reflectance, ρ_e is an average surface reflectance for the area surrounding the pixel in question, S is the spherical albedo of the atmosphere (accounting for the skylight photons), L_a^* is the upwelled radiance, A and B are surface independent coefficients that vary based on the atmospheric and geometric conditions. All of the variables are implicitly wavelength dependent.

The per-pixel A, B, S, and L_a^* variables are all determined empirically via MODTRAN. Once these values have been extracted, Equation 5 is used to solve for $\rho_e{}^2$:

$$L_e^* = \frac{(A + B)\rho_e}{1 - \rho_e S} + L_a^* \tag{5}$$

where L_e^* is a spatially averaged sensor reaching radiance value. Once ρ_e has been determined, Equation 4 can be solved for reflectance (ρ).

3. MAXIMUM NOISE FRACTION (MNF)

Only one algorithm for dimensionality/noise reduction will be looked at in this paper, the Maximum Noise Fraction (MNF).[3] Essentially, the MNF algorithm is the same as the Principle Components Analysis process with some "pre-processing" steps added. For the MNF transform, the covariance matrix of the noise present is needed. In order to obtain this (since it is not usually known), a spectrally flat field can be used. Once a flat field is identified within an image, the mean can be subtracted from the observed values of that field. The result is a representation of the noise present in the image. Once this "noise image" is found, its covariance matrix can be calculated. If a flat noise field cannot be found, the ENVI implementation of MNF allows for the noise statistics to be computed from the image.

The next step in this process is basically to perform PCA on the noise image to de-correlate the noise. This is done by calculating the eigenvectors of the noise image. By definition, the resulting eigenvectors are orthogonal to each other. The original image can now be "passed through" this orthogonal transformation. After this is done, the orthogonally transformed image can be whitened, or normalized, by the eigenvalues of the noise covariance matrix. The final result of all these steps is an image in which all of the noise present is orthogonal and identical. Now, PCA can be performed on the entire image.

The advantage of the MNF transform is that the "pre-processing" steps force the assumption used in PCA of variance representing information to be true. There is now equal variance present due to noise in the image. The PC bands are rank ordered by the signal to noise ratios and the highest bands are retained. The result is a greatly reduced dimension space with little noise to interfere with target detection.

4. ENDMEMBER SELECTION

There are two overarching ways to characterize the background of an image: structured or unstructured. While an unstructured matched filter (GLRT) was used in determining the targets and for comparison purposes, this paper will focus on matched filters that employ a structured background. Two related endmember selection algorithms will be compared in this paper.

4.1. Pixel Purity Index (PPI)

The first end member selection algorithm is the Pixel Purity Index (PPI) method.[4] Using PPI, all of the pixel values in an image are projected onto randomly selected vectors. Each time a projection is accomplished, the extreme points are noted. Spectrally pure pixels will consistently be extreme points of these random projections. A major assumption of this method is that spectrally pure pixels in the image are end members. This is generally a good assumption. This method is also rather susceptible to noise, so steps (*e.g.* thresholding) should be taken to help lessen the negative effects of noisy pixels. Figure 4 is an example projection of data points. ENVI uses a threshold factor to help mitigate noise interference. The threshold factor determines the number of pixels other than strictly the extreme pixels considered - pixels within the threshold factor of the extreme pixels are considered extrema as well.
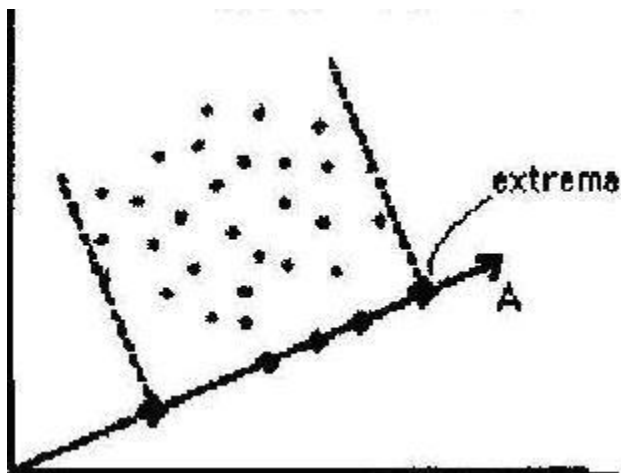

Figure 4. Example PPI projection[4]

For this paper, the PPI implementation in ENVI was used. Using ENVI, the PPI endmember selection process is not fully automated and does require a "man-in-the-loop". Once the desired number of iterations (normally several thousand) are completed, every pixel that was found as an extrema at least once is returned. It is up to the user to determine how many times a pixel must be found to be considered significant. Once this is completed, the significant pixels must be clustered. The amount of spectral clusters determine the number of endmembers in the image. All of this is done by the user. Finally, an endmember is extracted by taking the mean of each cluster.

4.2. MAXD

The second end member selection technique is MAXD.[5] Similar to PPI, MAXD also uses projection to help sort out the end members of an image. Given a set of data points, MAXD begins by selecting the points that have the maximum and minimum Euclidean distance from the origin. These two points are the first two end

members. Once these points are selected, all of the data points are projected onto a vector orthogonal to the difference vector between the first two, and the pixel the maximum distance from the common point projection of the first two is selected as the next endmember. This process is repeated until the desired number of end members is found. Figure 5 pictorially explains how this process works. Unlike PPI, MAXD is used as a fully automated process where the only user input is the number of endmembers to be found.
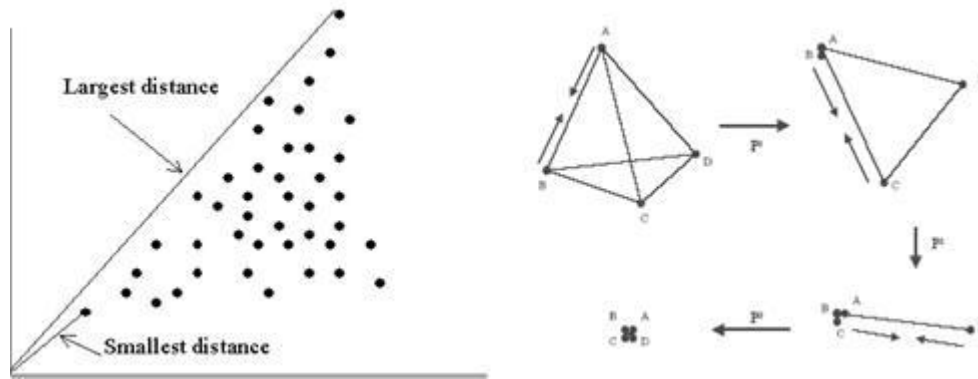

Figure 5. MAXD end member selection algorithm[5]

5. MATCHED FILTERS

The ultimate goal of the target detection algorithm chain is *target detection*. The means to accomplish target detection in this case are matched filters. The two similar but different structured background matched filters compared in this paper are the Orthogonal Subspace Projector (OSP) and the Adaptive Subspace Detector (ASD).

5.1. Orthogonal Subspace Projector (OSP)

The first matched filter discussed is OSP. The OSP detector is written as

$$T_{OSP}(x) = \frac{dP_b x}{d^T P_b d} \tag{6}$$

where d is the desired target vector and $P_b = I - U(U^T U)^{-1} U^T$ with I being the (bands x bands) identity matrix and U being the matrix of endmember vectors.[6]

OSP projects each pixel into a space orthogonal the the background of the image which is represented by the endmembers extracted. This, in theory, forces the target pixels to "stick out" from the background. The pixels that most closely match the given target spectra receive a high T_{OSP} value.

5.2. Adaptive Subspace Detector (ASD)

A second implementation of a structured background model detector is ASD. This detector is represented mathematically by

$$T_{ASD}(x) = \frac{x^T (P_b - P_S) x}{x^T P_S x} \tag{7}$$

where $P_S = I - S(S^T S)^{-1} S^T$ with $S \equiv [dU]$ the matrix fromed by the concatenation of the target spectra and the endmember spectra, and the other variables are the same as those defined for Equation 6.[6] In words, $P_b x$ is the projection of the pixel, x, onto the vector orthogonal to the background subspace.

Both OSP and ASD suppress the image background by projecting each interrogated pixel onto a space orthogonal to the background via the P_b operation. The major difference between OSP and ASD is the P_S term in ASD. This is the projection of the pixel onto a space orthogonal to the background *and* the target. The difference between these two projections is a measure of just how "target-like" the pixel is.

6. RESULTS

6.1. Experiment

The first step in the algorithm chain is atmospheric compensation. The "Forest Radiance I" data set was atmospherically compensated using both the ELM and FLAASH methods discussed in Section 2. The second step is noise and dimensionality reduction. This is accomplished using the MNF transform outlined in Section 3. The implementation available in ENVI was used for this paper. The option to "estimate noise statistics from data" provided by ENVI was used to obtain a noise spectrum for the transform. For both atmospherically compensated images, over 98% of the total variability was retained resulting in keeping 136 and 137 bands for the ELM and FLAASH images, respectively.

As mentioned in Section 4, PPI and MAXD were used to extract endmembers to characterize the background of the image. The procedure for usage of the PPI application in ENVI discussed in Section 4.1 was followed using 30,000 iterations and a threshold factor of 2. Seven different classes were found in the clustering step, resulting in a background matrix consisting of 7 endmembers. The MAXD algorithm is not currently available in ENVI, so IDL was used. The only user input necessary for MAXD is the number of endmembers to be found, for this image, 20 endmembers were found. Research has shown that there is an optimal number of endmembers that can be found via MAXD for a scene. Selection of either too many or too few has a negative impact on detection results reached using the background characterized by MAXD.[7]

Finally, each image/background combination was used with the matched filters presented in Section 5. Figure 6 shows the ROC curves that resulted from each combination for the two targets as well as the SAM and GLRT results for comparison purposes.

Figure 7 depicts the AFAR values associated with each of the ROC curves in Figure 6. These AFAR values can be used to preliminarily evaluate the performance of each algorithm chain.

6.2. Conclusions

Overall, there were no results that seemed particularly telling. For Target A, the "easy" target, all of the chains performed comparably with the exception of one, the FLAASH-MNF-MAXD-OSP chain. Looking at the Target B results, this chain underachieved a little there as well. This logically leads to the conclusion that that group of algorithms do not work well together. On the other hand, there did not seem to be one particularly stellar preforming chain of algorithms either.

It was expected that the chains using ELM would have greatly outperformed the chains using FLAASH due to the fact that good ground truth went into the ELM calculation. While this is generally true for the two targets, a much greater difference would have been expected. In most cases, these initial results demonstrate an expected improvement in performance with user intervention and knowledge. The chains using FLAASH for instance, while generally performing somewhat poorer, are the chains involving the least user knowledge.

It would be interesting to repeat the experiment using ELM on the scene without using the collected ground truth and see if the results change drastically. Another intriguing question is what the results would look like for a much more cluttered (*i.e.* urban) scene.

7. SUMMARY AND FUTURE WORK

This paper is intended to be a beginning point for the exploration of different algorithm chains, lending itself to more ongoing research. A very few number of algorithms were compared in this paper relative to the great number currently in general use. Some future work stemming from this research may include looking at many more distinct and complex chains. The ultimate goal is to *pre*-determine the *best* set of algorithms to solve a target detection problem for any given image and/or target. In solving this problem, a methodology for analysis
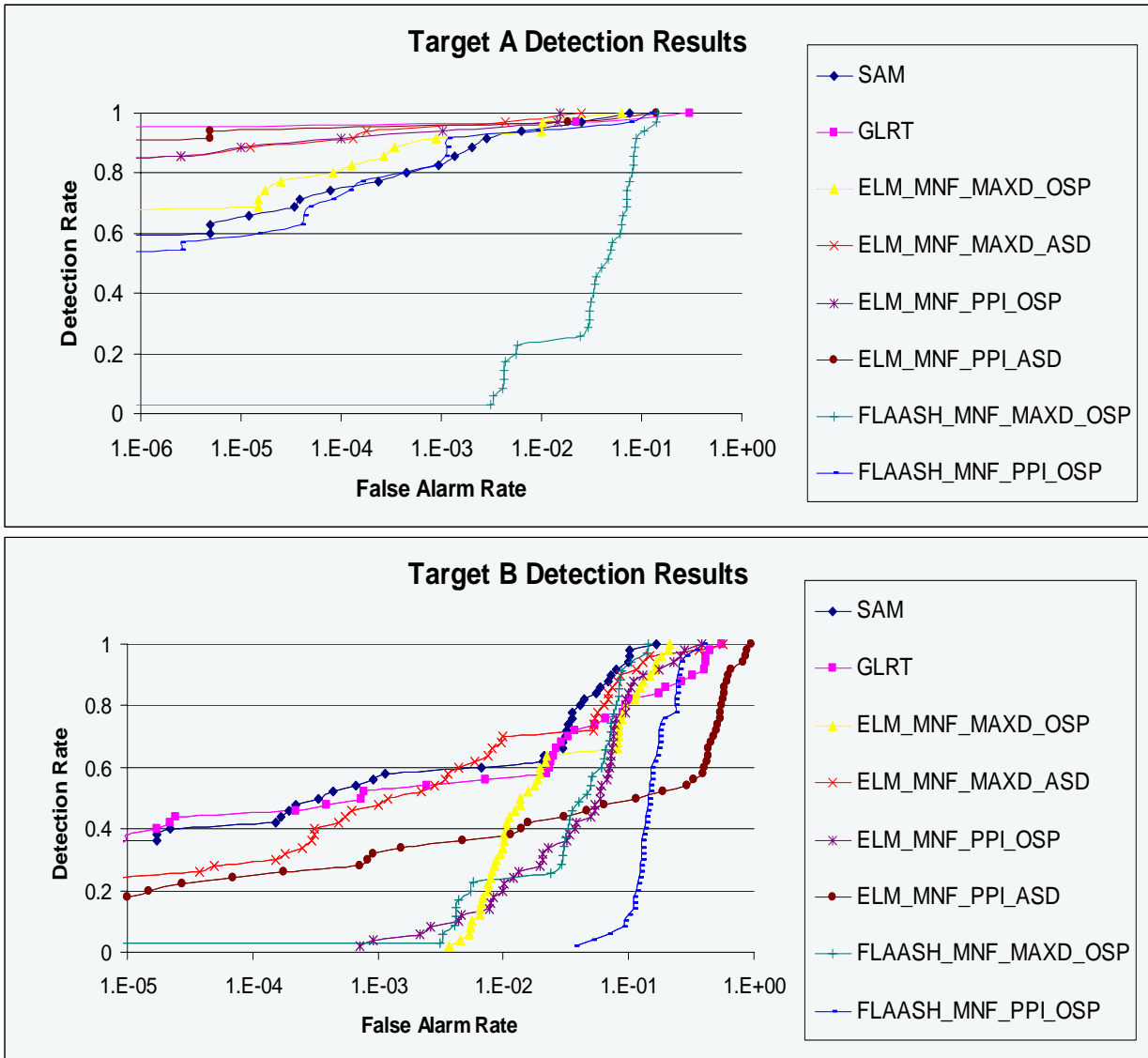
Figure 6. Resulting ROC curves and AFAR for the various algorithm chains.
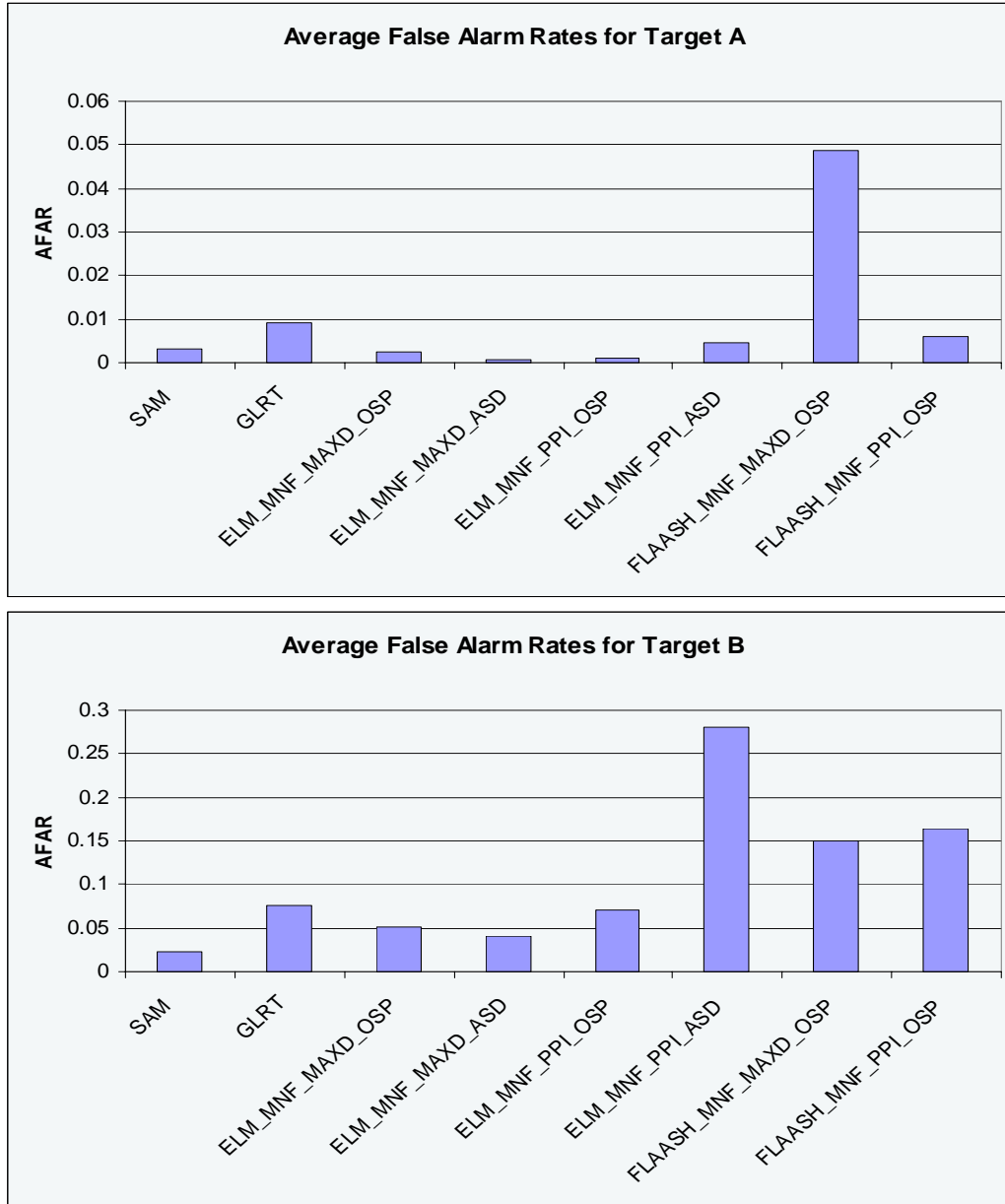
Figure 7. Resulting AFAR values for the various algorithm chains.

of the extensive unique algorithm chains would be extremely helpful. This paper is aimed at starting along the path for developing such a methodology.

DISCLAIMER

The views expressed in this article are those of the authors and do not reflect the official policy or position of the U. S. Air Force, Department of Defense, or the U. S. Government.

REFERENCES

1. P. Bajorski, E. J. Ientilucci, and J. R. Schott, "Comparison of basis-vector selection methods for target and background subspaces as applied to subpixel target detection," in *Proceedings of SPIE*, **5245**, pp. 97–108, August 2004.
2. A. Berk, S. M. Adler-Golden, A. J. Ratkowski, G. W. Felde, G. P. Anderson, M. L. Hoke, T. Cooley, J. H. Chewynd, J. A. Gardner, M. W. Matthew, L. S. Bernstein, P. K. Acharya, D. Miller, and P. Lewis, "Exploiting MODTRAN radiation transport for atmospheric correction: The FLAASH algorithm," in *Proceedings of the Fifth International Conference on Information Fusion*, **2**, July 2002.
3. A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," in *IEEE Transactions or Geoscience and Remote Sensing*, **26**, pp. 65–74, January 1988.
4. J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data," in *Fifth JPL Airborne Earth Science Workshop*, **1 of *JPL Publication 95-1***, pp. 23–26, 1995.
5. K. Lee, *A Subpixel Target Detection Algorithm for Hyperspectral Imagery.* PhD thesis, Rochester Institute of Technology, May 2004.
6. D. Manolakis, C. Siracusa, D. Marden, and G. Shaw, "Hyperspectral adaptive matched filter detectors: Practical performance comparison," in *Proceedings of SPIE*, **4381**, pp. 18–33, August 2001.
7. P. Bajorski and E. J. Ientilucci, "Geometric basis-vector selection methods and subpixel target detection as applied to hyperspectral imagery," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, September 2004.