

# **Feature Extraction Algorithms for Automated Detection in Hyperspectral Imagery**

Katherine Lee

Naval Research Laboratory

November 18, 2012

## **Abstract**

*Hyperspectral imaging offers the ability to classify objects in a scene based on their distinct spectral signatures. As a result, hyperspectral imaging technology has been widely used in a number of classification tasks spanning geology, biology, engineering and various other disciplines. The volume of data produced by hyperspectral imaging is far greater than that which can be interpreted visually; hence a significant amount of research is being devoted toward automated classification algorithms. This study focuses on the use of several such algorithms including the matched filter, Kernel Principal Component Analysis (KPCA), and sub-spectral identification for classifying different objects of interest in hyperspectral imagery. The effectiveness of each algorithm is quantified by the probability of correct classification. Although computationally expensive, it is found that the KPCA algorithm performs best on the imagery used in this study. On the other hand, sub-spectral identification is found to perform well when computational cost is considered, suggesting that only a few spectral bands are needed for the same (or better) detection.*

# Feature Extraction Algorithms for Automated Detection in Hyperspectral Imagery

Katherine Lee

## I. Introduction

Automated classification of objects in image data is receiving attention across a variety of disciplines. Hyperspectral imaging in particular has proven useful in biological, agricultural, and geological detection and classification applications. The development of improved imaging systems, digital image post-processing, and advanced algorithms can help better identify objects of interest. Hyperspectral imaging (HSI) has been shown to be successful for object detection in a maritime environment (Judd, 2012). These environments prove particularly difficult due to the lack of regularity in shapes for object detection (e.g., boat wakes) (Islam, 2009), and the environments themselves. In particular the maritime environment is challenging due to varying atmospheric conditions, glare from the water, marine haze, and clutter (deJong, 2009).

Advances in shortwave infrared (SWIR) sensor technology are helping to close the gap between traditional visible (and visible near-infrared [VNIR]) sensors and thermal band IR detectors. (Driggers, 1999). Shortwave infrared light has a longer wavelength than visible light which prevents it from being scattered as much by maritime haze. This allows systems that operate in the shortwave infrared band (0.9-1.7  $\mu\text{m}$ ) to discriminate objects more clearly in the presence of marine haze. Images obtained from SWIR sensors closely resemble those obtained from visible cameras making interpretation of a scene intuitive as opposed to the thermal band where interpretation maybe hindered by insufficient thermal contrast and angular resolution. This lends support for the use of SWIR technology in a HS sensor for maritime applications.

Spectral imaging captures spectral information by dispersing the light from a scene to create a spectrum of wavelengths for each pixel. We are interested in HSI, as applied to the shortwave infrared waveband, where the data are divided into approximately 50-350 spectral bands. Theoretically, using hyperspectral imaging over broad-band sensors would enhance object detection and discrimination by comparing spectral features of objects of interest within a scene to either a pre-defined dictionary or a trained set of features within a scene. For this presentation, hyperspectral data cubes acquired from a field exercise carried out at Pt. Loma, San Diego with a scanning shortwave infrared hyperspectral system were used.

This research seeks to find an effective means of feature detection in HS images. Several methods are considered for this including the matched filter (McDonough, 1995), and Kernel Principal Component Analysis (Hoffmann, 2006). The success of the algorithms will be determined through empirical testing with multiple hyperspectral images of representative scenes. These scenes will include various distractions including clutter and marine glare. The results of the tests will be evaluated with Receiver Operating Characteristic curves (ROC curves). In addition, a method of circumventing the Hughes Phenomenon, associated with high-dimensionality, is discussed.

## **II. Sensor information**

A shortwave infrared Hyperspectral (SWIR HS) sensor developed for the Naval Research Laboratory was used as a research tool for exploring the parameter space of an imaging spectrometer system for marine environments. The system is designed around a commercial 640 x 512 InGaAs detector sensitive to radiation in the 0.9-1.7 micron waveband, with a compact optical design consisting of a Dyson spectrometer and a commercial color corrected 50 mm F/1.4 telescope (Figure 1). The spatial data are aligned with the 640 pixel direction and the system has

an IFOV of 0.5 mrad. The spectral resolution is approximately 8.7 nm with 111 of the 512 bands illuminated. The data are stored in data cubes (widthx640x512. The width varies as the scan time changes.) A spectrometer mounted on a pan-tilt head is scanned and rotated through an angular sector to generate each data cube.

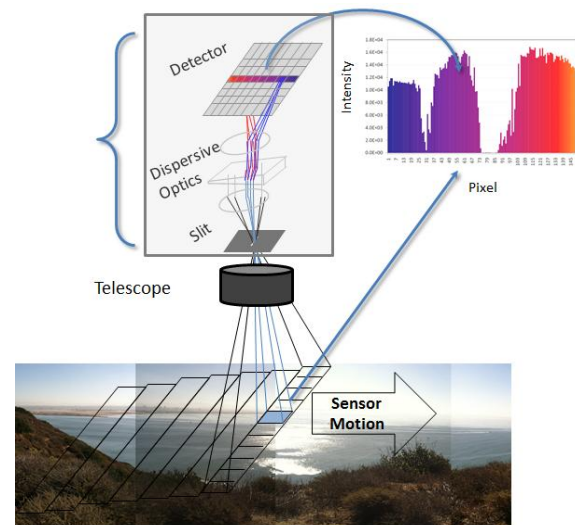


Figure 1: Hyperspectral scanning system.

### A. Field Data

As mentioned, the challenging nature of maritime sciences is used in this study to explore the robustness of the algorithms. Maritime scenes under different lighting and atmospheric conditions were collected at Point Loma, San Diego. The sensor was located along a cliff that provided a panoramic view consisting of Naval Air Station Coronado, the mouth of the harbor, San Diego proper and out to open ocean (see figure 1). The port is very active and provides a diverse array of recreational watercraft (primarily sailboats), charter vessels, and, as a major naval port, a plethora of military vessels.

### B. Spectral libraries

Utilizing collected data, a spectral library of normalized spectra taken from randomly selected sample imagery various image sources was formed. This enabled the later classification of spectra of interest by comparing against the known spectra in the library.

Spectra of interest in a maritime environment include boat wakes, sails, and concrete (figure 2). Due to the maritime environment and the SWIR range of the sensor, it was anticipated that all spectra under examination would have responses that modulated the solar irradiance curve (figure 3). The solar irradiance curve at sea level has three primary absorption bands

within the defined shortwave infrared region (0.9 to 1.7 microns) corresponding to  $H_2O$  and  $CO_2$ . All data show valleys at the wavelengths of the three absorption bands.

Wavebands from the center of the valleys in between the three absorption bands (~1578nm, ~1261nm, and ~1042nm) were taken to generate a false-color image. Outside these regions, the pictures become fuzzy because of a low signal-to-noise ratio. Four different types of vegetation are easily identified within this false-color image. The spectra pulled from representative pixels of the four different types of vegetation are shown in (figure 4).

The false-color image also shows that there are two distinguishable types of wake spectra. The first occurs closer to the propulsions systems and/or stern of a boat and has been named “new wake.” The second occurs farther away from the propulsion systems and/or stern of a boat and is labeled “old wake.” The agitated water relaxes as it ages and may contribute to the distinction between the two types of wake spectra (figure 5). In every trial, a matched filter algorithm was able to distinguish between new wake and old wake (figure 6)

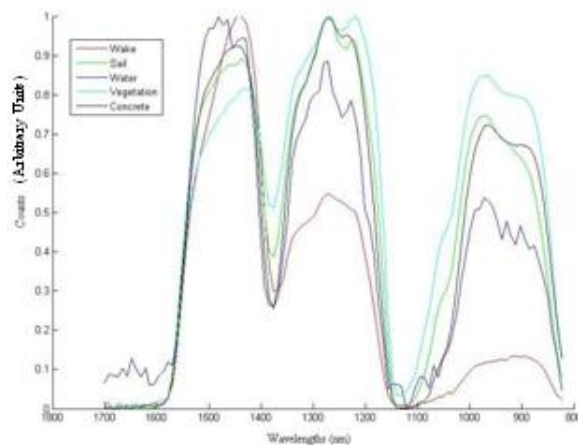


Figure 2: Normalized representative spectra from four wake pixels, four sail pixels, four water pixels, four vegetation pixels, and four concrete pixels.

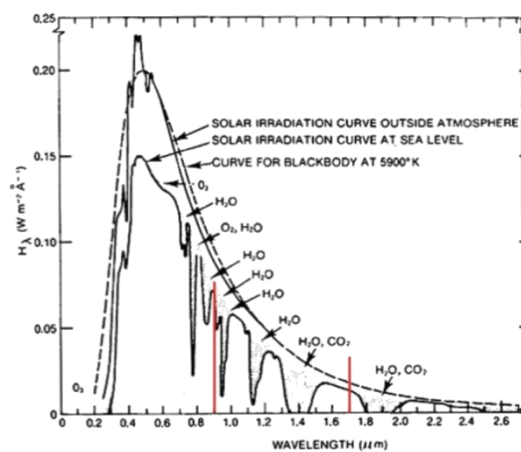


Figure 3: Solar reflectance curve at sea level. The range of the sensor used for data collection is noted in red. This image appeared in (Judd et. al., 2012).

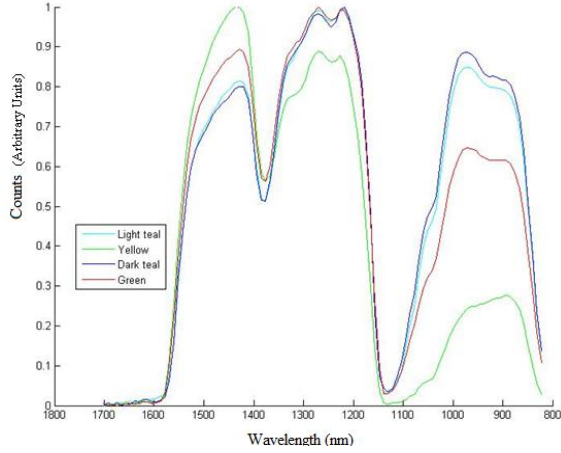


Figure 4: Normalized spectra from the four different types of vegetation present in an image.

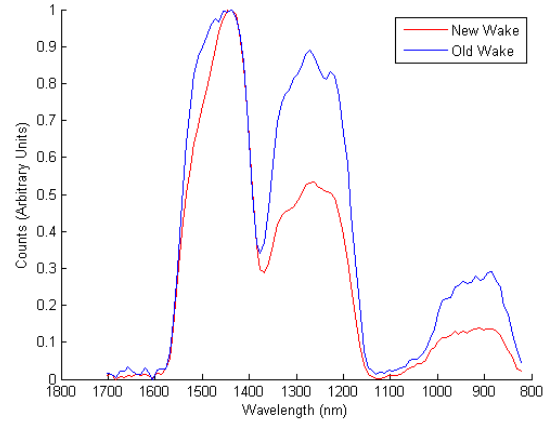
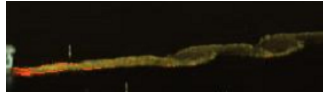
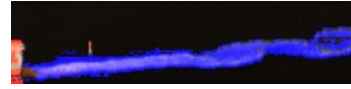


Figure 5: The two different types of wake spectra, new and old wake.



(a)



(b)

Figure 6: Here the new wake (a) is detected by cross-correlating the spectra of pixels in this image with new-wake spectra from another image. The detected pixels are highlighted in red. In (b), old-wake pixels are detected by cross-correlating the spectra of pixels in this image with old-wake spectra from another image. Pixels identified as old wake are highlighted in blue. The ship is highlighted in red as another set of spectra (for sails) was also cross correlated with this image and everything highlighted by red was identified as a ship.

### III. Methods

#### A. Normalization

Before any analysis was done with the spectra, all spectra were normalized with respect to themselves. While it is generally not necessary to normalize spectra when comparing an object's spectrum to another in the same image, normalization is required to reduce the effect of lighting and atmospheric variations. Normalization also allowed for better categorization of images. Data collected in the field study did not have accompanying calibration results, therefore normalizing the spectra with calibrated spectra was not an option in this study. Normalization was performed with the following formula:

$$S' = \frac{S - \min(S)}{\max(S) - \min(S)} \quad (1)$$

Where  $S$  is the spectrum,  $\min(S)$  and  $\max(S)$  are the minimum and maximum values of  $S$ , respectively, and  $S'$  is the normalized spectrum.

### *B. Matched filter*

Initial analysis of the data was done using matched filters. In an automated detection mechanism, a known spectrum is correlated with a new spectrum and compared with a predefined threshold level to determine the similarity of the new spectra with the known spectra. To accomplish this, the spectra are vectorized and the cross-correlation (inner product) between a spectrum (or vector) of interest and a known spectrum (or vector) is taken. The formula for the inner product is defined as:

$$R = \int_{-\infty}^{\infty} f(t) \cdot q(t) dt \quad (2)$$

Where  $f(t)$  is the function of interest and  $q(t)$  is the function  $f(t)$  is being correlated against. This can be approximated as a summation, as in this study, by the following formula as there is no explicit function that models the individual spectra.

$$R = \frac{1}{(N) \cdot \sigma_X \sigma_Y} \sum_{t=1}^N X_t \cdot Y_t \quad (3)$$

Where  $X_t$  is the  $t$ -th value of  $X$ ,  $Y_t$  is the  $t$ -th value of  $Y$ ,  $\sigma_X$  is the standard deviation of  $X$ ,  $\sigma_Y$  is the standard deviation of  $Y$ , and  $N$  is the number of terms in  $X$  or  $Y$ . The factor in front of the summation normalizes the two spectra to yield a correlation coefficient that ranges between 0 and 1 where 1 signifies that the two spectra are exactly correlated.

Taking the cross-correlation of a sample spectrum from a field image with known spectral signatures yields a correlation coefficient that can be compared against a predefined threshold level to determine whether or not the given spectrum is a good fit for the known spectra samples. An example of a detection done on wake pixels is shown in figure 7.



Spectra can also be cross correlated as a matrix as opposed to correlating each spectrum individually. This can be accomplished by building a matrix of spectrum values and performing matrix multiplication. This process is described in detail in (Judd, 2012) A benefit to using a matched filter method in initial stages is the speed of analysis as the computational intensity is low (multiplication and addition). This is also beneficial for working in the field for real-time analysis. The ability of the matched filter to operate on a matrix of spectra further enhances the speed benefits.

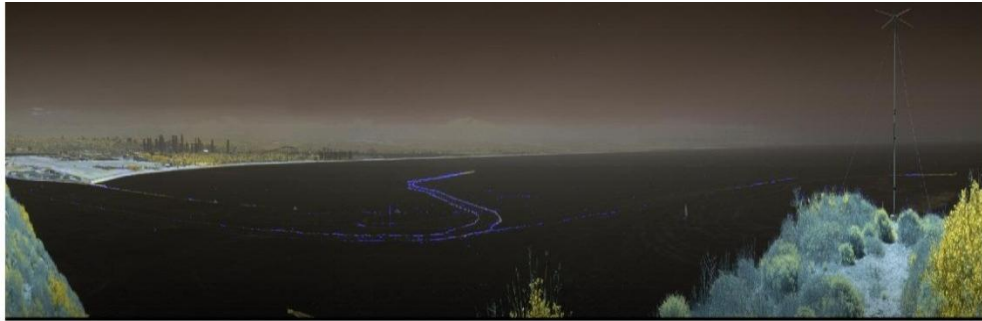


Figure 7: Cross-Correlation of a wake pixel from this image with every other pixel from the current image found the wake (colored in blue).

### *C. Kernel PCA*

Another method of classifying spectra is through Kernel Principal Component Analysis (KPCA). In this paper, spectra of a specific object (labeled as  $\mathbf{x}$ ) were pulled from the spectral library to create a PCA model (or a basis of eigenvectors) for analysis. The spectrum of interest (labeled as  $\mathbf{z}$ ) was then projected onto the model (eigenvectors) to compare the spectrum of interest with the model. This projection was done in a high-dimensional feature space that was accessed through the kernel trick. Performing PCA in the high-dimensional space is useful because linear decisions in the high-dimensional space yield non-linear (and more accurate) decision boundaries in the original space.

KPCA utilizes the kernel operator which transforms the data vectors into a high-dimensional “feature space”. Once in the feature space, a standard PCA is performed which yields a non-linear decision boundary in the low-dimensional space. Here, we use a Gaussian kernel which is parameterized by,  $\sigma$  (standard deviation of the Gaussian kernel) and  $q$  (number of dimensions the data is reduced to). If  $\sigma$  is too large, the KPCA becomes a regular singular value decomposition algorithm. If  $\sigma$  is too small, then the algorithm becomes trivial. The formula for finding the reconstruction error, or  $p(z)$ , the distance between the vector in the feature space and the projection onto the eigenvectors (as shown in figure 8 as given by (Hoffman, 2006) is as follows:

$$p(z) = \left[ k(\mathbf{z}, \mathbf{z}) - \frac{2}{n} \sum_{i=0}^n k(\mathbf{z}, \mathbf{x}_i) + \frac{1}{n^2} \sum_{i,j=1}^n k(\mathbf{x}_i, \mathbf{x}_j) \right] - \sum_{l=1}^q \left[ k(\mathbf{z}, \mathbf{x}_i) - \frac{1}{n} \sum_{r=1}^n k(\mathbf{x}_i, \mathbf{x}_r) - \frac{1}{n} \sum_{r=1}^n k(\mathbf{z}, \mathbf{x}_r) + \frac{1}{n^2} \sum_{r,s=1}^n k(\mathbf{x}_r, \mathbf{x}_s) \right]^2 \quad (4)$$

Where the first half of the equation is the vector in feature space as represented by the kernel, and the second half is the projection onto the eigenvectors,  $V^l = \sum_{i=1}^n \alpha_i \tilde{\Phi}(\mathbf{x}_i)$ .  $n$  is the dimensionality of the data,  $k(\mathbf{x}, \mathbf{x})$  is the gaussian kernel,  $\mathbf{z}$  is spectrum of interest,  $\mathbf{x}$  is a

spectrum, and  $\tilde{\Phi}(x)$  is the distance between the spectra in the feature space and the center of distribution in the feature space.

The KPCA analysis was highly successful in correctly categorizing sail pixels (figure 9), but was significantly more computationally intensive than the matched filter model because projections had to be calculated for each pixel's spectrum in a region of interest.

#### *D. Sub-spectral*

High-dimensional data like that obtained in HSI is plagued by the curse of dimensionality, or the Hughes Phenomenon

(Alonso, 2011). Algorithms like KPCA circumvent the Hughes Phenomenon, but sacrifice speed and computational simplicity in the process. To limit the effect of the Hughes Phenomenon and maintain a fast processing time, we attempt to use the matched filter algorithm to preprocess spectra. Previously, both linear PCA and non-linear KPCA were used by (Rodarmel, 2002) and (Fauvel, 2006) to preprocess spectra for feature detection. In this application, however, we attempt to decrease computing time and the rate of false positives by using the matched filter instead.

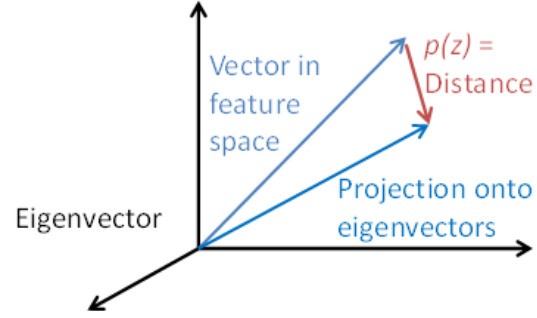


Figure 8: Graphical representation of formula (4)

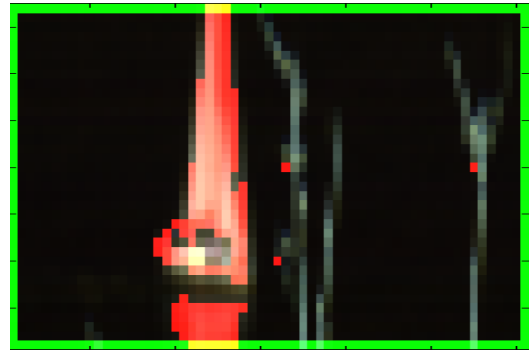


Figure 9: Region interrogated with the KPCA model created using the spectra from randomly selected images of sailboats. The model was able to properly categorize the sailboat along with the reflection with minimal false positives.

Due to the shortwave infrared range of the sensor (spectral response range of 0.9 to 1.7  $\mu\text{m}$ ) and, three regions were hypothesized to contain the most useful information about a scene. As previously mentioned, the region of interest, 0.9 to 1.7  $\mu\text{m}$ , contains three primary absorption bands corresponding to  $\text{H}_2\text{O}$  and  $\text{CO}_2$  (figure 3). The spectra show three distinct lobes where the most irradiance is transmitted in between the absorption bands.

To test this hypothesis, each wavelength band (where  $m$  is the number of bands) of  $n$  spectra of the same classification (e.g. all boat spectra, all wake spectra) were cross-correlated against each other with formula (3) to obtain a correlation coefficient for each wavelength band in the spectra. This yields a  $m \times (n!)$  matrix containing the correlation coefficients for each band with every other corresponding band in the other spectra.

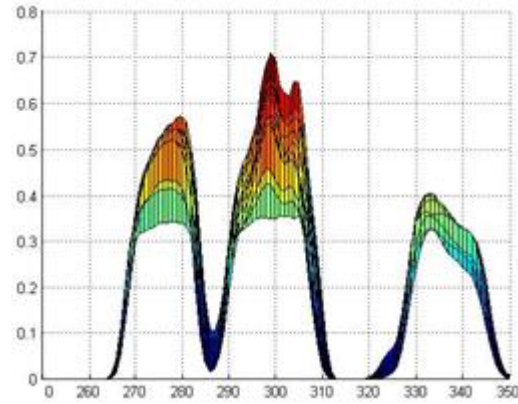


Figure 10: Cross-correlation coefficients for each waveband (indexed at 250) of a collection of sail spectra. The three peaks correspond to the three areas in absorption bands.

The top three most highly correlated regions (figure 10) were chosen as regions of interest for the original normalized cross-correlation method. These identified regions did fall between the absorption bands of  $\text{H}_2\text{O}$  and  $\text{CO}_2$  as expected. This allowed for better classification of the spectra of interest by only cross-correlating the sections of the spectra that are common to all other spectra of the same classification.

After preprocessing the spectra, a matched filter was again run on the selected bands of the spectra to find a new cross-correlation coefficient (figure 11).

## IV. Results

### A. Numerical ROC Curves

The ROC curves shown in figure 14 were generated numerically by first tabulating the cross-correlation coefficients or  $p(\mathbf{z})$  (the distance for KPCA) and sorting them into two lists. One list contains all the cross-correlation coefficients associated with known target pixels (figure 12) (referred to as “truth” coefficients), and the other list contains all the cross-correlation coefficients for the “non-truth” pixels.

The probability of detection for a given threshold is the fraction of “truth” coefficients that exceed threshold, while the probability of false alarm is estimated as the fraction of non-truth coefficient exceedances. The

problem with a numerical approach is that one is limited to discrete approximations to the probabilities with a resolution  $\frac{1}{M}$ , where  $M$  is the number of feature values observed.

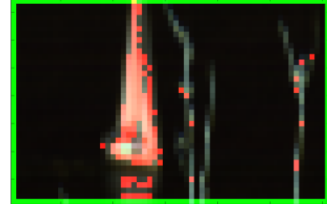


Figure 11: Sub-spectral analysis: Matched filter algorithm performed on the preprocessed spectral bands. This detection was done by using three bands in each of the three regions

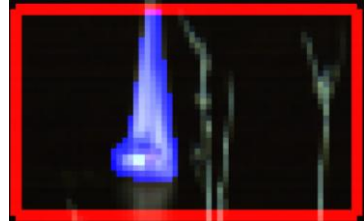


Figure 12: Pixels defined as truth for the sailboat.

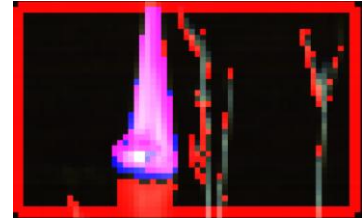


Figure 13: Red pixels were identified as sail pixels by the matched filter. Pixels in blue are those defined as truth as in figure 12.

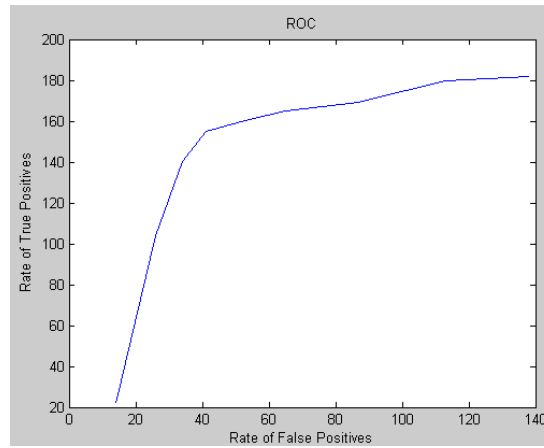


Figure 14: ROC curve detected by matched filter.

### B. Analytic ROC Curves

As an alternative, the “truth” and “non-truth” coefficient distributions may be fitted to a particular functional form and the needed fractions may be determined analytically by integration (i.e., generating the cumulative distribution function). In this work, the distribution of both the “truth” coefficients and the “non-truth” coefficients were found to be well approximated by the beta-distribution. Thus a rate of false alarms can be specified and

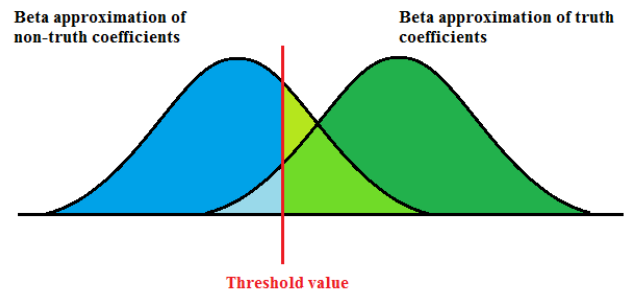


Figure 15: Here, two beta distributions are pictured. The beta on the left approximates the distribution of non-truth coefficients, and the beta on the right approximates the distribution of truth coefficients. The area colored in light and medium green represents the probability of false alarms, and the area colored in dark and medium green represents the probability of detection. The area colored in light blue represents the probability of falsely labeling a pixel as non-target, and the area in dark blue represents the probability of correctly labeling a pixel as a non-target.

the inverse of the beta cumulative distribution function (approximating the non-truth coefficients) can be found to yield the corresponding threshold level. This same threshold level can then be used to provide the area under the beta distribution associated with the truth coefficients, thus yielding a detection probability (figure 15).

The benefit of using an analytic ROC curve is to be able to explicitly specify a false alarm rate and then be able to find a threshold value and a corresponding rate of detection through integration. This allows us to compare algorithms at a specified percent of false alarms as opposed to being

restricted to the probabilities  $x_{\frac{i}{M}}^i, i = 1 \dots M$ .

Two analytic ROC curves are shown in

figure 16 that were showcase the differences between using three and six bands for detection in KPCA. The curve corresponding to using three total bands is very close to a random detection line where the rate of false alarms equals the detection probability (a diagonal line from (0, 0) to (1, 1)).

### C. KPCA optimization

The optimization of the two variables  $\sigma$  and  $q$  can only be done through computational means. The optimization of  $q$  is not shown here as  $q$  was taken to equal  $n$  for this study. However, the method of optimization for  $q$  is the same as the method for  $\sigma$  which is as follows.

Representative scenes were selected

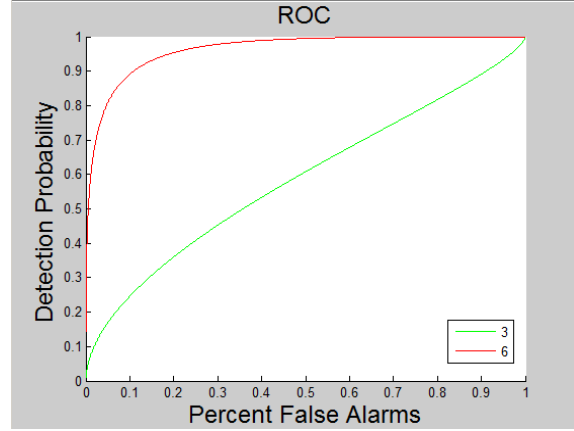


Figure 16: Two examples of analytic ROC curves. These curves were generated by using the matched filter to preprocess the spectra and then running the KPCA method of analysis. The numbers stand for the number of bands (total) used for the KPCA.

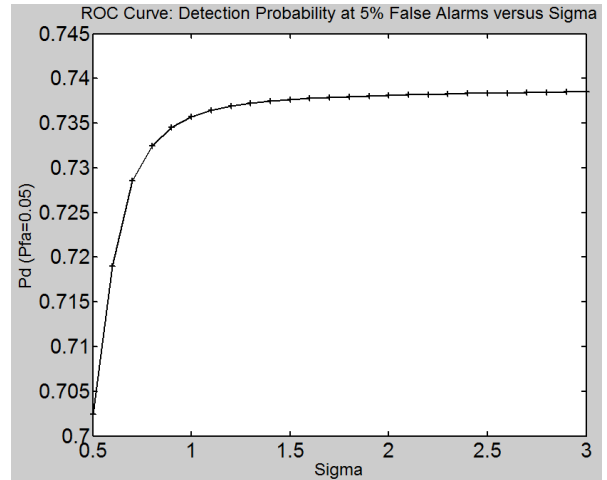


Figure 17: Detection probability at a 5% false alarm rate with respect to the value of the standard deviation of the Gaussian kernel (sigma).

to be cross-correlated with training spectra with first varying  $\sigma$ . Analytic ROC curves were then generated and the detection probability at a 5% false alarm rate. Figure 17 shows that the detection probability levels off near  $\sigma = 1.3$ . This signifies that the standard deviation for the Gaussian kernel need only be 1.3 to achieve an optimum detection rate.

### *Band number optimization*

The method of preprocessing begs the question of the optimal number of bands from each of the three most important regions. A similar method of optimization is used to determine the optimal number of bands as was done with  $\sigma$  for KPCA. The percentage of false alarms was again taken to be 5%. The KPCA was run with 3 to 29 bands with an even number of bands in each of the three highly correlated regions (figure 18). This suggests that it is unnecessary to use more than three bands in each of the three highly-correlated regions.

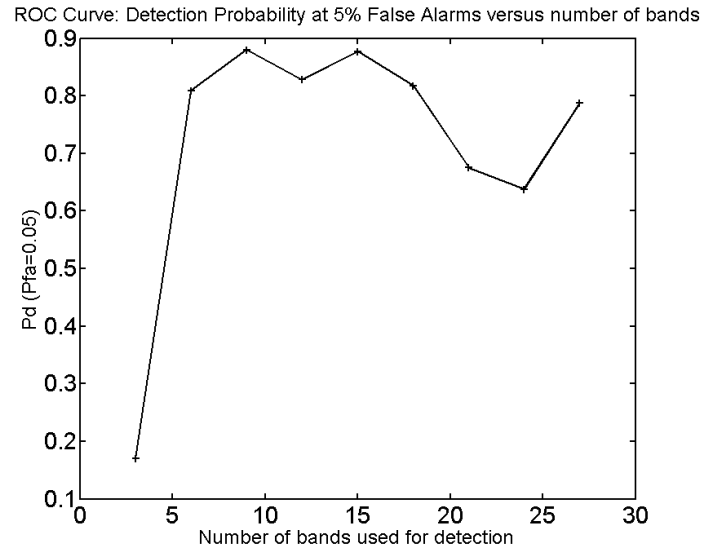


Figure 18: Detection probability at a 5% false alarm rate with respect to the number of bands used for analysis.



#### D. Algorithm Comparison

The KPCA method (figure 9) of characterization out-performed both the matched filter and the sub-spectral analysis (figure 11) at minimizing false positives while maximizing true positive in the interrogation region. The matched filter though, takes over the sub-spectral

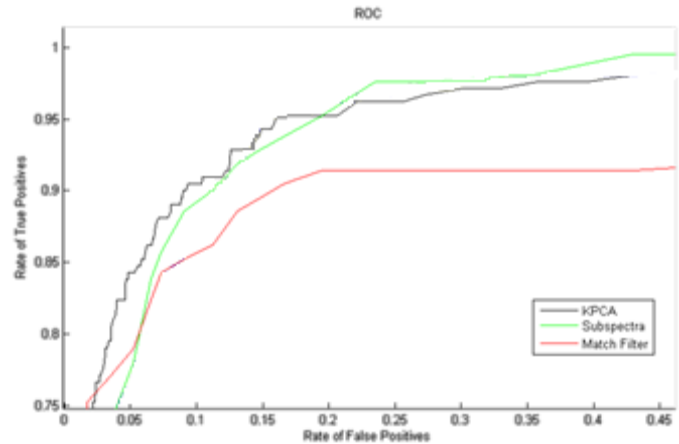


Figure 19: ROC curves of KPCA, sub-spectra, and matched filter analysis, zoomed in to show the differences.

analysis and the KPCA when we reach a less than 3% rate of false alarms. Otherwise, sub-spectral analysis outperforms the matched filter and is on par with the KPCA. A note taken is that figure 19 is zoomed into display just a small section of the entire ROC curve. For many practical applications, a false alarm rate of 5% is sufficient. At that rate of false alarm, the KPCA analysis performs best with sub-spectral slightly behind. Sub-spectral analysis, however, has the added benefit of speed and less computational complexity thus making it easier to implement.

#### V. Further Research

Further research in this area could create more complex models of the spectra to better understand and classify spectra of interest. The process of selecting spectra to train on is difficult and somewhat convoluted as adjacent pixels often gather information from an overlapping area. This is known as the adjacency problem. Selecting a pixel that is adjacent to one with a better match can sometimes cause the matched filter algorithm to not detect similar spectra in the interrogation region of a different picture (assuming the same threshold value). Improvements in the hardware of the sensor could address this problem to a certain degree, but refining spectral signatures in post-processing could also be useful.

Additionally, finer spectral resolution HSI sensor may improve detection. This could help distinguish further between the different spectra by characterizing finer spectral features that may be used to better classify spectra of interest. However, a finer spectral resolution HSI sensor may also face a signal-to-noise reduction as the light would be dispersed too much to have adequate levels of radiation reaching the cells of the sensor. Additionally, it would be useful to create a sensor where the user would be able to specific bands to retain information for, thus substantially reducing size of current data files and decreasing computation time. Additional field tests could involve calibrations with substances with known spectral signatures to provide additional atmospheric information to calibrate and normalize the extracted spectral signatures.

## **VI. Conclusions**

Matched filters, sub-spectral identification, and KPCA methods of classifying spectral signatures all worked well in distinguishing targets of interest from background noise (be it water or clutter). Hyperspectral imaging is very promising in terms of detecting and classifying objects within a scene. Additionally, only small numbers of bands contain useful differentiating information which helps to reduce complexity.

## **VII. Acknowledgements**

The author would like to thank Drs. K. Peter Judd (Naval Research Laboratory (NRL)), Colin Olson (NRL), Jon Nichols (NRL), Jim Waterman (NRL), John Dell (Thomas Jefferson High School for Science and Technology), and Mr. Max Conant (NRL) for their help and support in working on this project, in addition to the SEAP program and the Naval Research Laboratory for hosting her this summer. The SEAP program is supported by the Office of Naval Research.

## References

- [1] Alonso, M. C.; Malpica, J. A.; de Agirre, A. M. (2011) Consequences of the Hughes Phenomenon on some Classification Techniques. *ASPRS 2011 Annual Conference*
- [2] de Jong, A. N.; Schwering, P. B. W.; Fritz, P. J.; Gunter, W. H. (2009) Optical characteristics of small surface targets, measured in the False Bay, South Africa, June 2007. *Proc. SPIE, 7300*.
- [3] Driggers, R.; Cox, P.; Edwards, T. (1999). Introduction to Infrared and Electro optical Systems. Artech House.
- [4] Fauvel, M., Chanussot, J., Benediktsson, J. A. (2006). Kernel Principal Component Analysis for Feature Reduction Hyperspectral Images Analysis. *Signal Processing Symposium, 2006. NORSIG 2006. Proceedings of the 7th Nordic*
- [5] Hoffmann, H. (2006). Kernel PCA for novelty detection. *Pattern Recognition*, 40.
- [6] Islam, M. M.; Islam, M. N.; Asari, V.; Karim, M. A. (2009). Anomaly Based Vessel Detection in Visible and Infrared Images. *Image Processing: Machine Vision Applications II*, 7251.
- [7] Judd, K. P.; Nichols, J. M.; Howard, J. G.; Waterman, J. R.; Vilardebo, K. M. (2012). Passive shortwave infrared broadband and hyperspectral imaging in a maritime environment. *Optical Engineering*, 51(1).
- [8] Mahesh, P. M. P. (2004). Assessment of the effectiveness of support vector machines for hyperspectral data. *Future Generation Computer Systems*, 20.
- [9] McDonough, R. N.; Whalen A. D. (1995) Detection of Signals in Noise, Academic Press, San Diego.

- [10] Rodarmel, C; Shan, J. (2002). Principal Component Analysis for Hyperspectral Image Classification. *Surveying and Land Information Systems*, 62.
- [11] Rumpf, T. et al. (2010). Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Computers and Electronics in Agriculture*, 74.