New Improvements in the ORASIS Algorithm[a,b]

Jeffrey Bowles[c], David Gillis,[d,e] and Peter Palmadesso[d]
Naval Research laboratory
4555 Overlook Avenue, SW
Washington, DC 20375
202-404-1021
Jeffrey.Bowles@nrl.navy.mil

Abstract—We describe recent improvements to the Optical Real-time Adaptive Spectral Identification System (ORASIS). The NEMO satellite will use ORASIS as a compression algorithm and first pass analysis algorithm. One improvement concerns the quantization aspect of ORASIS. In this step, the hyperspectral data are put through a process that creates a set of spectra (called exemplars) that represent the entire data set within the specified error tolerance. Each exemplar has a high dimensional "sphere" of influence. Any spectra within that sphere are replaced by a reference to that exemplar. However, the spheres tend to overlap. Results are improved by selection of the optimum exemplar. The second improvement is a method that relates sensor noise characteristics to spectral angle. The noise parameters of the sensor can be used to calculate an error tolerance, for each spectrum, dependent upon spectral shape and intensity. This method provides a physical justification for the error tolerance.

TABLE OF CONTENTS

1. INTRODUCTION

Remote sensing methods employing hyperspectral-imaging promise great advances in the ability to find, discriminate, and monitor the condition of materials. The use of hyperspectral sensors is becoming more wide spread as the ability of the measurement systems increases and costs decrease. Already, aircraft or satellite-based hyperspectral systems have been used to find minerals,[1] monitor forests,[2] and measure ocean properties.[3,4] Additionally, new applications and new sensor systems are being developed constantly.

Hyperspectral systems generally measure between 20 and 400 spectral bands. Spatial resolution ranges from less than a millimeter, in microscope systems to greater than 100 meters. Data produced by these systems is complex, voluminous, and full of detailed information. The ability to process and extract relevant information from the data in a timely manner is paramount to the continued development of the field.

A number of algorithms exist for analysis[5] and compression[6] of hyperspectral data. ORASIS[7] has been under development at the Naval Research Laboratory (NRL) for over 5 years. It was originally developed as an analysis algorithm[8] but was discovered to be an excellent compression algorithm[9] as well. The goal of ORASIS development has always been to create a real-time system that is capable of tactical or other time critical use. Thus, ORASIS is usually discussed in this light. It is believed that in non-time-constrained situations, the results from ORASIS could be improved by subtlety changing how it runs.

The ORASIS algorithm is actually a series of independent algorithms, which are run sequentially under most circumstances. After some preprocessing, which includes flat fielding and various tests on the spectra, the exemplar selection process (ESP) is run. The goal of the ESP is to screen the data for the unique spectra, creating a representative set of spectra. This set is called the exemplar set. The process is based on spectral angle. An angle is calculated by considering the spectra to be vectors with dimensionality equal to the number of measured bands. The "spectral" angle between two spectra is the arccosine of their normalized dot product. The primary parameters that need to be set when running ORASIS are the settings that determine a spectral error angle. If two spectra are within this error angle, then the differences between them are ignored for the purposes of running the rest of the algorithm. The smaller the error angle the more computationally intensive ORASIS becomes and the more exemplars that are found.

Following the ESP, an algorithm to determine the dimensionality and endmembers of the exemplar set is run. Finally, the full data set, or sometimes just the exemplar set, is projected into the space defined by the endmembers.

The rest of this paper proceeds as follows: The next section deals with improvements made in the quantization aspect of the ORASIS algorithm. Improvements in the reconstructed data cubes are investigated statistically. Next, the relationship between the spectral error angle and the sensor noise parameters is investigated. The noise vs signal

[a] Work funded by the Office of Naval Research

[c] Remote Sensing Division, Code 7212
[d] Plasma Physics Division, Code 6700.3
[e] ASEE Postdoctoral Fellow

relationship is investigated for a particular sensor. Methods of using this calculated error angle to change ORASIS output are investigated. Finally, a brief discussion is presented.

2. IMPROVEMENT IN QUANTIZATION

The goal of the prescreener is to reduce the volume of data with minimal loss of important information. The reduced data set is then passed on to the next algorithm that determines the endmembers in the data set. In most cases, this algorithm is too time consuming to run on the entire data set. By reducing the data volume in the prescreener, the approximate endmembers may be found quickly. These endmembers are only slightly different from what would be found using the entire data set.

The prescreener builds up the reduced data set from scratch. Normally, the first spectrum in the scene is a member of the reduced data set. The members of the set are called exemplars. Each subsequent spectrum is tested to see if it is within the specified spectral error angle of any of the exemplars in the reduced data set. As soon as a match with an exemplar is found, the search stops. If no exemplars are found to be within the error angle, the spectrum being tested becomes an exemplar.

It is proper to think of the prescreener as a single pass Learned Vector Quantization (LVQ) process. The exemplars are representative of the entire data set. All spectra in the original data are within the error angle of at least one exemplar. This is the quantization aspect of the prescreener. Because every spectrum is represented to within the same set error angle, the claim can be made that ORASIS is not statistical in nature. A unique spectrum from the scene has the same limits of representation as the most common spectra in the scene. This one feature differentiates ORASIS from many of the statistical methods being used.

The exemplar set is thus created. The nature of the exemplars and their relationship to the data are indicative of the complexity of hyperspectral data. Exemplars have a high dimensional "sphere of influence." That is, a volume in which any present spectrum would have an angle, with regard to the exemplar, of less than the error angle. Consider the case where a spectrum from the scene is just outside the exemplar's sphere. That spectrum becomes an exemplar. There now exists a region of overlap between the exemplars (see figure 1.) In this region, a spectrum may be within the error angle of both exemplars. This turns out to be very common. In fact, as the error angle decreases, the likelihood that there is overlap increases. In some situations, the average number of overlaps has exceeded ten. While it may be difficult to envision this large a number of overlaps in low dimension, remember that the data lives in a very high dimensional space.

The overlap does not impact the next step of the algorithm, the endmember determination. The overlap has no effect on which spectra become exemplars. The exemplars depend only upon the order in which the spectra are processed. However, for ORASIS as compression tool, this overlap influences the quality of the compressed data. As the incoming spectra are compared to the exemplars, the order in which the exemplars are compared to the spectra also becomes important. The first exemplar found to match a

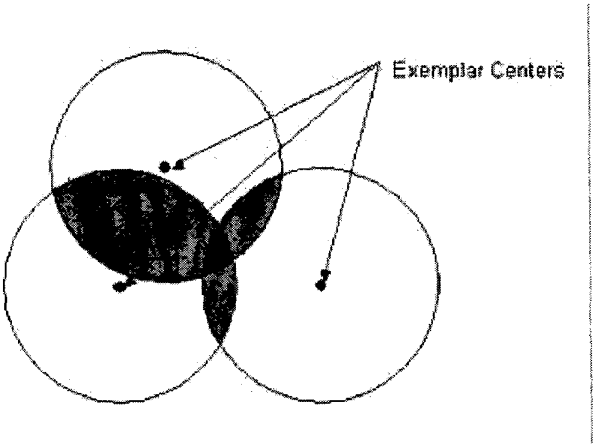

Figure 1: An example in 3 dimensions of exemplar overlap. The black, solid circles represent the center of the exemplars. The medium (red) shaded regions are areas that can be represented by two exemplars. The dark shaded (black) area has 3 exemplars overlapping. Although multiple exemplars can represent a spectrum in the overlapped areas one of the exemplars is the best match.

spectrum might not be the best fit in the set. Thus, more error than is necessary might be present if ORASIS stops after finding the first exemplar within the error angle for each spectrum in the scene. When ORASIS stops after finding the first fit within the error angle we refer to the process as first fit (FF).

If ORASIS continues to process even after a first fit is found, it is sometimes possible to find another exemplar that is closer to the spectrum under test. If an exhaustive search was done we would call this process the Best Fit (BF.) However, as stated above, as many as 10 exemplars on average may be positioned within the error criterion. In this paper, we provide a study to determine the benefits of using best fit. There will clearly be a time penalty for using any implementation of best fit, but we believe that in the end we may be able to develop a method that has only a modest time penalty. The exact details of the final implementation of BF are yet to be determined but will be published later.

In order to test this new aspect of ORASIS, AVIRIS data were used. Data sets from Cuprite NV and the Florida Keys were compressed using ORASIS. The results of reconstructing the data are shown here using three methods. First, figure 2 shows a randomly chosen spectrum from the original and reconstructed Cuprite scene. The total compression in this case was 31:1. Although it is hard to tell, the original spectra and the two reconstructed sections lie almost right on top of each other. The second test method shows the decrease in RMS error between the original and reconstructed data as a function of compression ratio. The compression ratio is controlled by changing the error angle setting. The third test is shown in figure 4 where reconstructed RGB images of the Florida Keys scene are shown. In some cases, the reconstructed single bands of the images can show streaking. The streaking is caused by using the same exemplar to represent many nearby spectra. This streaking is very small and is probably insignificant for most applications. However, the human eye is very good at detecting the streaking. Note the decrease in streaking in the figure when using BF as compared to FF.

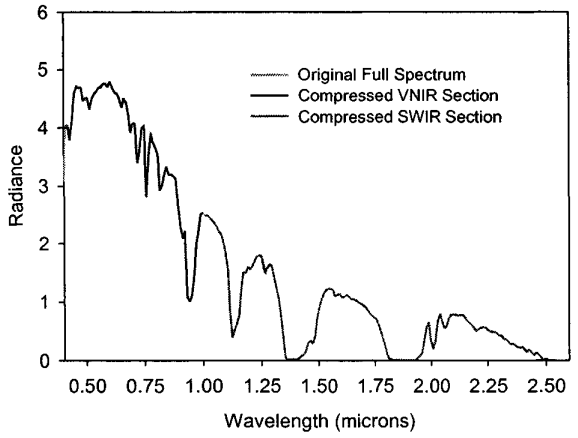

Figure 2: This figure shows a randomly chosen spectrum from the AVIRIS Cuprite data set. Overlain, is the reconstructed spectrum from the ORASIS compressed data. In this example, the VNIR and SWIR sections were compressed independently.

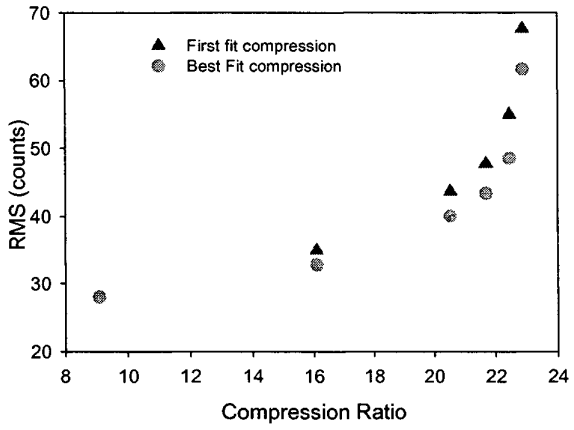

Figure 3: This figure shows the improvement in RMS error between the processing with FF and processing with an implementation of BF.

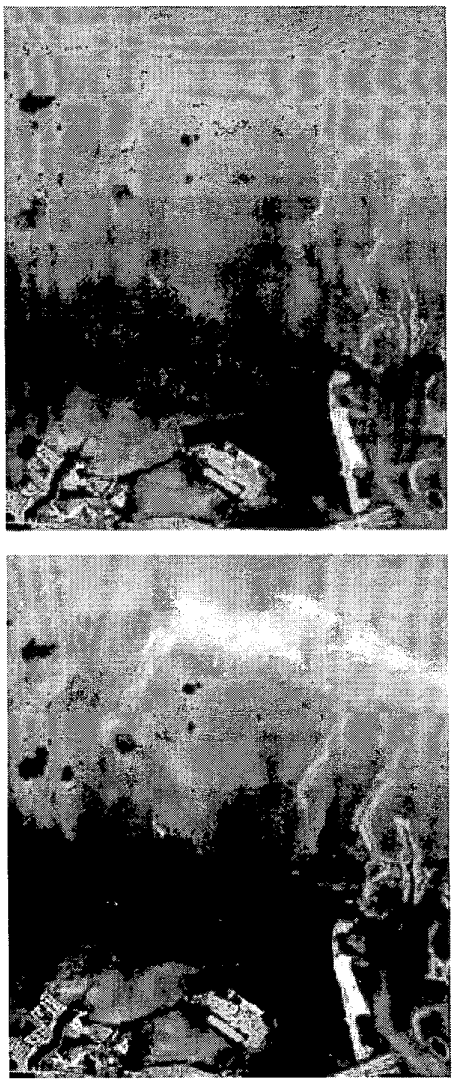

Figure 4: This figure shows reconstructed RGBs of the Florida Keys data. The top picture was created using FF and shows significant streaking. The bottom picture is from data that was compressed with the exact same settings except that it used BF. Only minimal streaking is apparent

3. Relating Sensor Noise to ORASIS Parameters

Up until recently the error angle was set by trial and error. If too many exemplars were found, the error angle was increased and vise versa. The effect of the error angle setting on analysis is complicated and somewhat difficult to explain in this short paper, however, the effect on compression is significant and much more straightforward. In compression, it is desirable to retain as much information as possible. It is intuitive that data from sensors that have an overall high signal to noise ratio (SNR) have a higher information content than data from low SNR sensors. Thus, it would seem likely that it should be possible to relate the sensor noise characteristics to the error angle in a physically meaningful way.

It would be ideal to be able to set the error angle at a value just below the noise, retaining as much information as possible while ignoring the differences that are truly due to noise. The ability to do this would have two benefits. First and most obvious, the error angle would become less arbitrary – switching between sensors would not be as difficult. One could tell where the sampling was occurring relative to the noise without any trial and error. The second benefit, which is subtler, is that setting the error angle separately for each spectrum can allow control over the sampling process. As is discussed below, it is possible to use the details of the spectrum to determine how the spectrum, and others like it, is sampled.

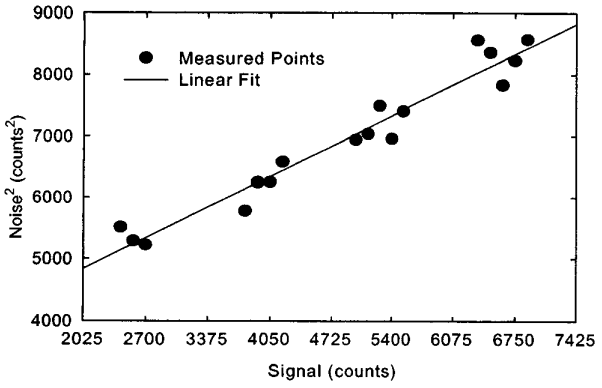

Figure 5: This graph demonstrates the relationship between the signal level and the level of the square of the noise. The relationship appears linear as expected from shot noise theory

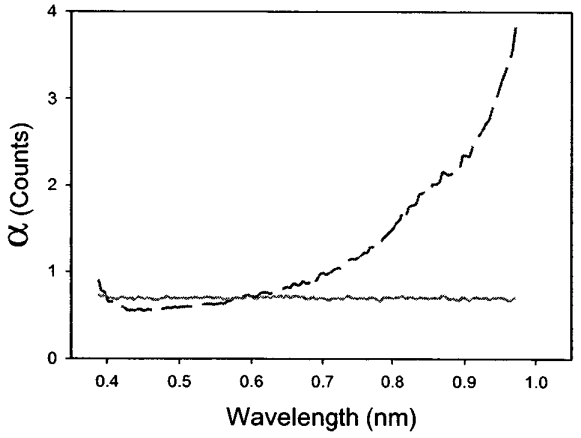

Figure 6: This figure shows how the slope of the line (shown in figure 5) varies with wavelength. When the calculation is done with the data in counts the slope is constant and does not change with wavelength. When calculated after the data is in radiance the slope is a strong function of wavelength and may be related to the efficiency of the grating.

To test this idea for coupling sensor noise characteristics to ORASIS parameters we used data from NRL's Ocean Portable Hyperspectral Imagers for Low Light Spectroscopy (Ocean PHILLS) system.[10] All PHILLS spectrometers are of the pushbroom variety. In operation, the foreoptic images a line of the scene onto a slit. The image of the slit is dispersed by a grating and focused onto a two-dimensional CCD array. A line, or frame, of data is taken every integration time. The data consists of spectral information in one direction and spatial informati on, along the slit, in the other direction on the array.

Figure 5 shows the relationship between the signal strength and the noise of data taken with our Ocean PHILLS system. To make this figure 1024 frames of data were taken with the system imaging an integrating sphere at NRL's Calibration Facility.[11] The frames were averaged together to create the signal level. The standard deviation is interpreted as the noise level. Note that the plot is Noise^2 on the y-axis and signal level on the x-axis. The line intercepts the positive y-axis giving a measure of the fixed, or dark, noise level. The fact that the line is linear on this plot is indicative of shot noise.

The fit to the plotted line allows the noise level to be estimated for any signal level. The slopes can be calculated for each wavelength channel. An error angle is calculated based on the slope and the fixed noise level. Figure 6 shows the slope, of the line, as a function of wavelength for two cases. The first case is when the calibration data is used in its raw, uncalibrated form. The dashed line is found when the data has been radiometrically calibrated. The shape of the second line is not well understood. It is not a direct result of the change in quantum efficiency. However, it does appear to be qualitatively similar in shape of the inverse of the grating efficiency. However, we cannot confirm that the grating efficiency is responsible for this shape at this time. Further work is necessary.

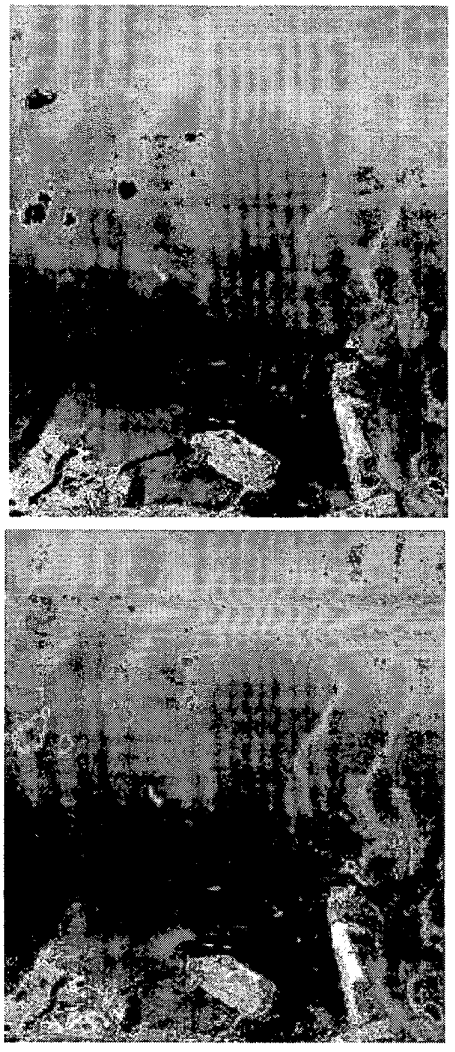

Figure 7: This figure shows the results of using the noise model input to control sampling. The yellow dots show the location of exemplars. The top figure shows an ORASIS run using the noise model. The bottom picture shows the results of using the noise model to tailor the sampling preferentially to dark pixels.

Using the information from the noise model a calculated error angle for each spectrum replaces the fixed error angle

in the ORASIS ESP. Use of the noise model allows ORASIS to compress the data based on the noise level in each individual spectrum. In practice, ORASIS uses a fixed multiple of the calculated error angle. The multiple controls the sampling and affects the fidelity of the reconstructed data.

It is typically true that darker spectra have larger calculated error angles, resulting from a lower signal to noise ratio. Conversely, the brighter spectra have smaller calculated error angles. If a single error angle is used for all spectra, as the error angle is reduced ORASIS will primarily add dark spectra to the exemplar set. In this case, ORASIS is sampling the noise surrounding the darker pixels.

If the calculated error angle is used, and the multiple of the error angle setting is reduced, then ORASIS samples equally into the noise of both the darker and brighter spectra. Additions to the exemplar set are based on whether the spectra are different from the exemplars within some level of noise (physically meaningful and calculated from the sensor characteristics). There is no preference concerning the intensity or shape of the spectra.

Calculation of the individual error angles also allows a more subtle control of the sampling. It is sometimes helpful if the processing, whether it is analysis or compression, can be focused on a particular type of spectrum (e.g. figure 7.) The top image shows a regular run of ORASIS using a calculated noise angle. The yellow dots show the location of the exemplar origins. The bottom image shows what happens when ORASIS is instructed to process bright spectra as if the calculated error is much bigger than it is while leaving dark spectra at the calculated error angle. The result is that the dark spectra are reconstructed at high fidelity while the bright spectra are reconstructed at lower fidelity (lower than would be justifiable with this sensor.) this is apparent on close inspection of figure 7. The images are from the reconstructed data cubes. Note that the bottom image has more structure in the water (particularly noticeable in the top half of the figure). Less obvious, but present, is the lack of detail in the bright land areas in the lower image. There are several applications for this method. For example, in the case where a scene of composed for two disparate materials, there may be interest only in one of them. ORASIS could be instructed to concentrate most of the exemplar selection process on one or the other of the materials.

A comparison was made between the original data cube and those that are represented in figure 7. The angle between the original and the recalculated spectrum was calculated for the entire scene. Histograms for both cubes are shown in figure 8.

4. Discussion

ORASIS continues to improve both in processing speed and output quality. The BF improvement allows for enhancement in the fidelity of compressed data. Other improvements in the quantization process have been formulated but still have to be investigated.

The connection between ORASIS parameters and sensor characteristics will also help with both the compression and analysis aspects of ORASIS. For analysis, the noise model will allow ORASIS to focus on particular family of spectra. For compression, the link to the noise model will allow a more physically justifiable and easier to use process and better overall reconstructed data.

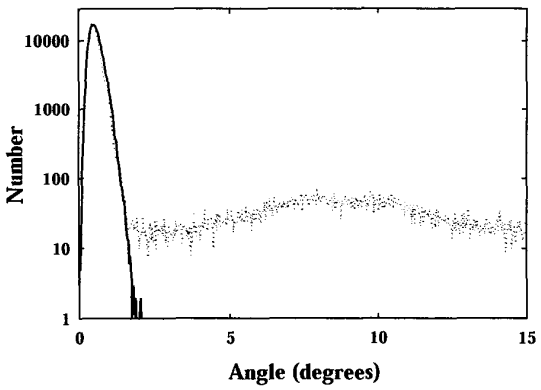

Figure 8: The solid black line represents the regular noise model approach to sampling. The blue (lighter colored) dotted line represents the results obtained when additional logic is used to control how the scene is sampled as described in the text.

5. Acknowledgements

This work was funded by the office of Naval Research. The authors would like to acknowledge M. Carney for her review of the text and Shannon OPray for her technical assistance.

Author Biographies

Jeffrey Bowles: In 1993, Dr. Bowles received a Ph.D. in physics from the University of California at Irvine. His primary research area is that of optical remote sensing. He is the ORASIS lead for the NEMO satellite project. Additional work involves development of hyperspectral imagers deployed from aircraft and methods to analyze the data.

David Gillis: Dr. Gillis received his mathematics Ph.D. in 1996 from the University of Houston, where he studied nonlinear differential equations under Dr. Martin Golubitsky. He is currently an ASEE postdoctoral fellow at the Naval Research Laboratory, where is helping to develop algorithms in hyperspectral imagery as part of the ORASIS project.

Peter Palmadesso: In 1970, Dr. Palmadesso received a Ph.D. from Stevens Institute of Technology in Hoboken, New Jersey. He is currently head of the Special Project in Nonlinear Science at the Naval Research Laboratory in Washington, DC. His areas of expertise include nonlinear algorithms for information processing, nonlinear dynamics and chaos, and nonlinear effects associated with the generation and propagation of waves in plasmas. He has

authored or co-authored more than 90 articles on subjects within these technical areas.

---

[1] FF Sabins, "Remote sensing for mineral exploration", *ORE Geology Reviews*, **14**, 157-183, Sept 1999

[2] PM Treitz and PJ Howarth, "Hyperspectral remote sensing for estimating biophysical parameters of forest ecosystems," *Progress In Physical Geography*, **23**, 359-390, Sept 1999.

[3] ZP Lee, KL Carder, CD Mobley, et al., "Hyperspectral remote sensing for shallow waters: 2. Deriving bottom depths and water properties by optimization," *APPL OPTICS*, **38**, 3831-3843, June 1999

[4] H Holden and E Ledrew, "Hyperspectral identification of coral reef features,", *INT J Remote Sens*, **20**, 2545-2563, Sept 1999

[5] C Brumbley and CI Chang, "An unsupervised vector quantization-based target subspace projection approach to mixed pixel detection and classification in unknown background for remotely sensed imagery," *Pattern Recogn.*, **32**, 1161-1174, JUL 1999

[6] MJ Ryan and JF Arnold "Lossy compression of hyperspectral data using vector quantization", *Remote Sens Environ*, **61**, 419-436, Sept 1997.

[7] Palmadesso et al., in preparation.

[8] Bowles et al., "Real time analysis of hyperspectral data sets using NRL's ORASIS algorithm," *Proceedings of the SPIE*, **3118**, 38, 1997.

[9] Bowles et al., "New results from the NEMO/ORASIS compression algorithm," *Proceedings of the SPIE*, **3753**, 226, 1999.

[10] Davis et al., "Calibration, characterization, and first results with the Ocean PHILLS hyperspectral imager," *Proceedings of the SPIE*, **3753**, 160, 1999.

[11] Bowles, et al., "Calibration of Inexpensive Pushbroom Imaging Spectrometers," *Metrologia*, **35**, 657, 1998.