DERIVATIVE BASED HYPERSPECRAL ALGORITHM FOR BATHYMETRIC MAPPING

David D. R. Kohler, William D. Philpot,
Civil and Environmental Engineering, Cornell University
Ithaca, NY14853

Curtis D. Mobley
Sequoia Scientific Inc., Mercer Island, WA 98040

ABSTRACT

The emergence of passive coastal hyperspectral imagery brings with it the possibility of deriving more accurate bathymetric maps. Past endeavors in remotely sensed bathymetry suggest that multispectral techniques could be extended to accommodate this new data. These procedures, though, do not take total advantage of the true spectral shape that these new data offer. A new procedure that does take this extra information into account shows promise. The technique, derivative analysis, significantly improves the accuracy of the depth estimates. It also is more robust when dealing with the misclassified pixels.

Keywords: bathymetry, hyperspectral remote sensing, derivative analysis

INTRODUCTION

Bathymetric maps derived from passive remotely sensed images have the attractive features of being readily available for most areas and relatively affordable. Unfortunately, relatively low accuracy limits the value of the bathymetric data derived in this way. With hyperspectral data, there is potential for improving the accuracy of classification of coastal waters and with that, to reduce the error in the depth estimates.

The quest to develop accurate bathymetric algorithms that utilized passive digital imagery was begun almost thirty years ago by scientists at Environmental Research Institute of Michigan[1, 2]. This group developed a bathymetry algorithm based on bottom reflectance. This algorithm, which is the basis of almost all studies to date, is expressed as:

$$L_d = L_b e^{-gz} + L_\infty \qquad (1)$$

where, L_d is the radiance received at the airborne detector; L_b is a term which represents the contrast between bottom reflectance and water reflectance at the bottom and the amount of water in the water column; L_∞ is the radiance that would be received at the airborne detector if the water were optically deep; g is the total diffuse attenuation coefficient; and z is the depth of the ocean floor.

It is reasonable to expect that techniques originally applied to multispectral data in past investigations should produce improved results on hyperspectral imagery simply because there will now be more data to work with, providing better error control[3,4]. But hyperspectral data is more than just an increase in the number of wavebands; it represents the spectral character of the signal.

This investigation will compare the results of applying the bathymetric algorithm to a hyperspectral data set using two methods. The first method is typical of traditional methods in which the wavebands are assumed to be independent of each other. The second method is one approach to incorporating a characterization of the spectral shape of the signal using derivatives.

DATA SET DEVELOPMENT

To properly compare different classification techniques for hyperspectral imagery, a controlled data set is needed. An ideal image would be one in which the true values for the depth, the water type, and the bottom type were known for every pixel. In the absence of an image fitting this description, a synthetic data set was developed. The ocean optic program, HYDROLIGHT 3.1, was employed to predict the upwelling spectral radiances needed for this investigation. The program essentially solves a radiative transfer equation (RTE) through numerical methods. While HYDROLIGHT requires the input of the water's inherent optical properties (IOP's), the programmer has supplied the user with default settings so the user need not have data to make a run[5]. Although the output from this model consists of a detailed array of apparent optical properties (AOP's) at various depths within the water column and exiting the water surface, only the water leaving radiance was required for the present work.

Due to a lack of available input data, HYDROLIGHT runs were all made with the default data. The spectral output was chosen to span from 350 to 700 nanometers with a bandwidth of 10 nanometers. Two different output bottom types were chosen: coral sand and green algae. The depths of interest selected were: 0.1m, 0.5m, 1.0m, 1.5m, 2.0m, 2.5m, 3.0m, 3.5m, 4.0m, 4.5m, 5.0m, 5.5m, 6.0m, 7.0m, 8.0m, 9.0m, 10.0m, 13.0m, 16.0m and infinitely deep water (50 meters). A HYDROLIGHT run was performed for each of these depths for each of the bottom types.

The default water column data for HYDROLIGHT uses a chlorophyll concentration that is variable by depth. While this was not ideal for the present purpose (equation 1 was derived assuming a homogenous water column) it can be viewed as a realistic condition. The assumption of a homogeneous water column is necessary, if unrealistic, for solving the remote sensing problem. Rarely can one be sure of the water column parameters when viewing remotely sensed image, thus the inclusion of a data set with a heterogeneous water column will test the robustness of the model.

TRADITIONAL PROCEDURE

Applying the bathymertic methods that were developed for multispectral image data to hyperspectral data is a straightforward procedure. While the techniques were

originated for images with only a few spectral bands, the procedure can easily handle any number of wavelengths [3,4].

Following the assumption that the water is vertically homogenous, the standard procedure is to linearize equation (1) to yield:

$$Ln(L_d - L_\infty) = Ln(L_b) - gz \qquad (2a)$$

when $L_d > L_\infty$, and

$$Ln(L_\infty - L_d) = Ln(-L_b) + gz \qquad (2b)$$

when $L_\infty > L_d$.

Given the spectral response at any two distinct, known depths and given an image in which the signal for optically deep water is present, the spectral curves of g and L_b can be found by solving a set of simultaneous equations. The calculated curves, g and L_b, can then be inserted so that the depth values for pixels of the same water and bottom type can be determined. Philpot[4] demonstrated that Principal Components Analysis (PCA) could be employed to weight the wavebands and thus produce more accurate depth estimates. With the principal components calculated, the next step is to classify the image. The derived principle component vector, a, is multiplied by equation (2a and b) to produce:

$$z = \frac{-\left[a \bullet Ln(L_d - L_\infty) - a \bullet Ln(L_b)\right]}{a \bullet g} \qquad (3)$$

This properly weights the different wavelengths so to generate the most accurate depth estimation.

HYPERSPECTRAL PROCEDURE

Hyperspectral data approximates the true spectral response of an object. Thus, the wavebands no longer need to be treated as if they were independent of each other. Therefore, procedures that exploit the wavebands' relationship to one another should be investigated. The approach taken for this investigation is to use spectral derivatives. This approach was chosen for two reasons. First, derivatives describe spectral shape and capture the relationship that exists between wavebands. Derivatives also amplify subtle changes in the spectral curvature and thus, make it easier to differentiate between different spectral signals.

This procedure will again make use of equation 1. Taking the derivative of this equation with respect to wavelength yields:

$$\text{If } L_d > L_\infty \quad \frac{\frac{\partial}{\partial \lambda}\left(\ln\left[\frac{(L_d - L_\infty)}{(L_b)}\right]\right)}{\frac{\partial}{\partial \lambda}(-g)} = z \qquad (4)$$

PCA was performed on equation (4), but with limited success. The relatively poor performance of PCA is probably related to the fact that PCA is designed to minimize

variance in a multinormal distribution and the distributions produced by the derivatives were not sufficiently normal.

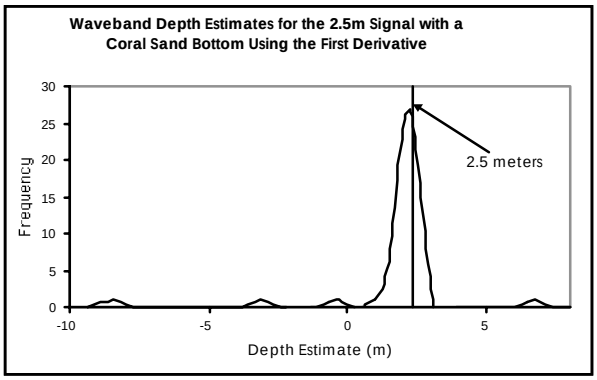

Figure 1

Since PCA proved disappointing for the derivatives, it was decided to look at the depth estimates individual wavelength derivatives produced. Figure 1 displays the distribution depth estimates from the 35 spectral derivatives derived from a single spectrum. This graph resembles the shape of a normal distribution. While not quite a normal distribution, it does demonstrate many of the normal distribution's characteristics; thus, it will be call a pseudo normal curve. Similar curves are present at all depths, using both bottom types, and over all derivatives up to and including the fifth. This trend may continue beyond the fifth derivative but only the first five derivatives were examined for this study. Ignoring the outliers, all the curves are centered on a reasonable estimate of the depth of the water.

Because of the tendency of the derivative derived depth estimates to adhere to a pseudo-normal curve, one might expect that taking the mean of the estimates over all the wavelengths would produce an acceptable estimate of the signals' true depth. But as can be seen in figure 1, several outliers are present and when these outliers are left unattended that can have enormous influence over the mean.

OUTLIER REMOVAL

Determining which outliers to remove is not a simple process. Because of the shear volume of signals in an image and the varying number and position of the outliers present for each signal, a sensitive yet automated approach must be developed.

The obvious approach would be to test the sample farthest from the mean of the sample. If it failed a statistical outlier test, it would be removed. The mean would then be recalculated for the newly trimmed sample and this procedure would be repeated with the next most distant sample. This would continue until the most distant sample from the mean passed the test, and therefore statistically belonged to the trimmed distribution. Then the mean of this outlier free sample would be reported as the signal's depth estimate[6].

While this process may seem fairly robust, it is not. A statistical phenomenon called the *masking effect* could cause the outlier removal process to fail. The *masking effect* occurs when a several closely clumped outliers are so far from the sample's mean that they greatly inflate the sample's standard deviation. The inflated standard deviation in turn renders the outlier test useless, and thus all outliers are considered part of the true distribution. The outlier equation is a simple t-test represented by:

$$t = \frac{|x_i - \bar{x}|}{s} \qquad (5)$$

where, x_i is the ith observation, $\bar{x}$ is the sample mean, and s is the standard deviation of the sample[6,7].

To deal with the potential problem caused by the *masking effect*, an "outwards" outlier test was employed. While this procedure still utilizes equation (5) to test for outliers, this method starts with a core sample and works its way outwards from this center. Rather than removing unfit data points, this method works by determining whether or not any new observation truly belongs to the trimmed distribution. After each new observation is included the mean and standard deviation are recalculated and used to evaluate the next closest observation to the new core.

Obviously, the selection of an appropriate core sample size is vital. After testing several traditional methods, a hybrid approach was chosen, which is best suited for remotely sensed data. This method starts by using the middle 10% of the sample as its core. If this core size had a standard deviation that was so small that no new observations would be included, the innermost outlying observation was added to the core and the core statistics were recalculated. If this new core was also so selective as not to include at least one new observation, the next closest outlier was added to the core. This procedure continued enlarging the core until at least one new observation passed the t-test and was included in the trimmed sample.

RESULTS

This investigation was centered on the comparison of two methods of developing bathymetric maps using passive hyperspectral data. The first method utilized the procedures created for multi-spectral imagery, while the second employed new procedures that were specifically developed for hyperspectral data. The question that needs to be answered is: which technique is better?

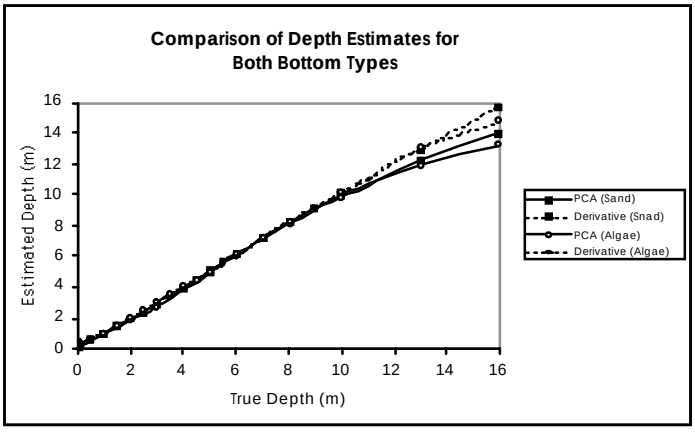

Figure 2

In order to answer this question, graphs of the original technique and the new procedure were constructed. Figure 2 shows the results of the depth estimation for all three evaluated bottom types. By viewing the graphs it should be clear that for both the green algae bottom and the coral sand bottom, the derivative method achieved better results.

Before drawing any final conclusions, the topic of pixel misclassification should be addressed. Up until now all the analysis has been performed assuming that the signal's bottom type was known. In complex scenes, however, such information may possibly be unknown. For example, a scene may contain a grass bed that is so small that it was overlooked. Or the existence of the bed may be noted, but since, the grass is confined to a very narrow range of depths, a L_b for the bottom type can not be determined. To examine

the robustness of the bathymetric equations when handling such problems, a classification was performed on the green algae signal L_b for coral sand. Figure 3 display the results of this analysis. As expected, the PCA approaches run parallel to the true depth[4]. Thus if at least one depth is know for the bottom type, a very accurate estimate can be derived. But, the point of this exercise was to determine how well the methods estimate depth if the pixel's bottom type is unknown. In this case, the derivative prevails, except for the deepest depth estimate.

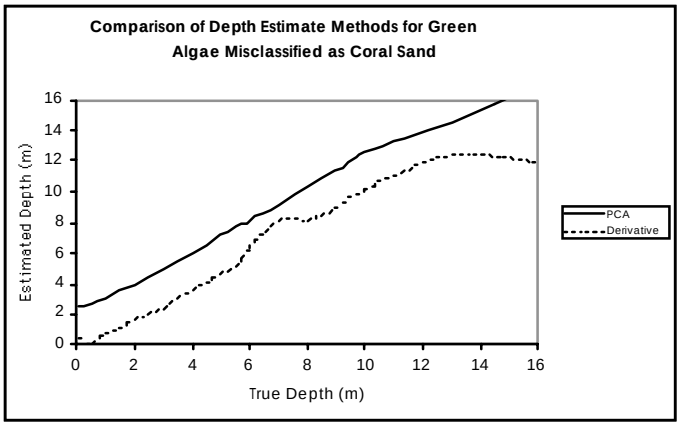

Figure 3

So which approach is better?

While it appears that the derivative procedure's outcome has surpassed the results of the PCA approaches, it is too early to declare it the method of choice. The PCA methods did produce acceptable results and therefore should not be dismissed. The fact that all of these tests have been performed using synthetic data must be kept in mind. When these techniques are applied to real data, unforeseen weaknesses in the approaches may become apparent.

REFERENCES

[1] F. C. Polcyn and I. J. Sattinger, "Water depth determination Using Remote Sensing Techniques," Proceedings of the 6th International Symposium on Remote Sensing of Environment, 1017 (1969).

[2] F. C. Polcyn and D. R. Lyzenga, "Calculations of water Depth from ERTS-MSS Data," Symposium of Significant Results from ERTS-1, NASA Spec, Publ. SP-327 (1973).

[3] D. R. Lyzenga, "Passive Remote Sensing Techniques for Mapping Water depth and Bottom Features," Appl. Oct. 17, 379 (1978).

[4] W. D. Philpot, "Bathymetry Mapping with Passive Multispectral Imagery," Appl. Oct. Apr. 15, 1569 (1989).

[5]C. D. Mobley, "Hydrolight 3.0 Users' Guide," Final Report, SRI Project 5632. SRI International, Menlo Park, CA (1995).

[6]V.Barnett and T. Lewis, Outliers in Statistical Data, John Wiley and Sons: New York (1994).

[7]D. M. Hawkins, Identification of Outliers, Chapman and Hall: New York, (1980).