Effect Of Radiance-To-Reflectance Transformation And Atmosphere Removal On Maximum Likelihood Classification Accuracy Of High-Dimensional Remote Sensing Data*

Joseph P. Hoffbeck and David A. Landgrebe

School of Electrical Engineering
Purdue University, West Lafayette, IN 47907-1285
Phone: (317) 494-3486 Fax: (317) 494-6440
hoffbeck@ecn.purdue.edu landgreb@ecn.purdue.edu

ABSTRACT

Many analysis algorithms for high-dimensional remote sensing data require that the remotely sensed radiance spectra be transformed to approximate reflectance to allow comparison with a library of laboratory reflectance spectra. In maximum likelihood classification, however, the remotely sensed spectra are compared to training samples, thus a transformation to reflectance may or may not be helpful. The effect of several radiance-to-reflectance transformations on maximum likelihood classification accuracy is investigated in this paper. We show that the empirical line approach, LOWTRAN7, flat-field correction, single spectrum method, and internal average reflectance are all non-singular affine transformations, and that non-singular affine transformations have no effect on discriminant analysis feature extraction and maximum likelihood classification accuracy. (An affine transformation is a linear transformation with an optional offset.) Since the Atmosphere Removal Program (ATREM) and the log residue method are not affine transformations, experiments with Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data were conducted to determine the effect of these transformations on maximum likelihood classification accuracy. The average classification accuracy of the data transformed by ATREM and the log residue method was slightly less than the accuracy of the original radiance data. Since the radiance-to-reflectance transformations allow direct comparison of remotely sensed spectra with laboratory reflectance spectra, they can be quite useful in labeling the training samples required by maximum likelihood classification, but these transformations have only a slight effect or no effect at all on discriminant analysis and maximum likelihood classification accuracy.

INTRODUCTION

Remote sensing platforms measure the up welling radiance from the earth in several different spectral bands. In remote sensing applications where the scientist is interested in deriving information about the surface of the earth, it is desirable to adjust for the variations in the measured radiance due to the solar output, atmosphere, and sensor noise, in an attempt to only retain variations caused by the reflectance of the material on the surface. Several radiance-to-reflectance transformations have been proposed that are designed to remove the effect of the solar output and atmosphere in order to estimate the reflectance of the surface.

If an estimate of the surface reflectance is available, it may be possible to identify some materials by comparison to laboratory reflectance curves. We have found this approach useful in labeling training samples for a maximum likelihood classifier when the material has obvious spectral features (Hoffbeck et al., 1993). The question this paper addresses is: Once the training samples have been labeled, is it better to classify using estimates of reflectance or the original radiance data?

In this paper, we show that the empirical line approach (Kruse et al., 1990), LOWTRAN7 (Rast et al., 1991), flat-field correction (Rast et al., 1991), single spectrum method (Bosch, 1990), and internal average reflectance (Bosch, 1990) are non-singular affine transformations, and that non-singular affine transformations have no effect on discriminant feature extraction and maximum likelihood classification. Furthermore, experiments with real data showed that ATREM (CSES et al., 1992) and the log residue method (Rast et al., 1991) slightly decreased classification accuracy.

NON-SINGULAR AFFINE TRANSFORMATIONS

Let x be an n by 1 vector containing the spectral measurements of a pixel. A non-singular affine transformation is any function of the data that can be written in the form $y = A^T x + B$ where A is a constant, non-singular n by n matrix, and B is an optional, constant n by 1 vector.

In the empirical line approach, LOWTRAN7, flat-field correction, single spectrum method, and internal average reflectance, the data in each spectral band is multiplied by a scale factor and in some cases shifted by a constant offset. The same scale factor and offset are used for every pixel in the scene. Therefore, each of these transformations from the original radiance data x, to the estimate of reflectance y, can be written as $y = A^T x + B$ where A is a diagonal matrix whose elements are the scale factors for each band, and B is a vector of the offsets for each band. Since the determinant of a diagonal matrix is the product of the diagonal elements, the determinant of A will be non-zero if all the scale factors are non-zero, which implies that A is non-singular. Thus the empirical line approach, LOWTRAN7, flat-field correction, single spectrum method, and internal average reflectance are non-singular affine transformations.

GAUSSIAN MAXIMUM LIKELIHOOD CLASSIFICATION

Next we show that non-singular affine transformations have no effect on Gaussian maximum likelihood classification. The decision rule in this case is to label the sample x as class j if the density of class j has the highest likelihood value of any of the classes: Choose ω_j if $\arg\max_i [p_x(x|\omega_i)] = j$

where $p_x(x|\omega_i) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_{xi}|}} \exp\left[\frac{-1}{2}(x - m_{xi})^T \Sigma_{xi}^{-1} (x - m_{xi})\right]$,

m_{xi} is the mean vector of class i, and Σ_{xi} is the covariance matrix of class i. Suppose a non-singular affine transformation is applied to the data: $y = A^T x + B$. The mean vector of class i in the transformed data y is $m_{yi} = A^T m_{xi} + B$, the covariance matrix is $\Sigma_{yi} = A^T \Sigma_{xi} A$, and the density is

$$p_y(y|\omega_i) = \frac{1}{\sqrt{(2\pi)^n |\Sigma_{yi}|}} \exp\left[\frac{-1}{2}(y - m_{yi})^T \Sigma_{yi}^{-1} (y - m_{yi})\right]$$

$$= \frac{1}{\sqrt{(2\pi)^n |A^T \Sigma_{xi} A|}} \exp\left[\begin{array}{l} \frac{-1}{2}\left((A^T x + B) - (A^T m_{xi} + B)\right)^T \cdots \\ \cdots (A^T \Sigma_{xi} A)^{-1} \left((A^T x + B) - (A^T m_{xi} + B)\right) \end{array}\right]$$

$$= \frac{1}{\sqrt{(2\pi)^n |A^T| \, |\Sigma_{xi}| \, |A|}} \exp\left[\begin{array}{l} \frac{-1}{2}\left(A^T (x - m_{xi})\right)^T \cdots \\ \cdots A^{-1} \Sigma_{xi}^{-1} (A^T)^{-1} \left(A^T (x - m_{xi})\right) \end{array}\right]$$

$$= \frac{1}{\sqrt{|A|^2} \sqrt{(2\pi)^n |\Sigma_{xi}|}} \exp\left[\frac{-1}{2}(x - m_{xi})^T \Sigma_{xi}^{-1} (x - m_{xi})\right]$$

$$= \frac{1}{\sqrt{|A|^2}} p_x(x|\omega_i)$$

Note that the value of the density of the transformed data y is simply the value of density of the original data x divided by a positive number. Since the density of each class is divided by the same positive number, the class with the largest likelihood value in the original data x will also have the largest likelihood value in the transformed data y. Therefore the classification of the vector x and y is the same, which implies that a non-singular affine

* Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS'94), CD-ROM pp 3289-3294 , Pasadena, CA, August 8-12, 1994.

transformation has no effect on the results of the classification.

DISCRIMINANT ANALYSIS

Discriminant analysis (Fukunaga, 1990) is a feature extraction method that finds a linear combination of the original features that maximizes the separation of the classes. Next we show that if the original data x is changed by a non-singular affine transformation, the discriminant vectors will also be changed by a non-singular affine transformation. Therefore, if the discriminant vectors are classified using a Gaussian maximum likelihood classifier, it will make no difference whether the original data or the transformed data is used.

Let L be the number of classes and p_i be the a priori probability that class i will occur. The global mean is defined as $m_{xo} = \sum_{i=1}^{L} p_i m_{xi}$, the between-class covariance matrix is defined as $\Sigma_{xb} = \sum_{i=1}^{L} p_i (m_{xi} - m_{xo})(m_{xi} - m_{xo})^T$, and the within-class covariance matrix is defined as $\Sigma_{xw} = \sum_{i=1}^{L} p_i \Sigma_{xi}$. Each eigenvalue λ_{xi} and the corresponding eigenvector e_{xi} of $\Sigma_{xw}^{-1}\Sigma_{xb}$ satisfies

$$\left(\Sigma_{xw}^{-1}\Sigma_{xb} - \lambda_{xi} I\right) e_{xi} = 0. \tag{1}$$

Since the rank of $\Sigma_{xw}^{-1}\Sigma_{xb}$ is L-1, there are L-1 eigenvectors whose eigenvalues are non-zero. Let $\Phi_x = \left[e_{x1}\ e_{x2} \cdots e_{x(L-1)}\right]$ be a matrix whose columns are the L-1 eigenvectors. The L-1 by 1 discriminant vectors are defined as $z_x = \Phi_x^T x$.

Now consider the effect of a non-singular affine transformation $y = A^T x + B$. Using the transformed data, the global mean, the between-class covariance, and the within-class covariance are as follows.

$$m_{yo} = \sum_{i=1}^{L} p_i m_{yi} = \sum_{i=1}^{L} p_i \left(A^T m_{xi} + B\right) = A^T m_{xo} + B$$

$$\begin{aligned}
\Sigma_{yb} &= \sum_{i=1}^{L} p_i \left(m_{yi} - m_{yo}\right)\left(m_{yi} - m_{yo}\right)^T \\
&= \sum_{i=1}^{L} p_i \left(A^T m_{xi} + B - \left(A^T m_{xo} + B\right)\right)\left(A^T m_{xi} + B - \left(A^T m_{xo} + B\right)\right)^T \\
&= \sum_{i=1}^{L} p_i A^T (m_{xi} - m_{xo})(m_{xi} - m_{xo})^T A \\
&= A^T \Sigma_{xb} A
\end{aligned}$$

$$\Sigma_{yw} = \sum_{i=1}^{L} p_i \Sigma_{yi} = \sum_{i=1}^{L} p_i A^T \Sigma_{xi} A = A^T \Sigma_{xw} A.$$

The eigenvectors and eigenvalues satisfy

$$\begin{aligned}
\left(\Sigma_{yw}^{-1}\Sigma_{yb} - \lambda_{yi} I\right) e_{yi} &= 0 \\
\left(\left(A^T \Sigma_{xw} A\right)^{-1} A^T \Sigma_{xb} A - \lambda_{yi} I\right) e_{yi} &= 0 \\
\left(A^{-1}\Sigma_{xw}^{-1}\Sigma_{xb} A - \lambda_{yi} A^{-1} I A\right) e_{yi} &= 0 \\
\left(\Sigma_{xw}^{-1}\Sigma_{xb} - \lambda_{yi} I\right) A e_{yi} &= 0
\end{aligned} \tag{2}$$

Comparing equations (1) and (2), we see that the affine transformation did not change the eigenvalues: $\lambda_{yi} = \lambda_{xi}$. Also, each eigenvector from the transformed data is a simple function of the eigenvector from the original data: $e_{yi} = A^{-1} e_{xi}$. Let $\Phi_y = \left[e_{y1}\ e_{y2} \cdots e_{y(L-1)}\right]$ be a matrix whose columns are the eigenvectors. The discriminant vectors of the transformed data are

$$\begin{aligned}
z_y &= \Phi_y^T y \\
&= \left[A^{-1}e_{x1}\ A^{-1}e_{x2} \cdots A^{-1}e_{x(L-1)}\right]^T \left(A^T x + B\right) \\
&= \left(A^{-1}\left[e_{x1}\ e_{x2} \cdots e_{x(L-1)}\right]\right)^T \left(A^T x + B\right) \\
&= \left[e_{x1}\ e_{x2} \cdots e_{x(L-1)}\right]^T \left(A^{-1}\right)^T \left(A^T x + B\right) \\
&= \Phi_x^T x + \Phi_x^T \left(A^{-1}\right)^T B \\
&= z_x + \Phi_x^T \left(A^{-1}\right)^T B
\end{aligned}$$

Note that the discriminant vectors of the transformed data z_y are simply the discriminant vectors of the original data z_x plus a constant offset. Since adding a constant offset is a non-singular affine transformation (where A=I), the Gaussian maximum likelihood classification of the discriminant vectors will be the same regardless of whether the original data or the transformed data is used.

EXPERIMENTS USING ATREM AND LOG RESIDUE

In the ATREM and log residue methods, the transformation is not constant across the scene. For example in ATREM, the amount of water vapor in each pixel is estimated from the values of the data near the water absorption bands, and this estimate is used in the transformation of that pixel. In the log residue method, each pixel spectrum is divided by the spectral average of that pixel. Since the transformation for each pixel depends on the value of the pixel, these transformations are not affine and, and so they may or may not effect classification accuracy.

In order to evaluate the effect of the ATREM and log residue transforms on classification accuracy, experiments were performed with real AVIRIS data. Pixels whose class was known were selected from the radiance data. A certain percentage (12.5, 25, or 50%) of these pixels were selected at random to be training samples, and the rest were used as test samples. The training samples were used to compute the discriminant eigenvectors and to train a Gaussian maximum likelihood classifier. The discriminant vectors were extracted from the test samples and classified. The average classification accuracy for 10 trials using random training sets was computed. Then the whole experiment was repeated using the reflectance data and identical training sets. The details of each experiment and the average classification accuracy are presented below.

Effect Of ATREM On Classification Accuracy

Experiments were performed with ATREM data from two sites: Jasper Ridge, California, and Tippecanoe County, Indiana. The 1992 AVIRIS radiance data taken over Japer Ridge was converted to reflectance using the ATREM program. Classes from this site were identified by comparing the scene to a ground cover map made from aerial photographs (Gamon et al., 1993). In all, 3207 pixels were selected from the following classes: evergreen woodland, serpentine grassland, greenstone grassland, deciduous woodland, chaparral, and water. The 193 spectral bands (0.40-1.34, 1.43-1.80, and 1.95-2.47μm) outside the water absorption bands were used, and the average classification accuracy using 12.5, 25, and 50% of the pixels for training samples is shown in Figure 1. Note that in the trials, the average classification accuracy was slightly lower for the ATREM data than for the original radiance data.

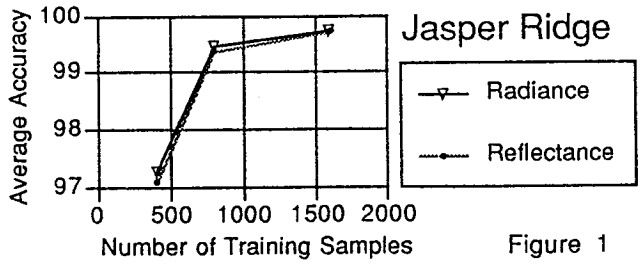

Figure 1

The 1992 AVIRIS scene taken over Tippecanoe County, Indiana was also processed using the ATREM program. Ground observations were used to identify a total of 2521 pixels from the following classes: beans with no residue, beans with corn residue, corn with no residue, corn with bean residue, corn with wheat residue, and wheat with no residue. The spectral bands used (0.42-1.34, 1.43-1.80, and 1.95-2.47μm) were outside the water absorption bands and totaled 191 bands. Again, the average classification accuracy was slightly lower for the ATREM data than for the original radiance data (see Figure 2).

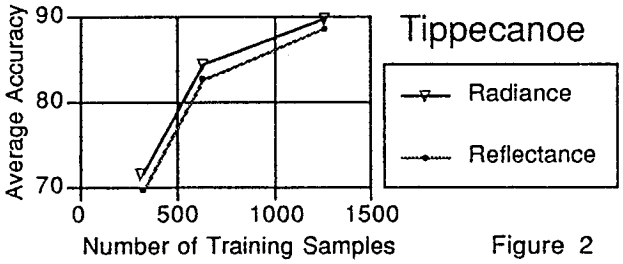

Figure 2

Effect Of Log Residue On Classification Accuracy

The log residue transform was computed using the 1992 AVIRIS scene taken over Cuprite, Nevada. Four of the classes (alunite, buddingtonite, kaolinite, and quartz) had easily identifiable absorption features and were identified by comparing the log residue spectrum to laboratory reflectance curves (Goetz et al., 1985). The classes alluvium, argillized, tuff, and playa were identified by comparing the scene to a geology map produced from ground observations (Abrams et al., 1977, and Kruse et al., 1990). The experiment used 2744 pixels and 191 bands (0.40-1.34, 1.43-1.80, 1.96-2.46μm). The average classification accuracy was slightly lower for the log residue data than for the radiance data (see Figure 3).

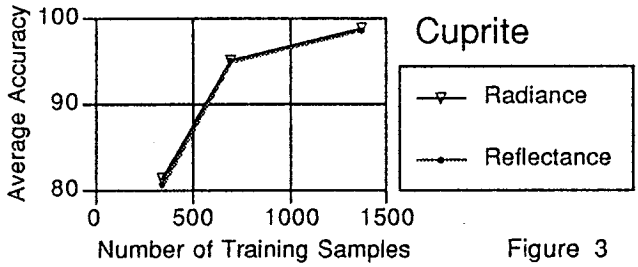

Figure 3

When only the 4 classes which were identified using the log residue method and the 31 bands (2.05-2.35μm) surrounding the absorption features of these minerals were used, the average accuracy of the log residue data was slightly higher than that of the original radiance data (see Figure 4). This improvement did not occur, however, when the other classes and the other bands were used.

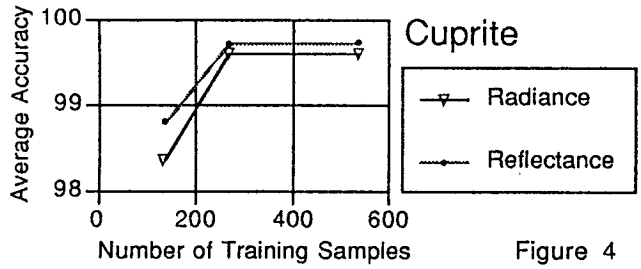

Figure 4

SIMILARITY OF CLASSIFICATIONS

Another experiment was performed with each of the three scenes to measure how many pixels would be classified differently when the reflectance data was used instead of the radiance data. In this experiment, all the samples of known class were used to train a Gaussian maximum likelihood classifier, and a large portion of the scene was classified. The percentage of pixels whose classification differed between the radiance classification and the reflectance classification was 2.7% (Jasper Ridge), 8.4% (Tippecanoe County), and 5.6% (Cuprite).

CONCLUSION

Although an estimate of the reflectance of a pixel can be quite useful in labeling training samples for a maximum likelihood classifier, it matters little if the actual classification is performed on the original radiance data or data that has been transformed to approximate reflectance. This result suggests that Gaussian maximum likelihood classification is insensitive to the variations caused by the solar curve and the atmosphere, so that attempts to correct for these changes are unnecessary.

REFERENCES

Abrams, M.J.; Ashley, R.P.; Rowan, L.C.; Goetz, A.F.H.; and Kahle, A.B. "Mapping of Hydrothermal Alteration in the Cuprite Mining District, Nevada, Using Aircraft Scanner Images for the Spectral Region 0.46 to 2.36 μm." *Geology* 5(1977):713-718.

CSES (Center for the Study of Earth From Space) and CIRES (Cooperative Institute for Research in Environmental Sciences), University of Colorado, Boulder, *ATmosphere REMoval Program (ATREM) User's Guide*, 1992, pp. 5-8.

Bosch, J. van den, "Application of LOWTRAN 7 as an Atmospheric Correction to Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Data," in Proceedings of the Second Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Workshop, JPL Publication 90-54, 1990, pp. 78-81.

Fukunaga, K. *Introduction to Statistical Pattern Recognition.* Boston: Academic Press, Inc., 1990, pp. 445-455.

Gamon, J.A.; Field, C.B.; Roberts, D.A.; Ustin, S.L.; Valentini, R. "Functional Patterns in an Annual Grassland during an AVIRIS Overflight." *Remote Sensing of the Environment* 44(1993):239-253.

Goetz, A.F.H. and Srivastava, V. "Mineralogical Mapping in the Cuprite Mining District, Nevada," in Proceedings of the Airborne Imaging Spectrometer Data Analysis Workshop, JPL Publication 85-41, 1985, pp. 22-31.

Hoffbeck, J.P. and Landgrebe, D.A. "Classification of High Dimensional Multispectral Image Data," in Summaries of the Fourth Annual JPL Airborne Geoscience Workshop, JPL Publication 93-26, 1993, pp. 93-96.

Kruse, F.A.; Kierein-Young, K.S., and Boardman, J.W. "Mineral Mapping at Cuprite, Nevada with a 63-Channel Imaging Spectrometer." *Photogrammetric Engineering and Remote Sensing* 56(1990):83-92.

Rast, M.; Hook, S.J.; Elvidge, C.D.; and Alley, R.E. "An Evaluation of Techniques for the Extraction of Mineral Absorption Features from High Spectral Resolution Remote Sensing Data." *Photogrammetric Engineering and Remote Sensing* 57(1991):1303-1309.