A FEATURE EXTRACTION SCHEME FOR QUALITATIVE AND QUANTITATIVE ANALYSIS FROM HYPERSPECTRAL DATA — DATA INTEGRATION OF HYPERSPECTRAL DATA —

Sadao FUJIMURA and Senya KIYASU

The University of Tokyo, Graduate School of Engineering
Department of Mathematical Engineering and Information Physics
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, JAPAN

Commission II

KEY WORDS : Data compression, Classification accuracy, Significance weighted, Hyper-dimensional, Quantitative estimation , Measurement

ABSTRACT

Extracting significant features is essential for processing, storing and/or transmission of a vast volume of hyperspectral data. Conventional ways of extracting features are not always satisfactory for this kind of data in terms of optimality and computation time.
We present here a purpose-oriented feature extraction scheme from hyperspectral data designed both for qualitative analysis (supervised classification) and quantitative analysis (density estimation). In processing or analysis of data we have some kinds of purpose or intention. We introduce subjective significance explicitly into feature extraction. We assume that we can get typical data describing the properties of objects: that is, training data for supervised classification and spectral data for various values of a quantity can be obtained. After all these data are orthogonalized and reduced by principal component analysis, a set of appropriate features for prescribed purpose or range is extracted as linear combinations of the reduced components. The features describe the properties of classes in classification and the variation of spectra of the quantity in density measurement. The feature specifies weights for the data of each hyperspectral dimension. Thus the dimension is fused and integrated according to the extracted features, which means new channels are generated from the hyperspectral data.
The performance of the method is examined, and the validity of the feature extraction algorithm is confirmed. We used our ground-based imaging spectrometer for obtaining hyperspectral (about 400 channel) data, and applied our method for feature extraction to supervised classification and density estimation of a pigment. It is shown that our method yielded satisfactory results.

1. INTRODUCTION

Recently the dimension of remotely sensed data becomes higher and higher because of higher spectral resolution, increasing number of sensors, and multi-temporal observations. For example, the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) has 224 spectral bands in the 0.4 - 2.5μm region (Vane,1988). In order to efficiently obtain necessary information from these hyperspectral data, or to transmit the data through a communication channel, the quantity of data must be reduced. This can be achieved by extracting significant features.

One of the conventional methods of feature extraction utilizes an exhaustive search for selecting the best subset of sensor channels (Landgrebe,1978). However, when it is applied to hyperspectral data, astronomical computation time is required to evaluate all the combinations of channels. Hyperspectral data contain useful information spread in many channels. Therefore it would be better to fuse the data from many dimensions and generate new channels to extract useful information. Principal component analysis (Ready,1973) or canonical analysis (Schowengerdt,1983) are the methods which can be used for fusing many dimensions of data. Though they extract features which yield high average separability among classes, they are not always suitable for purpose-oriented classification, because they are not selected from the viewpoint of particular discrimination.

We present here a purpose-oriented feature extraction scheme designed both for qualitative analysis (supervised classification) and for quantitative analysis (density measurement). In processing or analysis of data we have some kinds of purpose or intention. This leads us to introduce subjective significance explicitly into feature extraction (Fujimura,1994,1997). The destination of our feature extraction in classification is to extract a set of features which optimally separate one class from another among a particular set of important classes. As for quantitative estimation using hyperspectral data, our destination is to estimate the quantity in a particular range as accurately as possible. We regard the range setting in measurement or estimation as such like intention or significance, and features describing the variation of spectra are extracted by using the data for various values of a quantity.

We assume that we can get typical data describing the

properties of objects: that is, training data for supervised classification and spectral data for various values of a quantity can be obtained for analysis. After all these data are orthogonalized and reduced by principal component analysis, a set of appropriate features for prescribed purpose are extracted as linear combinations of the reduced components. These features describe the properties of classes in classification and the variation of spectra of the quantity in quantitative estimation. In this way the features specify weights for the data of each hyperspectral dimension. Thus the dimension is fused and integrated according to the extracted features, which means that new channels are generated from hyperspectral data.

The performance of the method is examined, and the validity of the feature extraction algorithm is confirmed. We use our ground-based imaging spectrometer for obtaining hyperspectral (about 400 channel) data, and apply our method for feature extraction to supervised classification and density measurement of a pigment. It is shown that the method proposed in this paper yields satisfactory results.

2. QUALITATIVE ANALYSIS — Supervised Classification —

2.1 Description of Data

First of all and as usual, we assume that we can get training data for almost all the classes in an image to derive feature with: that is, we can estimate the characteristics of most classes included in the image.
We denote hyper-dimensional data (N dimension) by a vector $\mathbf{y} = (y_1, \cdots, y_N)'$ ($'$: transpose), and suppose that they are classified into one of, say, n classes. Then, $\mathbf{y}$ can be decomposed into class mean $\mathbf{y}_a$ and within-class dispersion $\mathbf{y}_e$: that is, $\mathbf{y}$ is written as

$$\mathbf{y}_{ij} = \mathbf{y}_{a_i} + \mathbf{y}_{e_{ij}} \quad (i = 1 \sim n,\ \ j = 1 \sim m_i), \tag{1}$$

(see Fig. 1), where $\mathbf{y}_{ij}$ is j-th data of class i. We write the covariance matrix of $\mathbf{y}$, $\mathbf{y}_a$ and $\mathbf{y}_e$ as $\mathbf{C}_{yy}$, $\mathbf{C}_a$ and $\mathbf{C}_e$ respectively. We call $\mathbf{C}_a$ and $\mathbf{C}_e$ between-class and within-class covariance matrix, respectively. Here, we assume that the covariance matrix of each class is identical. This assumption is rather reasonable from the view point of the generality of training data.

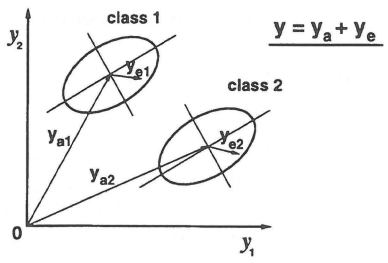


Figure 1 Description of data

2.2 Feature Extraction

Here, for simplicity we consider two cases where one and two most important classes should be discriminated from all the other classes. In general, classification accuracy increases as the separability[1] of classes increases. We use separability to evaluate the performance of features extracted. We extract the features which maximize the separability of a particular pair of classes that we wish to discriminate.

Our method proposed here consists of two steps of processing: pre-processing and feature extraction.

In the pre-processing, hyper-dimensional data $\mathbf{y} = (y_1, \cdots, y_N)'$ are reduced and normalized to m ($m \ll N$) components $\mathbf{z} = (z_1, \cdots, z_m)'$ by a linear transformation $\mathbf{z} = \mathbf{A}'\mathbf{y}$. From the assumption on $\mathbf{C}_e$, the within-class dispersion of each class in the original space has the same ellipsoidal shape shown in Fig. 1. After transformation, they are normalized into an m dimensional sphere. This makes the space uniform: this means that the distance measured in terms of variance does not have directionality in the space.

In the second step, features are successively extracted until there remains no class which has distance from the particular classes less than the minimum distance obtained so far. Feature extraction is done by determining sub-space in the feature space: that is, by making a linear combination of $\mathbf{z}$ as $\mathbf{a}'\mathbf{z}$, where $\mathbf{a}$ is an m dimensional weight vector which we call here feature vector. Thus, feature extraction is no other than the determination of a feature vector. As the space is uniform now, the direction of an optimal feature vector which discriminates between two classes is obtained just by connecting the centers of these classes. The feature vectors obtained are orthogonalized to make independent.

The procedures for determining successive feature vectors is as follows:

(1) First, we set an optimal feature vector $\mathbf{a}_1$ between the two nearest classes among the prescribed classes.
(2) Next, we evaluate the separability on $\mathbf{a}_1$ for all the combination of the prescribed classes.
(3) If there is any pair of prescribed classes which does not have enough separability, we set an additional feature vector $\mathbf{a}_2$ between them. We ortho-normalize the new vector $\mathbf{a}_2$ with $\mathbf{a}_1$ as shown in Fig. 2, so that this feature is independent of the first one.
(4) Features are successively extracted in the same way until all the distance among the prescribed classes are larger than the minimum distance obtained so far.
(5) Then, we apply the procedures (2)$\sim$(4) to the distance among the prescribed and the other classes.

When only one class is prescribed, the procedure starts from setting a feature vector between the class and its nearest class in the feature space.

[1] We used the divergence as a measures of separability. We call it as distance in the rest of this paper.

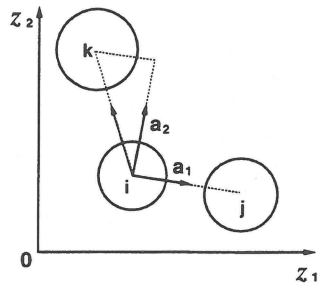


Figure 2 Feature vectors discriminating between two classes

A feature $\mathbf{a}_i' \mathbf{z}$ is equivalent to $(\mathbf{A}\,\mathbf{a}_i)' \mathbf{y}$ expression using original data $\mathbf{y}$, because $\mathbf{z} = \mathbf{A}'\mathbf{y}$, where $(\mathbf{A}\,\mathbf{a}_i)$ means the weighting factor for spectral data.

2.3 Experimental Results and Discussion

We acquired data for four growth-stages of tree leaves (A~D: from young to fallen), soil, stone and concrete by using a ground-based imaging spectrometer which we developed. It produces 411 dimensional hyper-spectral data. For estimating the mean and the variance of each class, 45 training data were used. Averaged relative reflectance is shown in Fig. 3. In the following, the covariance matrices of the classes are set identical.

After reducing and normalizing the data to 7 orthogonal components, features were extracted from one to another. We selected classes A and B as those significant to be classified. The first feature vector was set between class A and B. The next feature vector was set between A and the nearest class C. There remains no other classes whose distance from A or B is less than that between A-B. The two features characterize the weighting factor are shown in terms of wavelength in Fig. 4.

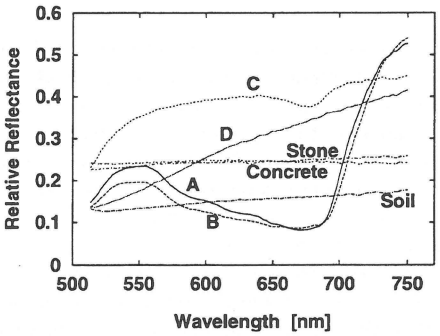


Figure 3 Spectral reflectance of objects (A~D : Leaves of plant)

In this case the distance of each class from class A and B is shown in Table 1: (a) in the original 8 dimensional space, (b) 1 dimension (the first feature), and (c) two dimensional space made by the first two features. From (c), it is seen that the minimum distance is that between A and B which was already obtained in (a). Thus, the two features are sufficient for this case.

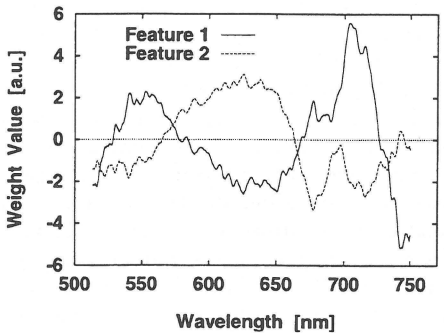


Figure 4 Weighting factors for the significant classes A and B

Table 1 Distance from A and B (Relative Distance) (A~D: Leaves, E: Soil, F: Stone, G: Concrete)

(a) Distance in 7 Dimension

	A	B	C	D	E	F	G
A	—	4.2	19.3	16.3	17.3	17.0	18.2
B	4.2	—	19.5	16.2	17.3	17.2	18.0

(b) Distance in 1 Dimension

	A	B	C	D	E	F	G
A	—	4.2	1.1	2.7	2.0	1.2	1.8
B	4.2	—	3.2	1.5	2.2	3.0	2.4

(c) Distance in 2 Dimension

	A	B	C	D	E	F	G
A	—	4.2	19.3	13.7	16.3	16.3	17.5
B	4.2	—	19.5	13.5	16.3	16.5	17.5

To confirm the validity of this method when compared with canonical analysis, the classification accuracy was estimated by test samples. Figure 5 shows the classification accuracy for the class A and B in terms of the number of features. The accuracy depends on the number of features used, and is higher than that by canonical analysis by about 35% (one feature) and 18% (two features). The confusion matrix is shown in Table 2.

3. QUANTITATIVE ANALYSIS — Density Measurement —

3.1 Feature Extraction Algorithm

The two graphs in Fig. 6 illustrate the characteristics of spectral transmittance for various density of a pigment (methylene blue). Fig. 6 (a) shows the spectra for density from 2 to 20%, and Fig. 6 (b) those for the density from 20 to 100%. It is seen that the most sensitive region of wavelength for density estimation depends on the density of solution. Fig. 6 suggests that it is reasonable to weight the data for wavelength according to the density to be measured.

Table 2 Confusion Matrix
(A~D: Leaves, E: Soil, F: Stone, G: Concrete)

(a) Classification Using First Feature

(%)

	A	B	C	D	E	F	G
A	69.4	0	15.8	6.1	2.6	2.0	4.1
B	0	83.2	0	16.8	0	0	0
C	31.6	2.0	7.1	30.1	15.3	6.6	7.1
D	0	0.5	1.0	45.9	25.5	8.2	18.9
E	0	0.5	4.1	25.5	28.1	8.2	33.7
F	7.7	0	23.0	3.1	17.4	20.9	28.1
G	1.0	0	9.2	27.0	30.6	10.2	21.9

(b) Classification Using First Two Features

(%)

	A	B	C	D	E	F	G
A	91.8	8.2	0	0	0	0	0
B	0	100.0	0	0	0	0	0
C	0	0	83.2	3.1	1.0	11.2	1.5
D	0	0	0	76.5	19.4	0	4.1
E	0	0	0	0	83.7	16.3	0
F	0	0	2.6	0	29.6	53.6	14.3
G	0	0	17.9	0	12.8	11.2	58.2

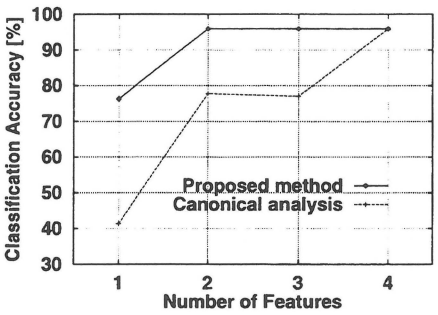
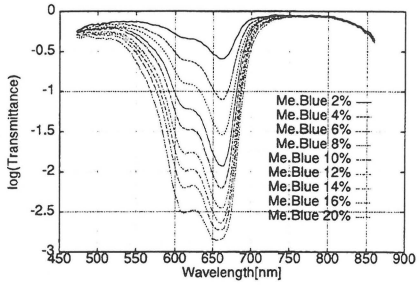


Figure 5 Classification accuracy of calsses A and B versus number of features

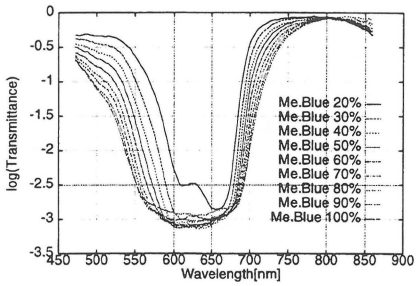
We assume that we get training data for several values of quantity and estimate the characteristics for each quantity. The data from an object (class) is considered to be defined with mean and covariance as described above.

We consider that a quantity in the region defined by the two values should be estimated with high accuracy. This corresponds to determining features to discriminate the two classes. We use a measure of separability between two classes and extract features which maximize the separability.

We applied the same algorithm above to the estimation of density. After the features are extracted for each range of quantity, an estimation curve was determined. We fitted a polynomial to the relationship between the value of a feature and the quantity to be estimated. The quantity of unknown samples was determined by using the estimation curve.



(a) $2, 4, \cdots, 20\%$



(b) $20, 30, \cdots, 100\%$

Figure 6 Spectral transmittance of methylene blue solution

3.2 Experiments

We acquired data of spectral transmittance about methylene blue solution for various density by using a the imaging spectrometer above.

We obtained 360 dimensional data between 470 and 860 nm. For estimating the mean and the variance, 110 sets of (training) data were used for each class. Averaged relative transmittance is shown in Fig. 7. The transmittance is expressed in log scale. We divided the range into two ranges, one is from 2 to 20%, and the other from 20 to 100%. Two features shown in Fig. 8 were extracted for each range. This figure shows how we weight and fuse the data.

Fig. 9 shows the relationship between the density and the value of the features. The dashed lines are the quadratic polynomials fitted for density estimation. The polynomials are written by eq.(2), where D is the density and S the value of a feature obtained from measurement. From eq.(2), the density is determined as shown in eq.(3).

$$\mathrm{S(D)} = \begin{cases} 0.383D^2 - 19.3D + 12.6 & (2 \sim 20\%) \\ -0.00133D^2 - 2.30D + 101 & (20 \sim 100\%) \end{cases} \tag{2}$$

$$\mathrm{D(S)} = \begin{cases} -\sqrt{2.61S + 603} + 25.2 & (2 \sim 20\%) \\ -\sqrt{-750S + 8.22 \times 10^5} - 864 & (20 \sim 100\%) \end{cases} \tag{3}$$

Table 3 shows the results of density estimation in the region from 2 to 20%. The estimation results from two conventional methods of estimation are also shown in the table: one from the sum of the transmittance at two different wavelength and the other from the ratio of the two values. Compared to the conventional methods(Gordon,1983), the deviation of the quantity estimated by our method greatly reduced.

Table 3 Result of density estimation ($2 \sim 20\%$)

density (%)	conventional 1 $\lambda_1 + \lambda_2$		conventional 2 λ_1 / λ_2		proposed method	
	density	Std.Dev.	density	Std.Dev.	density	Std.Dev.
2	2.15	0.11	2.02	0.33	1.83	0.071
6	5.91	0.15	5.68	0.94	5.88	0.073
10	9.89	0.23	10.4	1.8	9.87	0.13
14	14.0	0.43	13.3	1.7	13.9	0.14
20	19.7	1.2	20.0	3.5	20.0	0.16

Std.Dev. : Standard Deviation

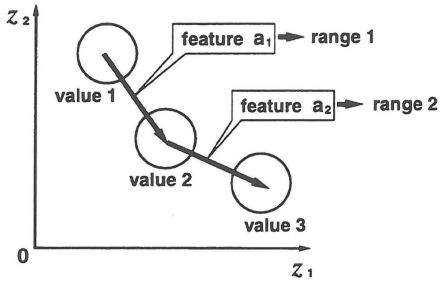


Figure 7 Feature extraction in normalized space

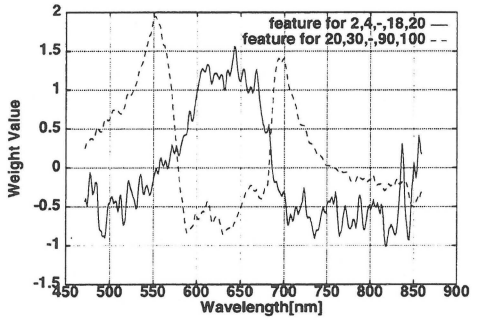
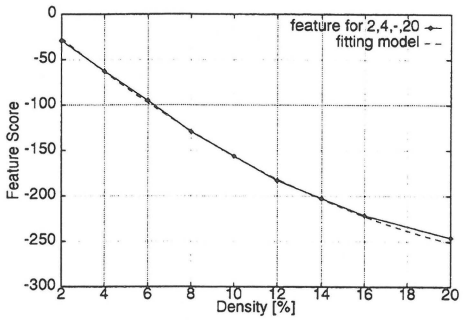
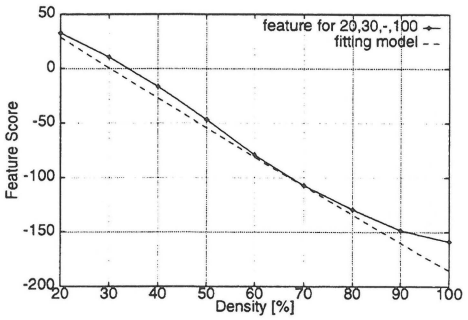


Figure 8 Extracted features



(a) $2, 4, \cdots, 20\%$



(b) $20, 30, \cdots, 100\%$

Figure 9 Estimation of density using the extracted features

4. CONCLUSIONS

We have proposed a purpose-oriented feature extraction scheme from hyperspectral data both for qualitative analysis (supervised classification) and for quantitative analysis (density measurement). By extracting features, hyperspectral data are integrated or fused to generate new channels.

The method was tested using about 400 dimensional hyperspectral data obtained by a ground-based imaging spectrometer we developed. It was found by numerical simulation using the data above that a small number of features to classify the prescribed significant classes were extracted and that classification accuracy of particular classes increased by more than several percents, compared with classification using the features extracted by canonical analysis.

We have also applied the approach to density measurement of methylene blue in a solution. Experimental equations for the estimation of the density of methylene blue were derived using the extracted features. The range they cover was determined dependently on the accuracy. It was found that precision of density estimation highly increased compared with the conventional methods using the sum or the ratio of the values at two wavelengths.

Application of this qualitative scheme to evaluating the performance of current multi-spectral sensors and to designing spectral bands of new sensors, application of this quantitative scheme to the data of transmittance with scattering and those of reflectance, and application to real remotely sensed data are subjects for future studies.

REFERENCES

References from Journals:

Ready,P.J. and Wintz,P.A.,1973. Information Extraction, SNR Improvement, and Data Compression in Multispectral Imagery. IEEE Trans. Comm., COM-21, pp.1123-1131.

Gordon,H.R., Clark,D.K., Brown,J.W., Brown,O.B., Evans,R.H. and Broenkow,W.W.,1983. Phytoplankton pigment concentrations in the Middle Atlantic Bight: comparison of ship determinations and CZCS estimates. Appl.Opt., 22, pp.20-36.

References from Books:

Landgrebe,D.A.,1978. Useful Information from Multispectral Image Data: Another Look, in Remote Sensing: The Quantitative Approach, Swain,P.H. and Davis,S.M. Eds. McGraw-Hill, pp.358-361.

Schowengerdt,R.A.,1983. Techniques for Image Processing and Classification in Remote Sensing. Academic Press, pp.159-167.

References from Other Literature:

Fujimura,S. and Kiyasu,S.,1994. Significance-Weighted Feature Extraction from Hyper-Dimensional Data. Proc. of SPIE, 2318, pp.63-68.

Fujimura,S. and Kiyasu,S.,1997. A Method for Object-Oriented Feature Extraction from Hyperspectral Data — Generation of New Channels by Fusion of Data—. Proc. IGARSS'97, 2, pp.975-977.

Vane,G.,1988. First results from the Airborne Visible/ Infrared Imaging Spectrometer (AVIRIS). Proc. of SPIE, 834, pp.166-174.