Land-use/Land-cover Classification with Multispectral and Hyperspectral EO-1 Data

Bing Xu and Peng Gong

Abstract

We compared the capability of the Earth Observing-1 (EO-1) Hyperion hyperspectral (HS) data with that of the EO-1 Advanced Land Imager (ALI) multispectral (MS) data for discriminating different land-use and land-cover classes in Fremont, California. We designed a classification scheme of two levels with level I including general classes and level II including more specific classes. Classification shows that the HS data does not produce better results than the MS data when we directly applied a Mahalanobis distance (MD) classifier.

We tested a number of feature reduction and extraction algorithms for the HS image. These algorithms include principal component analysis (PCA), segmented PCA (SEGPCA), linear discriminant analysis (LDA), segmented LDA (SEGLDA), penalized discriminant analysis (PDA) and segmented PDA (SEGPDA). Feature reductions were all followed by an MD classifier for image classification. With SEGPDA, SEGLDA, PDA, and LDA, similar accuracies were achieved while a segmentation-based approach we proposed (SEGPDA or SEGLDA) greatly improved computation efficiency. They all outperformed SEGPCA and PCA by 4 to 5 percent (level II) and 1 to 3 percent (level I) in classification accuracy.

For level II classification, overall accuracies obtained by using the features extracted from the HS image were 2 to 3 percent greater than those obtained with the MS image. For various vegetation class and impervious land use categories, the HS data consistently produced better results than the MS data. For level I classification, the HS image generated a thematic map that is <0.01 greater in kappa coefficient comparing to the MS image. When we collapsed the level II classification map to a level I map, 5 percent (HS) to 7 percent (MS) improvements were achieved.

Introduction

The EO-1 Hyperion is the only hyperspectral (HS) sensor operated in space (NASA, 1996). An HS sensor has contiguous narrow wavelength bands (about 10 nm each) that are able to capture more subtle spectral details of the objects on the ground than a multispectral (MS) sensor (about 100 nm each). However, it is often difficult to process the high dimension of such data (>200 bands). The limited number of training samples comparing to the high dimension of data and the high correlation among the adjacent bands will lead to inaccurate estimation of the covariance structures and degenerate ranks of spectral matrices, thus limiting the accuracy of classification (Hughes, 1968; Hsu *et al.*, 2002).

Land-use classification in an urban setting using multispectral data or panchromatic imagery has been widely studied. Spectral, contextual, texture, and structural information are extracted to aid the characterization of different and complex land surfaces and to improve the accuracy of identification (Gong and Howarth, 1990a and 1990b; Gong and Howarth, 1992; Deguchi and Sugio, 1994; Ridd, 1995; Xu *et al.*, 2003). Gong *et al.* (1997, 2001, and 2002) measured *in situ* HS data in order to establish a spectral library, to recognize different species of conifers, and to further extract ecological parameters by data transformation and selection of biophysically sensitive bands. Thenkabail *et al.* (2004) reported substantial improvement in rainforest type classification in Africa when using Hyperion data compared to the use of Advanced Land Imagery (ALI), Ikonos, and Landsat ETM+ data. However, analysis of imaging spectrometer data for urban land-use applications has rarely been explored. Roessner *et al.* (2001) applied the concept of linear spectral unmixing, considering spatial neighborhood using a procedure of iterative endmember selection to differentiate land surfaces in an urban area in Germany.

Therefore, two questions lead us to this research. First, does the HS image contain more information than the MS image for urban land-use classification? This question has not yet been addressed probably due to the unavailability of simultaneous acquisition of both the MS and HS data. The launch of EO-1 made it possible for us to explore and compare the classification results from the two sensors. Second, as for HS data, what feature reduction/extraction methods would be better in terms of both computational efficiency and classification accuracy?

Direct spectral matching using binary coding and vectorization with known spectra in a spectral library to label the unknown pixels is possible because of the high spectral resolution (Jia, 1996; Goetz *et al.*, 1985). Due to the complexities caused by the high dimensional space of the data, feature extraction schemes such as PCA or fisher's LDA have been applied in transforming and reducing the data dimension by maximizing the ordered variance of the whole

Bing Xu is with the Center for Natural and Technological Hazards, Department of Geography, University of Utah, 260 S. Central Campus Dr., Rm. 270, Salt Lake City, UT 84112-9155 (bing.xu@geog.utah.edu).

Peng Gong is with the Department of Environmental Science, Policy, and Management, 137 Hilgard Hall, University of California, Berkeley, CA 94720 and the State Key Lab of Remote Sensing Science, Jointly Sponsored by Institute of Remote Sensing Applications, Chinese Academy of Sciences, and Beijing Normal University, Beijing, 100101, P.R. China (gong@nature.berkeley.edu).

Photogrammetric Engineering & Remote Sensing
Vol. 73, No. 8, August 2007, pp. 955–965.

0099-1112/07/7308–0955/$3.00/0

data set or the ratio of between-class variance and within-class variance of the training samples. Jia (1996) and Jia and Richards (1999) proposed a segmented principal component (PC) transformation to reduce the computation cost by selecting subsets of the covariance matrix in a lower segmented dimension. Penalized discriminant analysis (PDA) was suggested to more efficiently deal with the high correlation among the bands by penalizing the high within-class variance and to improve the performance of LDA (Yu *et al.*, 1999). Jimenez and Landgrebe (1999) segmented and transformed the whole spectrum into sub-spectra, estimated training statistics at the sub-spaces, iteratively updated an orthogonal projection matrix until getting a minimum Bhattacharyya distance (BD) among the classes.

Study Site, Data Description, and Georeferencing

Our study area is the city of Fremont, part of the Silicon Valley, located on the southeast side of the San Francisco Bay in California. Fremont is a city of over 208,000 people with an area of 235 km^2, making it the fourth most populous city in the Bay Area and California's fifth largest city in area (The City Government of Fremont, California, 2004). Figure 1 contains a EO-1 Hyperion image taken on 17 January 2001. The diverse landscapes in the study area range from the hilly regions in the north, the urban area in the middle, and the salt evaporators in the southern end of the San Francisco Bay. The urban part includes old residential areas (A), indicating that larger and taller trees are present; new residential areas (B) mainly in the north, covered by new and smaller trees with wider streets and roads; city parks and school lawns (C); industrial and commercial areas (D) in the northwest and south parts appearing bright due to the high reflectance of the concrete or sand cover occupying a relatively larger space; Highway 880 (E) running through the heart of the urban region from the northwest to the southeast; and the Quarry Lake (F) that appears dark, couching down south of the mountainous regional park in the north.

Figure 1. EO-1 Hyperion imagery of our study site in Fremont, California taken on 17 January 2001. (A) Old residential; (B) New residential; (C) Park; (D) Industrial; (E) Highway 880; (F) Quarry Lake.

An overview of instruments Landsat ETM+, EO-1, ALI, and Hyperion and their science applications were presented (Ungar *et al.*, 2003). The fourth track of the ALI data geographically overlaps with the long narrow Hyperion scene (256 pixels in width and 6,478 lines in length). Both images have a spatial resolution of 30 m. The panchromatic (PAN) scene covering the same area has a ground sampling distance of 10 m. The ALI, Hyperion, and PAN images were cut to cover only the study site (Figure 1). The Hyperion and the ALI data were geographically registered by carefully choosing 13 ground control points (GCPs) at a root mean squared error (RMSE) of 0.20 and 0.11 pixels in the vertical and the horizontal directions. They were further geometrically corrected with respect to two quadrangles of 7.5-minute topographic maps. Nine GCPs were used and the RMSE of the georeferencing were 0.50 pixels and 0.83 lines. The PAN band of the imagery was georeferenced with respect to three corresponding orthorectified aerial photographs at 1 m spatial resolution from the USGS Digital Orthophoto Quadrangle (DOQ) database. Fourteen GCPs were collected and the RMSE was 0.38 pixels and 0.41 lines. The PAN imagery was geo-matched with the Hyperion and the ALI data. The final georeferenced ALI and Hyperion images have 310 by 550 pixels while the PAN image is 930 by 1,650 pixels. The cubic convolution method was adopted for image resampling. The PAN image has been used along with the orthophotos and the topographic maps as references.

The Hyperion data has 242 bands (22 bands with overlapped wavelengths) with 12-bit quantization and a spectral range covering 400 nm to 2,500 nm wavelengths. There were 38 empty bands and 43 noisy bands that were removed, so that 161 bands were used for analysis. Those removed bands were mostly water vapor absorption bands. The bands of the ALI were designed in such a way that the central wavelengths of six bands matched those of the Landsat Thematic Mapper (TM).

Hierarchical Classification Scheme

A land-use and land-cover classification system with remotely sensed data has been developed by the USGS and adopted since the 1970s (Anderson *et al.*, 1976; Roessnner, 2001). A hierarchical classification scheme was adopted to make it possible and easier to merge and separate land-use classes based on different scales of studies (Table 1). In this research we developed a two level system for detailed land-use and land-cover mapping (level I and level II). Level I and level II are in an increasing order of detail. Field visits were made twice in March 2002. Selection of separate training and test samples was guided by the property description for each class and validated in the field. The number of training and test samples is also listed in Table 1.

TABLE 1. HIERARCHICAL CLASSIFICATION SCHEME WITH TWO MAPPING LEVELS AND THE NUMBER OF TRAINING TEST SAMPLES AND DESCRIPTION FOR EACH LAND-COVER AND LAND-USE CLASS OF LEVEL II MAPPING

Level I		Level II		Train #	Test #	Description
1	meadow land	1	meadow land	198	162	meadow land on the hill slope, moisture content varies with seasons, rather dry in the summer, green in the winter
2	lawn	2	wetland lawn	125	92	well trimmed with higher water content
		3	lawn	138	93	such as school lawns, frequently walked and interfered by humans, lower moisture
3	oak woodland	4	oak woodland	205	206	trees grown in the valley, mainly oaks
4	residential	5	old residential	187	157	appear dark red on the false color display of images, covered by houses, larger and more trees
		6	new residential	125	75	appear relatively lighter red on the false color display of images, covered by houses, smaller and less trees
5	dry grass/bare soil	7	dry grass/bare soil	131	105	perannual dry grass or bare soil
6	swamp	8	swamp	173	191	marshland embedded with grass and shrubs, appear brown in the false color display of images, along the edges of salt evaporators
7	lake	9	mud	102	108	mud shell, between the edges of lakes and swamps
		10	lake water	182	149	clear lake water
		11	mixed lake/ salt evaporator	163	144	appear dark blue and navy blue, in between clear lake water and salt evaporator, more salt concentration comparing to clear lake
8	salt evaporator	12	salt evaporator 1	215	161	appear light green, different stage of salt evaporation
		13	salt evaporator 2	187	174	appear bright green
		14	salt evaporator 3	168	144	appear dark green
9	commercial/ industrial	15	commercial	160	89	relatively smaller areas along the streets
		16	industrial	239	153	relatively larger areas and more concentratedly distributed
10	transportation	17	high way	177	132	high way 880, 84
		18	road	162	117	roads and streets in the city

While these classes are the common land-use and land-cover categories found in the study area, several classes were generally ignored in previous research. These include the two level II classes, road and highway. Previous research using satellite data from similar spatial resolution rarely distinguish them from the other urban classes. We included them in this research in order to test the capability of HS data in distinguishing them from the other urban land-use classes.

Classification Methods

Correlation Matrix

The high correlation among adjacent spectral bands creates a large amount of redundant information and causes the estimated covariance or correlation matrices with respect to continuous bands to degenerate and become unreliable. Let us denote the digital number of a pixel as $dn_{i,n}$ for band i and pixel n, $i = 1, \ldots, I$; $n = 1, \ldots, N$; I and N is the total number of bands and total number of pixels in an imagery, respectively. The vector representation for pixel n is $dn_n = (dn_{1,n}, \ldots, dn_{I,n})^T$. The band mean vector is written as $\mu = (\mu_1, \ldots \mu_i \ldots, \mu_I)^T$, where μ_i is the mean digital number of band i. The total covariance of the image is represented by $\sum_T = \frac{1}{N}\sum_{n=1}^{N}\left(dn_n - \mu\right)\left(dn_n - \mu\right)^T$.

Figure 2a and 2b show the band correlation matrix of the ALI and the Hyperion data. The diagonal line indicates the highest correlation, 1, which is represented in white. The darker the tone, the lower is the absolute value of the correlation. We can see that the contiguous bands along the diagonal line appear "in blocks" showing high correlation among them, especially for the Hyperion data. The correlation matrix of the Hyperion data looks quite similar to that of the airborne visible and near-infrared imaging spectrometer (AVIRIS) data (Jia, 1996). There are six major blocks along the diagonal line, with separation lines between blocks at band 1, 13, 28, 49, 85, 125, and 161 corresponding to wavelengths at 437, 560, 712, 925, 1,346, 1,810, and 2,365 nm. They are mostly absorption bands by water or oxygen.

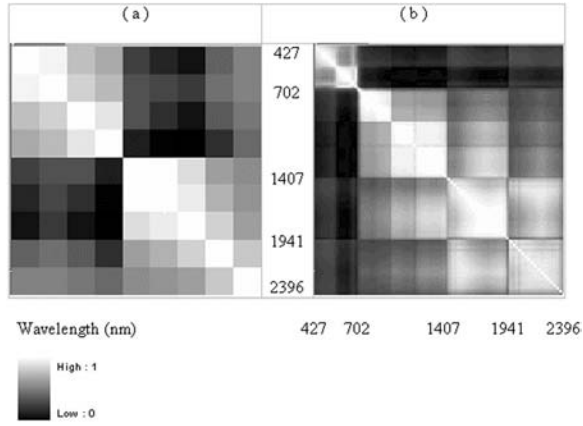

Figure 2. Band correlation matrix (in absolute values) presented in an image form: (a) ALI data, and (b) Hyperion data. Separation wavelengths between sub-blocks of the Hyperion correlation data are labeled. The diagonal line indicates the highest correlation, 1, which is represented in white. The darker the tone, the lower is the absolute value of the correlation.

Mahalanobis Distance Classifier

The estimates of training samples for the mean vector and within-class covariance matrix are denoted by $\mu(c)$ and Σ_W, $c = 1, \ldots, C$; where C is the total number of classes. Here, $\Sigma_W = \sum_{c=1}^{C} \pi_c \cdot \frac{1}{N_c} \sum_{n=1}^{N_c} \left(dn_n - \mu(c)\right)\left(dn_n - \mu(c)\right)^T$ where π_c is the proportion of class c samples in the total samples and Nc is the sample size of class c. In this study, we used Mahalanobis distance classifier for both the original ALI and the Hyperion datasets, and all the subsequent feature reduced images. A pixel dn_n will be classified into class c if the distance $\left(dn_n - \mu(c)\right)^T \Sigma_W^{-1} (dn_n - \mu(c))$ is minimized. The dn_n and $\mu(c)$ can be either obtained from the original image or from the transformed images by feature reduction techniques.

Feature Reduction Methods

PCA and Segmented PCA

To deal with the problem of low ratio of the number of training samples over the spectral bands and to reduce the computational cost, PC transformation is used to project the original data into a new orthogonal space with successive variance maximized. The result is that the few major components contain most of the variances in the original dataset so that further analysis can be pursued based on the few major components. The SEGPCA adopts the idea of the segmented PC transformation proposed by Jia and Richards (1999) in such a way that the processing operates on subgroups of the original data rather than on the full set of data as does the conventional PCA. Thus, it further reduces the computation time and mitigates the small training problem (Jia, 1996; Jia and Richards, 1999). Other efforts have been made on the generation of more reliable statistics in a lower dimensional space (Jimenez and Landgrebe, 1999; Jimenez *et al.*, 1999). In this study, to maintain the computation simplicity and efficiency, we did not select features according to the criterion of BD but rather choosing one or two major components from the six blocks according to three criteria: the visual inspection, order of variance explanations, and eigenvalues. We calculated the variance-covariance matrix from the entire image rather than using the training samples. Therefore, both PCA and the SEGPCA were kept in an unsupervised fashion. It is intuitively meaningful that we select a few major components from each highly correlated block to avoid choosing most of the major components from only one or two blocks.

LDA and Segmented LDA

Fisher's LDA searches for successive linear combinations of data to maximize the ratio of between-class variance over within-class variance of training samples in an expectation of spreading the means or the cluster centers of different classes as much as possible while keeping the within-class variation at a similar level for all classes (Yu *et al.*, 1999). It is based on an assumption of reliable estimation of training statistics. Let us denote the overall mean vector of training samples by $\mu_0 = \sum_{c=1}^{C} \pi_c \cdot \mu(c)$. The between-class covariance matrix is defined as $\Sigma_B = \sum_{c=1}^{C} \pi_c \cdot \left(\mu(c) - \mu_0\right)\left(\mu(c) - \mu_0\right)^T$. LDA looks for an I by p projection matrix $B_{LDA} = (\beta_1, \ldots, \beta_p)$ whose j^{th} column β_j is a transformation vector for producing the particular j^{th} component image. The number of total components p is determined by the intrinsic dimension of Σ_B corresponding to the number of non-zero eigenvalues, which is less than or equal to *C-1*. The ratio for the newly projected images $g = \frac{\sigma_B^2}{\sigma_W^2}$, where $\sigma_B^2 = \sum_{c=1}^{C} \pi_c \cdot \left(\beta^T \mu(c) - \beta^T \mu_0\right)^2 = \beta^T \Sigma_B \beta$, $\sigma_W^2 = \beta^T \Sigma_W \beta$ by definition. Differentiating g with respect to β and letting it be *0* yield $\Sigma_W^{-1} \Sigma_B \beta = g \cdot \beta$, where βs are the eigenvectors corresponding to the p non-zero eigenvalues of $\Sigma_W^{-1}\Sigma_B$ and maximizing g.

Segmented LDA first divides the whole spectrum into sub-blocks with each block containing a set of continuous highly correlated spectral bands as determined from the correlation matrix mentioned earlier (Figure 3). Denote the dimension of the k^{th} sub-block as I_k, and $I_1 + \cdots + I_k + \cdots + I_K = I$. For each sub-block of spectral bands, estimate the between-class covariance matrix and the within-class covariance matrix in a sub-space that has the dimension equal to the number of

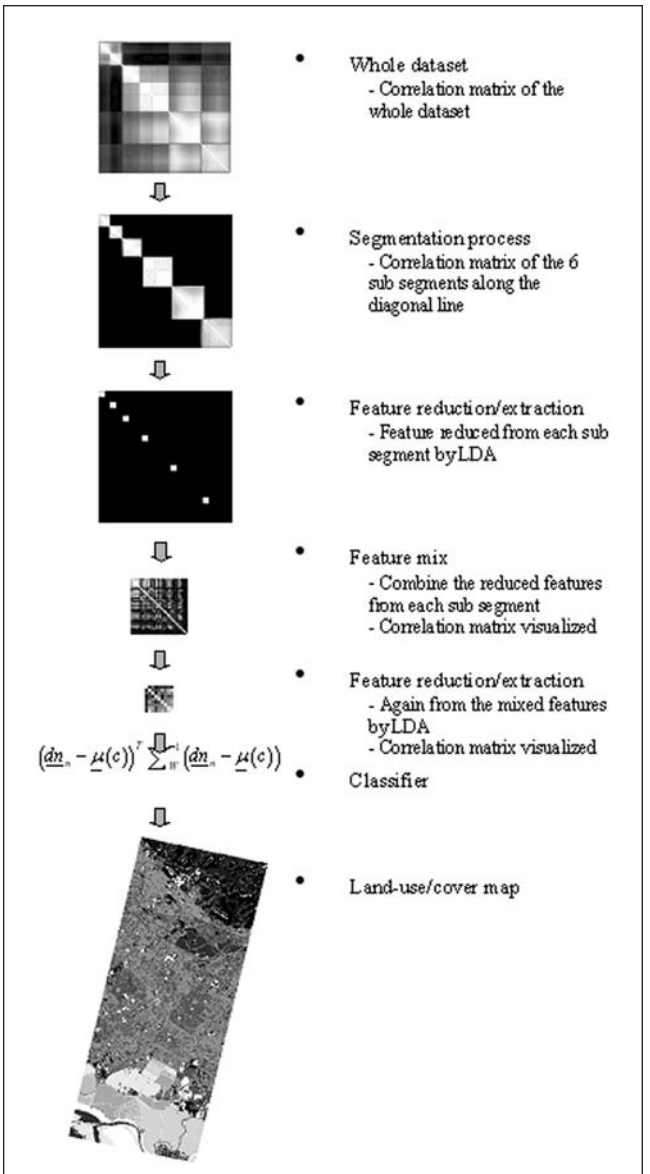

Figure 3. Flow chart of a Segmented LDA.

TABLE 2. KAPPA ACCURACY AND SIGNIFICANCE TEST FOR THE ALI AND THE HYPERION IMAGE CLASSIFICATION FOR BOTH LEVEL I AND LEVEL II CLASSES WITH DIFFERENT FEATURE REDUCTION ALGORITHMS

				Level I					Level II				
				MS		HS			MS		HS		
	Scenario	λ	Criterion	Kappa	Kappa sd	Kappa	Kappa sd	Z-value	Kappa	Kappa sd	Kappa	Kappa sd	Z-value
MD				0.929	0.00555	0.924	0.00569	0.63	0.855	0.00732	0.868	0.00703	1.28
PCA						0.905					0.834		
SEGPCA			1			0.924					0.848		
			2			0.927					0.847		
			3			0.921					0.843		
LDA				0.928	0.00556	0.935	0.00530	0.91	0.855	0.00733	0.880	0.00677	2.51
SEGLDA	1		1			0.891					0.822		
	1		2			0.883					0.825		
	1		3			0.888					0.828		
	2					0.932					0.876		
PDA		0.1		0.917		0.936			0.844		0.882		
		1		**0.930**	0.00551	**0.937**	0.00525	0.92	**0.857**	0.00730	0.884		
		10		0.923		**0.937**	0.00525		0.849		0.880		
		20				0.936					**0.885**	0.00664	2.84
		50				0.934					0.884		
		100				0.933					0.876		
SEGPDA	1	20	3			0.893					0.836		
	2					0.931					0.871		

bands in the sub-block. Then, apply LDA to each sub-block to generate new component images (features) (with a number of *min(C-1, I_k)*). This projection is supposed to spread the means of the classes as much as possible. With the newly projected images for each sub-block, we could either select the first few feature images for each sub-block to generate a combined pool of new features that can be subsequently used for classification (Scenario 1 in Table 2) or select more feature images (less than *min(C-1, I_k)*) from the k^{th} sub-block for $k = 1, \ldots, K$ to form a new sub-space (Scenario 2 in Table 2). For Scenario 2, we apply LDA once again to search for an optimal set of orthogonal sub-space for use in final MD classification.

Similar to the SEGPCA, SEGLDA significantly saves computation time. To calculate the symmetric within-class covariance matrix, the conventional LDA needs to fill in $O_{LDA} = I^*(I + 1)/2$ entries while the SEGLDA only needs to fill in $O_{segLDA} = \sum_{k=1}^{K} \frac{I_k(I_k + I)}{2}$ entries. For example, if $I = 200$, $K = 4$, $I_k = 50$ for $k = 1, \ldots, K$, then $O_{LDA} = 20100$ and $O_{segLDA} = 5100$. To apply LDA, the conventional method has to invert the matrix and to calculate the eigenvalues and eigenvectors at the full dimension while the SEGLDA completes these tasks at the sub-spaces. To do the image transformation, the conventional LDA must calculate at least I^*m multiplications and $(I-1)^*m$ additions where m is the number of features finally selected for classification while the SEGLDA only needs to calculate $\frac{m}{K} \cdot \sum_{k=1}^{K} I_k$ multiplications and $\frac{m}{K} \cdot \sum_{k=1}^{K} (I_k - 1)$ additions if the number of bands for each subblock is equally distributed. For example, if $m = 8$, then $I^*m = 1600$, $\frac{m}{K} \cdot \sum_{k=1}^{K} I_k = 400$. On the other hand, the SEGLDA avoids the estimation of the covariance at the full dimensional space. Instead it is applied to a small group of bands, thus making it easier to estimate reliable parameters from limited number of training statistics.

PDA and Segmented PDA

The PDA introduces a penalty matrix Ω to the within-class covariance matrix to penalize and limit the effect that a band with high within-class variation may have in the case of LDA but to reserve the low within-class variation band in the meantime. The function of the penalty matrix was geometrically interpreted (Yu *et al.*, 1999). It unequally smoothes the within-class variation for all the classes in the hyperspectral space. The realization of SEGPDA is similar to that of the SEGLDA in a sense of the segmentation process before applying PDA, except that it adds a penalty term to the estimation of the within-class covariance matrix.

Experiments, Results, and Discussion

To answer the question "Does the HS image contain more information than the MS image for urban land use mapping?," we tested classification on both the original images and the above feature extracted images. The spectral curve of the mean digital number for each of the 18 level II classes was plotted for both the ALI and the Hyperion data (Figure 4). We did not apply any atmospheric correction to the data because they will not affect classification performance (Song *et al.*, 2000). Therefore, we directly applied Mahalanobis distance classifier to both of the original ALI image (nine bands) and the Hyperion image (161 bands). The kappa coefficient from the nine MS bands for the level II mapping was 0.855 and 0.868 from the 161 HS bands (Table 2). The difference between the two images was only 1.3 percent. When we collapsed the fine and detailed level II classes by merging them into their corresponding coarser level I classes, the kappa coefficient for the MS bands was 0.929 and 0.924 for the HS bands. The HS image did not gain more from collapsing the detailed classes than did the MS image. To statistically test the significance of the accuracy differences, we conducted a t-test between the kappa coefficients of each pair of compared classification results. A Z-value can be determined by dividing the difference between two Kappa coefficients resulting from two classifications by the square-root of the sum of the corresponding Kappa variances (Gong and Howarth, 1990b). A Z-value above 2.37 indicates

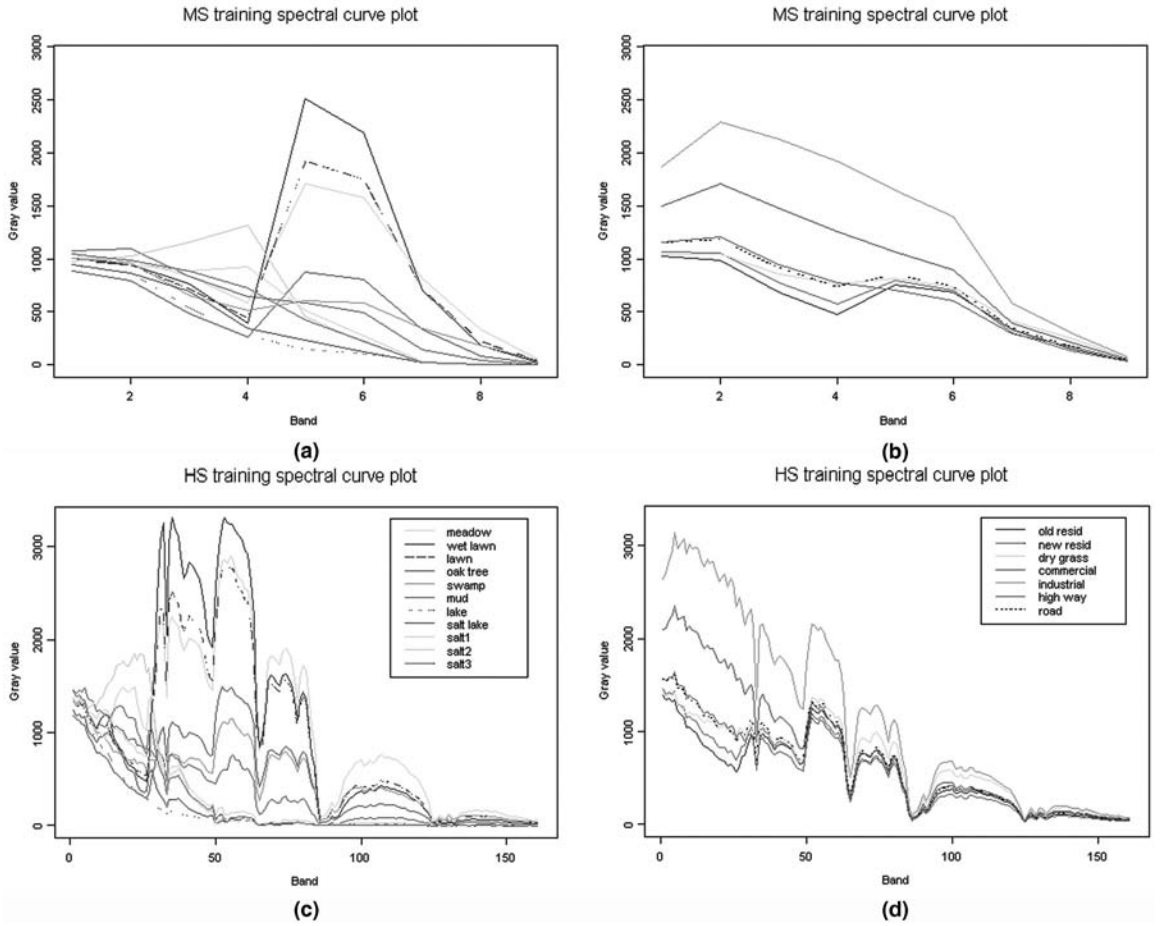

Figure 4. Spectral plot of the mean digital numbers of the 18 training classes: (a) ALI data, land-cover classes, (b) ALI data, land-use classes, (c) Hyperion data, land-cover classes and (d) Hyperion data, land-use classes.

that the two classification results are significantly different at the 0.99 confidence level. The resulting Z-values of 1.28 and 0.63 for the level II and level I classes indicated that there was no significant difference between the two images on the urban land-use classification.

Regarding the second question in feature reduction, the original Hyperion data was first projected into a transformed space by maximizing either the data variance (i.e., PCA) in an unsupervised fashion or separability among different classes (i.e., LDA, and PDA) in a supervised fashion (through the use of training samples). The newly projected component images were then both visually inspected and computationally tested for classification. We found that the first seven to nine features contain most of the variation of the data. Therefore, they were used for further classification. Due to the fact that most of the HS space was redundant, the useful information can be kept in a low dimensional space. The intrinsic dimensionality was generally less than ten (Harsanyi and Chang, 1994). Therefore, we finally reduced the dimension of the Hyperion data from 161 to 9, from which we could achieve the best accuracy and this coincidentally matched the dimension of the ALI space.

For the segmentation-based feature reductions of the Hyperion data, we divided the spectrum into six sub-blocks according to the visual inspection result of the correlation matrix as stated earlier (Table 3). We tested the blocking

Table 3. Data Segmentation Listing Block Number, Band Range, Wavelength Range, and Three Criteria for Feature Selection Using the segpca

Sub-block	Band	Wavelength(nm)	Criterion 1	PC1(%)	PC2(%)	Criterion 2	ρ1(10e6)	ρ2(10e5)	Criterion 3
1	1–13	437–560	1	97.50	1.5	1,2	0.78	0.12	1
2	14–28	570–712	2	95.80	2.2	3,4	1.32	0.31	2,3
3	29–49	722–925	3,4	98.20	1.4	5	3.70	0.52	4,5
4	50–85	962–1346	5,6	99.00	0.5	6	9.10	0.41	6,7
5	86–125	1417–1810	7,8	98.00	1.2	7	0.84	0.10	8
6	126–161	1971–2365	9	95.50	1.1	8,9	0.06	0.01	9

behavior of the correlation matrix obtained from two other study sites in Patagonia, Argentina and found that the separation lines at wavelength of 437, 560, 712, 925, 1,346, 1,810, and 2,365 nm surprisingly representative. Clearly, these separation lines are not scene dependent and therefore they may be applicable to other scenes without preexamination. For the SEGPCA, we generated a data covariance matrix from each sub-block and transformed each sub-block of images into new component images that maximized the variances. In other words, we applied PCA to each sub-block of bands rather than applying PCA to the whole dataset. Then, we selected one or two features from each segment of component images according to the three criteria listed earlier (Table 3). Criterion 1 is based on visual inspection. Criterion 2 is based on the consideration of choosing component images that could cover the data variation from each sub-group as much as possible. Criterion 3 was based on the ranking of the weighting contributions of the component images. The results from Table 2 showed that there was no much difference among these three criteria selections. Comparing the PCA with the SEGPCA, there was a 1 percent increase in the kappa accuracy using the SEGPCA for level II mapping and a 2 percent increase for level I mapping.

Utilizing the same strategy of feature selection as SEGPCA, we obtained a kappa coefficient of 0.822 to 0.828 for level II classification employing Scenario 1 for the SEGLDA. The results were 5 percent to 6 percent lower for level II and 4 percent to 5 percent lower for level I than those from the conventional LDA although it saved about five-sixths of the computation time. This was attributed to the fact that we ignored the high correlations among these selected features from each sub-block. The combination of the high separability features from each sub-group might not lead to high separability for the whole dataset. On the contrary, they might contain less information as it was possible that the high separability images from each subgroup were along similar directions, and they were highly correlated thus ignoring other useful features. Based on this observation we constructed a second experiment (Scenario 2) in which a reasonable number of features from each sub-group (it was five in this experiment as the remaining features contain mostly noise) was chosen. Those selected features were then combined to form a new sub-group (5*6 = 30 features) for a second round of LDA resulting in nine final features that have the maximum separability among the classes. In this case, we achieved a kappa accuracy of 0.876 that was at the same level as a conventional LDA while still greatly reducing the computation of a conventional LDA.

To penalize high local variation, we considered a second derivative-type penalty matrix Ω_D, defined as $\lambda D_K^T D_{K-1}^T D_{K-1} D_K$, where D_K denoted a k–1 by k dimensional first difference operator matrix and λ was the smoothing parameter (Yu *et al.*, 1999). Our within-class covariance matrix Σ_W was perturbed by adding this penalty matrix in the PDA. We let the smoothing parameter λ to vary at 0.1, 1, 10, 20, 50, and 100. The best classification accuracy we could achieve with PDA was 0.885 with $\lambda = 20$ for the level II mapping and 0.937 with $\lambda = 1$ *or 10* for the level I class mapping. Applying SEGPDA in a similar fashion as SEGPCA, we obtained a kappa value of 0.836 for level II classes when we selected the features based on the ranking of eigenvalues (Scenario 1, criterion 3). By pooling features from each subgroup and forming a new set of features for final SEGPDA (Scenario 2), we improved the kappa value to 0.871 for level II classes and 0.931 for level I classes. No significant difference existed between the results from the SEGPDA and the PDA (Z-value of 1.45 and 0.79). The choice of the penalty matrix can be critical, and further pursuit to the knowledge of different spectral behavior among different bands will make the penalty matrix to be more adaptable to the data and would possibly improve the classification performance.

Among the six feature reduction/extraction methods, SEGPDA, PDA, SEGLDA, and LDA obtained similar classification accuracy, although a PDA-based approach consistently achieved a slightly higher (less than 1 percent for both level I and level II classes) accuracy than an LDA-based approach. They all outperformed the SEGPCA and the PCA by 3 percent to 5 percent for level II classes and 1 percent for level I classes.

For the ALI imagery, we also tested LDA and PDA in an attempt to further reduce and extract features to improve the MD classification. However, it seemed that there was no room left for further improvements as a kappa value of 0.857 and 0.930 for level II, and level I classification was the best we could obtain comparing to 0.855 and 0.929 by directly using a MD classifier. When we compared the best kappa values obtained from the Hyperion data for both level II and level I mapping (0.885 and 0.937, respectively) with that obtained from the ALI data (0.857 and 0.930, respectively), we found that there was less than 3 percent difference for the level II classification and no difference for the level I classification at a Z-value of 2.84 (indicating a significant improvement) and 0.92, respectively. The resulting maps were shown in Plate 1.

Is it worthwhile to make this effort of using the HS data instead of the MS data in the classification of urban land uses? For which specific class does the HS perform better than the MS? When the conditional kappa coefficients of the best classification results using the ALI (PDA, $\lambda = 1$) and the Hyperion (PDA, $\lambda = 20$) data are plotted for each individual level II class (Figure 5), it can be seen that for various water class including swamp, mud, clear lake, mixed lake/salt evaporator, and three different staged salt evaporators, both the HS and the MS images satisfactorily (perfectly for the HS) classify them with kappa values close to 1. For meadow land, wetland lawn, lawn, and oak woodland that are mainly vegetation, the HS data perform obviously better than the MS data. For the level II impervious land-use classes such as commercial, industrial, highway, and concrete road, the HS image classification outperforms the MS image, particularly on the industrial. It implies us that hyperspectral imagery possibly not only have strength on identifying subtle spectral changes of vegetation, but also on mapping urban impervious surface. For road, none of the two images appeared to produce a satisfactory result. This can be explained by the high spectral confusion between road and several other classes such as highway, new residential, old residential (Figure 4). However, the MS classifies better with dry grass/bare soil and especially new residential, where less vegetation coverage in the neighborhood comparing to the old residential.

Two confusion matrices arranged in three main blocks along the diagonal entries are presented for the best MS (Table 4) and HS (Table 5) classification results. The category number is listed on the first row and the first column; row entries are references and column entries are classified results. The first block contains all vegetation classes. Both images achieved good results, and the HS image improves the MS classification by 7 percent on average. The second block contains all water-related and relatively pure cover classes. Both the images produce almost perfect classification results. The third block contains all human induced land-use classes, and they are mostly spectrally mixed classes. Generally, both the images achieved poorer results

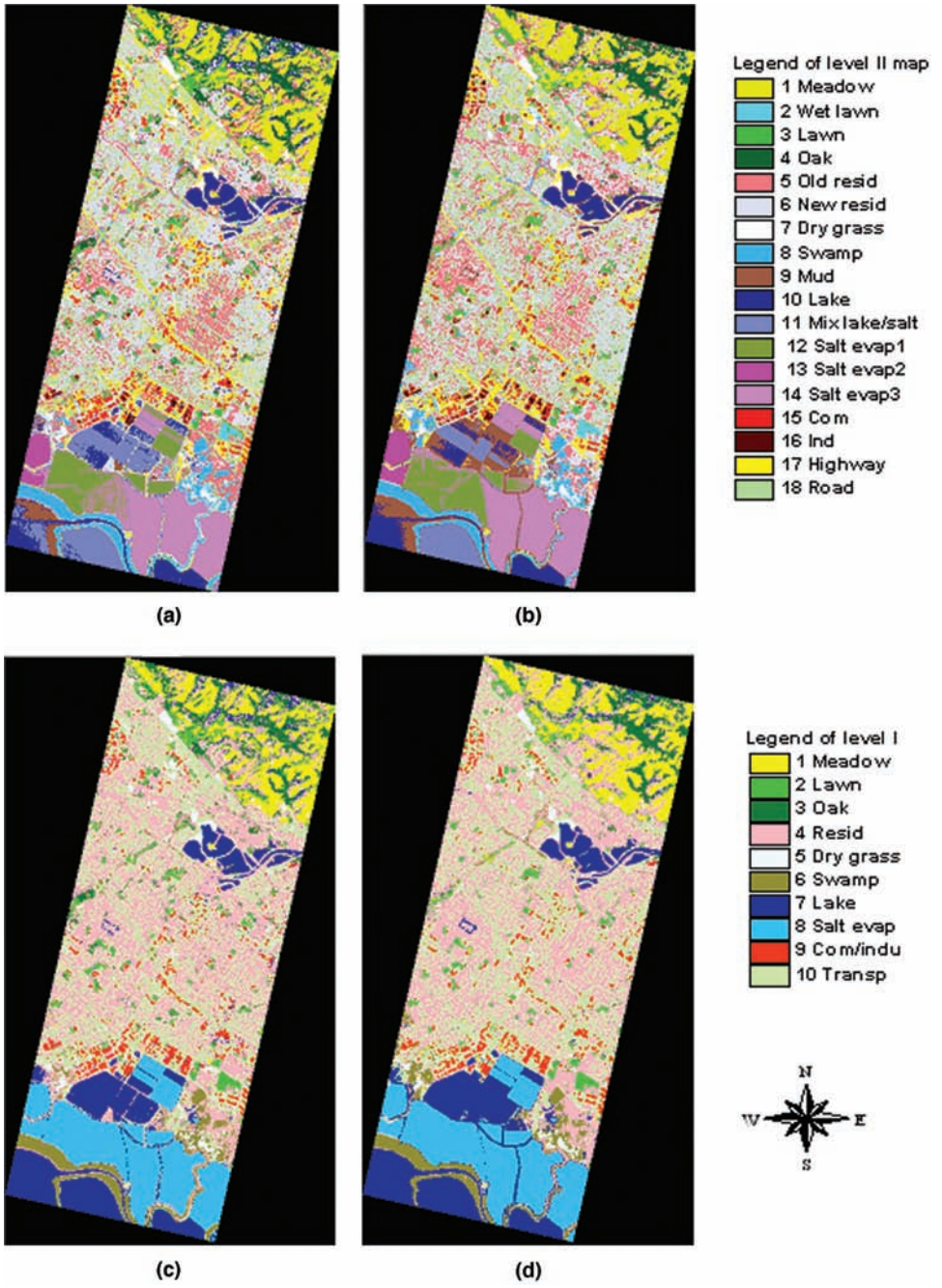

Plate 1. The best level II classification map using (a) the ALI data, and (b) the Hyperion data. The best level I classification map using (c) the ALI data, and (d) the Hyperion data.

with these classes comparing to the other land-cover classes. There is a great confusion among old residential, new residential, and the roads, as is between commercial and industrial, and between highway and road. The HS improves the classification accuracy by 5 percent on average for this third block. The left column of Figure 4 depicts spectral curves from land-cover classes listed in the first, and the second block and the right column depicts those from the land-use classes as listed in the third block. The use of hyperspectral data on vegetation type identification has been intensively investigated in the past, however, less explored on urban mapping. This research evaluates the capability of satellite-borne hyperspectral sensor on mapping urban environment, particularly the potential of mapping urban impervious surfaces.

What else can be done to further improve the overall HS mapping efficiency and/or accuracy? From a statistical point of view, the existing feature reduction methods could

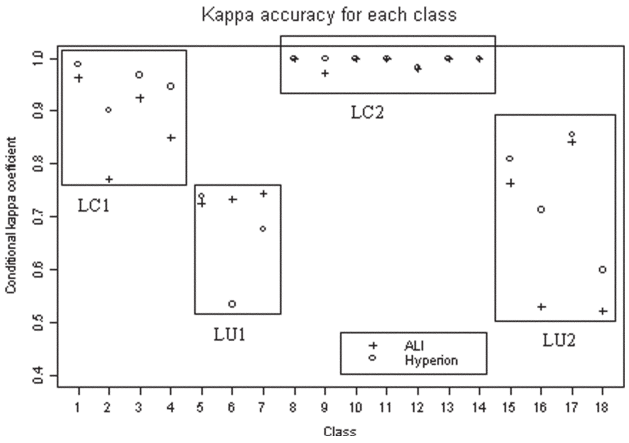

Figure 5. The conditional kappa coefficients of individual classes for the best level II classification using the ALI data and the Hyperion data. LC1 (land-cover type 1): vegetation class; LC2 (land-cover type 2): water class; LU1 (land-use type 1): residential and soil; LU2 (land-use type 2): impervious surface.

be adjusted. For example, the Noise Adjusted PCA or MNF could be further explored in an absence of the *a priori* information about the noise structure; it could be applied not only for rare target detection but also for an exhaustive image classification (Green *et al.*, 1988; Lee *et al.*, 1990; Chang and Du, 1999). For the PDA, the penalty matrix may be adjusted to fit the local band variation for a better performance. A fast and optimal feature selection process should also be pursued. It may be beneficial to use spatial information in the HS imagery to further improve the classification. All these can be further investigated.

Conclusions

With the satellite-borne hyperspectral sensor, Hyperion, recently coming into existence, we investigated ways to make better use of those abundant resources of spectral data to improve the land-cover and land-use mapping in an urban environment. Based on our exploratory analysis, we conclude:

1. There is no significant improvement (less than 1 percent in kappa coefficient) of classification accuracy when comparing Hyperion data (161 bands) with the ALI data (nine bands) if a Mahalanobis distance classifier is directly applied. However, the best feature reduction and extraction image (nine bands) from the original Hyperion data improves the ALI (nine bands) classification by 3 percent for level II classes (significant on the *t*-test) and less than 1 percent (insignificant on the *t*-test) for level I classes.
2. Feature reduction algorithms applied to the ALI data do not improve classification accuracy compared with the original ALI image. Feature reduction does not seem necessary for the ALI image.
3. Among different feature reduction algorithms applied to the Hyperion data, a segmentation-based approach basically improves the computation efficiency over a conventional one. There are no significant differences in terms of classification accuracy among SEGPDA, SEGLDA, PDA, and LDA although the PDA-based algorithms consistently showed a less than 1 percent improvement over the LDA-based algorithms. All these four supervised feature reduction methods outperform PCA by 5 percent (level II) and 3 percent (level I), SEGPCA by 4 percent (level II) and 1 percent (level I) and the direct use of the MD classifier by 2 percent (level II) and 1 percent (level I).
4. For the individual classes, the best feature reduction image from the HS data generally produced better results than the MS data except for new residential and dry grass. For pure land-cover (block 2 of the confusion matrices) classes, both images produced close to perfect results. For the vegetation classes (block 1) and the relatively mixed spectral classes (block 3), the HS improved the kappa accuracy by less than 3 percent. The HS particularly showed strength on classifying vegetation and urban impervious surface.
5. By collapsing the level II to the level I classes, we gained about 5 percent accuracy for the ALI image and 7 percent for the feature reduced Hyperion image.

TABLE 4. CONFUSION MATRIX FROM THE BEST ALI CLASSIFICATION. BLOCK 1: CLASS TYPE 1 TO 4 ARE VEGETATION; BLOCK 2: CLASS TYPE 8 TO 14 ARE WATER; BLOCK 3: CLASS TYPE 5 TO 7 AND 15 TO 18 ARE PERVIOUS AND IMPERVIOUS LAND-USE CATEGORIES, RESPECTIVELY

		Classification																		
	ALI	1	2	3	4	8	9	10	11	12	13	14	5	6	7	15	16	17	18	
Reference	1	156		6																162
	2		71	21																92
	3	5		86	1										1					93
	4			3	175		1	12						15						206
	8					191														191
	9						105	3												108
	10							149												149
	11								144											144
	12									158		3								161
	13										174									174
	14											144								144
	5				2	1							114	33					7	157
	6												6	55				4	10	75
	7														78			1	26	105
	15														1	68	7	12	1	89
	16					2	1								3	60	81	6		153
	17												1	2	4	1		111	13	132
	18					2							3	18				33	61	117
		161	71	116	178	196	107	164	144	158	174	147	124	123	87	129	88	167	118	2452

TABLE 5. CONFUSION MATRIX FROM THE BEST HYPERION CLASSIFICATION. BLOCK 1: CLASS TYPE 1 TO 4 ARE VEGETATION; BLOCK 2: CLASS TYPE 8 TO 14 ARE WATER; BLOCK 3: CLASS TYPE 5 TO 7 AND 15 TO 18 ARE PERVIOUS AND IMPERVIOUS LAND-USE CATEGORIES, RESPECTIVELY

		Classification																	
	Hyperion	1	2	3	4	8	9	10	11	12	13	14	5	6	7	15	16	17	18	
Reference	1	160		2																162
	2		83	9																92
	3	2		90										1						93
	4				195		3							8						206
	8					191														191
	9						108													108
	10							149												149
	11								144											144
	12									158		3								161
	13										174									174
	14											144								144
	5												116	30				1	10	157
	6												4	40					31	75
	7					15								16	71				3	105
	15															72	14	3		89
	16															37	109	7		153
	17												2			1		113	16	132
	18												7	24	2			14	70	117
		162	83	101	195	206	111	149	144	158	174	147	129	119	73	110	123	138	130	2452

Acknowledgments

Financial support from NASA's EO-1 Science Validation program through grant (NCC5-492) and a grant from Chinese Academy of Sciences from its "one hundred people" program.

References

Anderson, J.R., E.E. Hardy, J.T. Roach, and R.E. Witmer, 1976. *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*, U.S. Geological Survey Professional Paper 964, Washington, D.C. 20402.

Chang, C.-I., and Q. Du, 1999. Interference and noise-adjusted principal components analysis, *IEEE Transactions on Geoscience and Remote Sensing*, 37(5, Part 2):2387–2396.

Deguchi, C., and S. Sugio, 1994. Estimations for percentage of impervious area by the use of satellite remote-sensing imagery, *Water Science and Technology* 29(1–2):135–144.

Goetz, A.F.H., G. Vane, J.E. Solomon, and B.N. Rock, 1985. Imaging spectrometry for earth remote sensing, *Science*, 228(4704): 1147–1153.

Gong, P., and P.J. Howarth, 1990a, The use of structural information for improving land-cover classification accuracies at the rural-urban fringe, *Photogrammetric Engineering & Remote Sensing*, 56(1):67–73.

Gong, P., and P.J. Howarth, 1990b, An assessment of some factors influencing multispectral land-cover classification, *Photogrammetric Engineering & Remote Sensing*, 56(5):597–603.

Gong, P., and P.J. Howarth, 1992, Frequency-based contextual classification and gray-level vector reduction for land-use identification, *Photogrammetric Engineering & Remote Sensing*, 58(4):423–437.

Gong, P., R. Pu, and B. Yu, 1997. Conifer species recognition: An exploratory analysis of in situ hyperspectral data, *Remote Sensing of Environment*, 62(2):189–200.

Gong, P., R. Pu, and B. Yu, 2001. Conifer species recognition: Effects of data transformation, *International Journal of Remote Sensing*, 22(17):3471–3481.

Gong, P., R. Pu, and R.C. Heald, 2002. Analysis of in situ hyperspectral data for nutrient estimation of giant sequoia, *International Journal of Remote Sensing*, 23(9):1827–1850.

Gonzalez, R.C., and P. Wintz, 1987, *Digital Image Processing*, Second Edition, Addison-Wesley Publishing Company, Reading, Massachusetts.

Green, A.A., M. Berman, P. Switzer, and M.D. Craig, 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal, *IEEE Transactions on Geoscience and Remote Sensing*, 26(1):65–74.

Harsanyi, J.C., and C.-I. Chang, 1994. Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach, *IEEE Transactions on Geoscience and Remote Sensing*, 32(4):779–785.

Hastie, T., A. Buja, and R. Tibshirani, 1995. Penalized discriminant analysis, *Annals of Statistics*, 23(1):73–102.

Hsu, P.-H., T. Yi-Hsing, and P. Gong, 2002. Dimension reduction of hyperspectral images for classification applications, *Geographic Information Sciences*, 8(1):1–8.

Hughes, G.F., 1968. On mean accuracy of statistical pattern recognizers, *IEEE Transactions on Information Theory*, IT14 (1):55–63.

Jia, X., 1996. *Classification Techniques for Hyperspectral Remote Sensing Image Data*, Ph.D. dissertation, School of Electrical Engineering, University College, ADFA, University of New South Wales.

Jia, X., and J.A. Richards, 1999. Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification, *IEEE Transactions on Geoscience and Remote Sensing*, 37(1, Part 2):538–542.

Jimenez, L.O., and D.A. Landgrebe, 1999. Hyperspectral data analysis and supervised feature reduction via projection pursuit, *IEEE Transactions on Geoscience and Remote Sensing*, 37(6): 2653–2667.

Jimenez, L.O., A. Morales-Morell, and A. Creus, 1999. Classification of hyperdimensional data based on feature and decision fusion approaches using projection pursuit, majority voting, and neural networks, *IEEE Transactions on Geoscience and Remote Sensing*, 37(3, Part 1):1360–1366.

Lee, J.B., A.S. Woodyatt, and M. Berman, 1990. Enhancement of high spectral resolution remote sensing data by a noise-adjusted principal components transform, *IEEE Transactions on Geoscience and Remote Sensing*, 28(3):295–304.

NASA New Millenium Program (NMP), 1996. URL: http://eol.gsfc.nasa.gov (last date accessed: 21 May 2007).

Roessner, S., K. Segl, U. Heiden, and H. Kaufmann, 2001. Automated differentiation of urban surfaces based on airborne hyperspectral imagery, *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1525–1532.

Ridd, M.K., 1995. Exploring a V-I-S (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote-sensing - Comparative anatomy for cities, *International Journal of Remote Sensing*, 16(12):2165–2185.

Song, C., C.E. Woodcock, K.C. Seto, M.P. Lenney, and S.A. Macomber, 2000. Classification and change detection using Landsat TM data: When and how to correct atmospheric effects?, *Remote Sensing of Environment*, 75(2):230–244.

The City Government of Fremont, California, 2004. URL: http://www.ci.fremont.ca.us (last date accessed: 21 May 2007).

Thenkabail, P.S., E.A. Enclona, M.S. Ashton, C. Legg, and M.J. De Dieu, 2004. Hyperion, IKONOS, ALI and ETM+ sensors in the study of African rainforests, *Remote Sensing of Environment*, 90:23–43.

Ungar, S.G., J.S. Pearlman, J.A. Mendenhall, and D. Reuter, 2003. Overview of the Earth observing one (EO-1) mission, *IEEE Transactions on Geoscience and Remote Sensing*, 41(6): 1149–1159.

Xu, B., P. Gong, E. Seto, and R. Spear, 2003. Comparison of gray-level reduction and different texture spectrum encoding methods for land-use classification using a panchromatic IKONOS image, *Photogrammetric Engineering & Remote Sensing*, 69(5):529–536.

Yu, B., M. Ostland, P. Gong, and R. Pu, 1999. Penalized linear discriminant analysis for conifer species recognition, *IEEE Transactions on Geoscience and Remote Sensing*, 37(5): 2569–2577.

(Received 18 August 2004; accepted 19 August 2004; revised 19 March 2007)