

**Spectrally segmented principal component analysis of hyperspectral imagery for mapping invasive plant species**

F. TSAI, E.-K. LIN and K. YOSHINO

This is a preprint of an article submitted for consideration in the INTERNATIONAL JOURNAL OF REMOTE SENSING © [accepted in June 2006] [copyright Taylor & Francis]; INTERNATIONAL JOURNAL OF REMOTE SENSING is available online at: <http://journalonline.tandf.co.uk/>

*For internal use and private communication only. Please do not distribute.*

# Spectrally segmented principal component analysis of hyperspectral imagery for mapping invasive plant species

F. TSAI\*†, E.-K. LIN† and K. YOSHINO‡

†Center for Space and Remote Sensing Research, National Central University, Zhong-Li, Taoyuan 320, Taiwan (ROC)

‡Institute of Policy and Planning Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan

## Abstract

Principal component analysis (PCA) is one of the most commonly adopted feature reduction techniques in remote sensing image analysis. However, it may overlook subtle but useful information if directly applied to the analysis of hyperspectral data, especially for the discrimination among different vegetation types. In order to accurately map an invasive plant species (horse tamarind, *Leucaena leucocephala*) in southern Taiwan using Hyperion hyperspectral imagery, this study developed a spectrally segmented principal component analysis based on spectral characteristics of vegetation over different wavelength regions. The developed algorithm can not only reduce the dimensionality of hyperspectral imagery but also extract helpful information for differentiating the target plant species from other vegetation types more effectively. Experiments conducted in this study demonstrated that the developed algorithm performs better than correlation-based segmented principal component transformation (SPCT) and conventional PCA (Overall Accuracy: 86%, 76%, 66%; Kappa value: 0.81, 0.69, 0.57) in detecting the target plant species as well as mapping other vegetation covers.

**Keywords:** hyperspectral data; Hyperion; PCA; vegetation; mapping; spectral segmentation

## 1. Introduction

The intrusion of nonnative vegetation, especially those with aggressively invasive capability, have caused significant impact to biodiversity and ecosystem balances world-wide. In order to develop effective remedy strategies to battle against this threatening situation, it is necessary to understand the accurate spatial distribution and progression of invading alien species. Conventionally, this task relies heavily on field-based investigations. These methods are usually expensive and time-consuming. Furthermore, they can only provide limited information from sparse points of collection. Remote sensing, on the other hand, is an effective and economical alternative for gathering spatially distributed data over a large area. Aerial photographs and satellite images comprised of a single band or multiple spectral channels are common data sources for vegetation-related studies. Researchers have reported successful applications of these remotely sensed data to vegetation mapping and related applications (Franklin et al., 2000; Katoh, 2004; McCormick, 1999).

As sensors and associated technologies evolve, new types of remote sensing data have emerged and extended the horizons of remote sensing applications. The availability of hyperspectral remote sensing data provides researchers and scientists an opportunity to pursue more complex applications that are difficult, if not impossible, to accomplish with

---

\* Corresponding author; Fuan Tsai, Center for Space and Remote Sensing Research, National Central University, 300 Zhong-Da Rd., Zhong-Li, Taoyuan 320, Taiwan; e-mail: ftsai@csrsr.ncu.edu.tw

traditional multispectral data. With tens to hundreds of contiguous spectral bands and high spectral resolution, hyperspectral imagery can deliver an abundance of spectral information and has a great potential to achieve species-level plant target identification (Cochrane, 2000; Hirano et al., 2003; Underwood et al., 2003; Xiao et al., 2004). Nevertheless, the inherent differences between hyperspectral data and traditional multispectral images also bring up new challenges to the processing and analysis and sometimes may pose extraordinary obstacles to the success of various applications.

One of the most substantial difficulties in hyperspectral data processing and analysis has to do with the massively larger data volume and high data dimensionality. The two factors require excessively more computational resources to deal with and longer processing time. More importantly, they may result in the failure or poor performance of analysis algorithms. Therefore, data compression or feature reduction is often considered a practical necessity in hyperspectral data analysis. An obvious solution to this problem is to select only a few suitable bands from original hyperspectral data and continue to use traditional multispectral methods (Price, 1997). This might work in certain circumstances, but it does not take full advantage of hyperspectral data. Also, to obtain an optimal subset of spectral bands will require a complete examination on all feature combinations. Since the number of feature combinations increases exponentially, it will be next to impossible to derive the best band selection from a hyperspectral data set.

A better approach is to reduce the data dimensionality while trying to maintain the most vital and useful information in the data set. A few algorithms have been developed for this purpose, including band moments (Staenz, 1996), endmember selection (Acito et al., 2002), minimum noise fraction (MNF) transform (Green et al., 1988), discrete wavelet transform (Bruce et al., 2002) and derivative spectral analysis (Laba et al., 2005; Tsai & Philpot, 1998; Tsai & Philpot, 2002). Although varying in concepts, these novel methods all aimed to tackle a critical issue, i.e. to extract the most useful information from a hyperspectral data set in order to reconstruct a feature collection with a minimum number of bands to represent most of the desired data characteristics so the data dimensionality can be reduced.

Among various feature reduction methods, principal component analysis (PCA) is a commonly adopted technique to reduce the dimensionality of remote sensing data. However, Cheriadat and Bruce (2003) argued that despite its effectiveness in data compression, PCA might not necessarily be an optimal method for feature extraction of hyperspectral data, especially for target detection and supervised classification applications. Their primary reason was that PCA might fail to extract useful information for separating targets from certain data distributions because higher order principal components do not always retain desired distinct features. As a result, they suggested using other feature extraction methods for better performance in target identification and supervised classification with hyperspectral data.

Even so, PCA should still have advantages over other (more sophisticated) feature reduction techniques. In particular, PCA is simple, straightforward, easy to use, and most importantly, has been implemented in almost all available remote sensing image processing and analysis packages. It is possible to use PCA as the fundamental framework to develop an appropriate feature extraction system that is capable of collecting information most helpful to the discrimination of target classes. The purpose of this study was to develop a spectrally segmented principal component analysis system specifically designed for identifying a predetermined plant species from hyperspectral remote sensing imagery. The objective was to

correctly map an invasive plant species (horse tamarind, *Leucaena leucocephala*) that is spreading at an alarming rate in the Kenting National Park and vicinity located in the Hen-Chun peninsula of southern Taiwan (Chiang & Hsu, 2000).

## 2. Segmented PCA for plant target detection

Principal component analysis is a linear transformation that projects data onto a new orthogonal feature space in a way that the first few components in the new feature space will represent most of the variances in the original data set. The transformation is based on the second order statistics (covariance) of the original data. Assume  $\mathbf{x}$  represents the vector of a pixel's gray value in an image of  $N$  bands. The transformation is defined as

$$\mathbf{y} = \mathbf{A}^t \mathbf{x} \quad (1)$$

where  $\mathbf{A}$  is the matrix of normalized eigenvectors of the image covariance matrix, and  $t$  denotes the transpose operation of the matrix. The image covariance matrix,  $\mathbf{C}$ , is an  $N$ -by- $N$  matrix and can be constructed according to all pixels,  $\mathbf{x}_i$ ,  $i=1, 2, \dots, K$  and the mean vector  $\mathbf{m}$  as described in Eq-2 and Eq-3.

$$\mathbf{C} = E[(\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^t] = (K-1)^{-1} \sum_{i=1}^K (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^t \quad (2)$$

$$\mathbf{m} = E\{\mathbf{x}\} = \frac{1}{K} \sum_{i=1}^K \mathbf{x}_i \quad (3)$$

There are two tasks in a PCA. The first is an eigen-analysis to generate the transformation matrix  $\mathbf{A}$ ; and the second is the linear transformation for each pixel to project data onto the new orthogonal space,  $\mathbf{y}$ . More details about PCA are described in remote sensing image analysis texts (e.g. Richards & Jia, 1999) and are not repeated here.

If applied to conventional multispectral data, where spectral bands are few and discrete, PCA is efficient and usually yields satisfactory outcomes in extracting useful features. For hyperspectral imagery, the data distribution tends to be a hyper-ellipsoid and produces large local (within-class) variances (Lee & Landgrebe, 1993). Directly applying PCA to the entire data set of a high dimensional hyperspectral image may generate biased results in addition to the exponentially increased computational load and processing time. For instance, hyperspectral remote sensing images usually exhibit higher variances in the short wavelength bands; thus PCA will be dominated by visible and near-infrared bands. Also, because PCA operates on global statistics, it may overlook local variances that are helpful to the detection of targets. This shortcoming can be alleviated with an algorithm taking local variances into account, such as the canonical analysis that also transforms data to a new feature space but will maximize the ratio of among-class variance to within-class variance (Richards & Jia, 1999). However, canonical analysis usually requires a large number of training data in order to obtain reliable estimates of class variances, which is unfortunately hardly the case in most remote sensing applications.

Accordingly, it is worthwhile to further improve PCA for a better use with hyperspectral data, whether it is to accelerate the processing or for better feature extraction. A few researchers have worked on this track and proposed modified PCA algorithms. One way to improve PCA performance is to supplement it with other analytical techniques. An example of such augmented PCA systems was a concurrent spectral-screening algorithm that put together principal component transformation, spectral angle classification and human-

centered colour mapping for hyperspectral image fusion (Achalakul and Taylor, 2000). Another approach to enhance PCA is to further refine or extend the PCA algorithm and procedure. For example, a 'generalized principal component analysis' (GPCA) was constructed to deal with motion segmentation in computer vision and other data clustering problems (Vidal et al., 2003; Vidal et al., 2004).

Jia and Richards (1999) presented a segmented principal component transformation (SPCT) procedure to extract the best principal components determined from pairwise separability measures of Bhattacharyya distance for colour display and classification of hyperspectral images. They used a threshold (of 0.5) to identify high correlation bands of a hyperspectral image and examined the correlation matrix to search for edges in order to determine how to divide spectral bands into groups. This approach is useful and effective for general classification purpose. However, if the goal is to detect a specific plant type, it might not be an efficient way to segment spectral bands.

For plant target detection and classification with hyperspectral imagery, if PCA is applied to the entire data set, variances among vegetation and non-vegetation covered pixels will dominate the eigen-analysis. As a result, conventional PCA will most likely extract features for separating plants from other landcover categories instead of discriminating a specific plant species from other vegetation types. To eliminate or at least minimize these unwanted features, non-vegetation pixels should be excluded from the PCA operation. This concept is illustrated in Figure-1, which shows an example of three plant clusters and a non-vegetation class in two-dimensional scatter plots and principal component transformed space. In Figure-1a, the first principal component (PC) band accentuates the distinctions between vegetation and non-vegetation clusters. After the removal of non-vegetation data, PCA is more likely able to differentiate among different plants (Figure-1b).

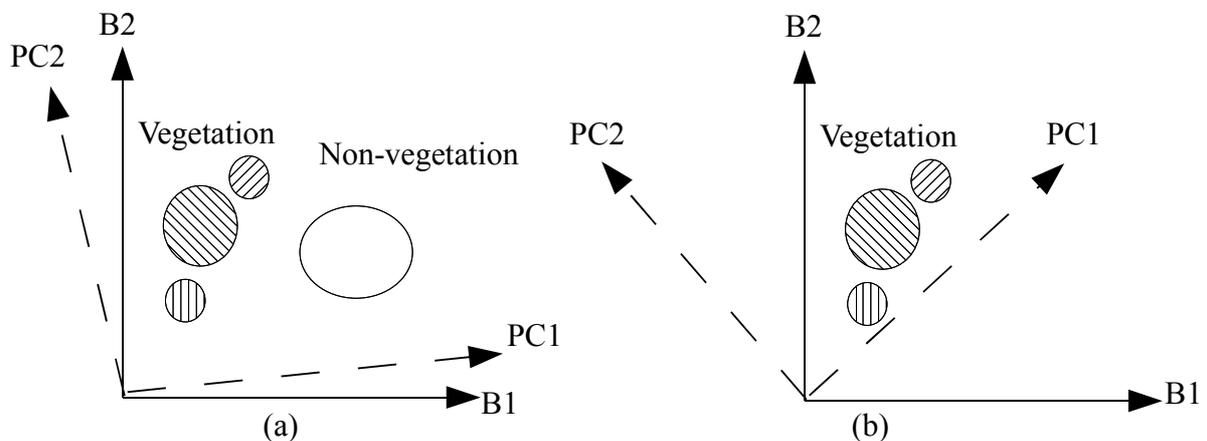


Figure 1: PCA for vegetation differentiation. (a) PCA is dominated by the difference between vegetation and non-vegetation; (b) PCA highlights distinctions among vegetation clusters in PC1.

The difficulty is that subtle features that are useful for distinguishing a specific plant species from other vegetation types may still be neglected and pushed back to lower order PC bands after the transformation. In this regard, it is of particular value to enhance the PCA algorithm for more effective and efficient extraction of implicit but distinctive information

from vegetation spectral data. Bell and Baranoski (2004) developed a piecewise PCA (PPCA) algorithm to analyze field-collected spectral data sets of vegetation. They divided collected spectra into several groups of different wavelength regions and applied PCA to each group independently. The basis of this spectral segmentation approach is that plants interact with solar radiation differently in absorption, reflection and transmittance over different regions of wavelength because of the constituents and structures of their leaves. The experiments of Bell and Baranoski on the reconstruction of reflectance and transmittance curves for plant specimens indicated that (conventional) PCA worked effectively in visible to near-infrared regions and PPCA further improved the efficiency and accuracy-cost ratio in data compression and restoration of plant spectra.

Inspired by these achievements, this study adapted the spectrally segmented PCA approach to reinforce conventional PCA for plant target detection with hyperspectral remote sensing imagery. The system developed in this study was a modification of SPCT (Jia and Richards, 1999) described above, but specifically designed for distinguishing among different vegetation types. Instead of conducting an exhaustive examination on the correlation matrix to divide bands into several groups, the spectral segmentation was determined based on known characteristics of plant leaves over different regions of wavelength. This approach should provide a more efficient and effective spectral segmentation scheme in detecting specific plant targets. The rest of this paper describes the developed system and its application to map an invasive plant species using a satellite hyperspectral (Hyperion) image. It also compares the analysis results of spectrally segmented PCA to conventional PCA and SPCT and discusses the effects of different spectral segmentation schemes.

### **3. Material and methodology**

The study site of this research included parts of the Kenting National Park and vicinity of southern Taiwan as shown in Figure-2. Being one of the most important biological preserves in Taiwan, the national park has a rich vegetation population. Besides the target plant species (*Leucaena leucocephala*), there are three other dominant vegetation types within the study site, including Taiwan acacia (*Acacia confusa*), chaste tree (*Vitex negundo*) and grassland (Liu and Chen, 2002). Consequently, these four plant types were selected as classes to classify in this study and labeled as Leucaena, Acacia, Vitex and Grass, respectively.

The primary image used in this study was an EO-1 (Earth Observing-1) Hyperion hyperspectral image acquired in January, 2004. Hyperion is a space-borne imaging spectrometer. It collects 242 spectral bands (220 unique channels) of solar spectrum covering from 355 nm to 2577 nm in wavelength at a nominal ground resolution of 30 m (Beck, 2003; Pearlman et al., 2003; Ungar et al., 2003). The original image data were calibrated for sensor characteristics (level-1R), but not geometrically corrected. In addition to the Hyperion image, other data used in this study (primarily for data correction, training data selection and analysis validation) included a vector layer of Leucaena distribution, 1/5000 photo-maps, aerial photographs, high resolution satellite images and field-collected data (land-use maps and distributions of other vegetation types produced with GPS surveying). The Leucaena distribution map was created in a 1996 survey conducted by a local forestry research institute by means of interactive interpretation of aerial photographs. The map plotted polygons of vegetation areas with different Leucaena population densities. Each polygon was labeled as one of the four Leucaena density levels: L1 (0% - 25%), L2 (25% - 50%), L3 (50% - 75%), and L4 (75% - 100%). Because the map was more than eight years old, it was corrected with

more up-to-date aerial photographs (of year 2002), high resolution satellite images and field-verifications. The (corrected) vector map and field-investigated data were used as primary ground truth in this research for selecting training data and evaluating analysis results.

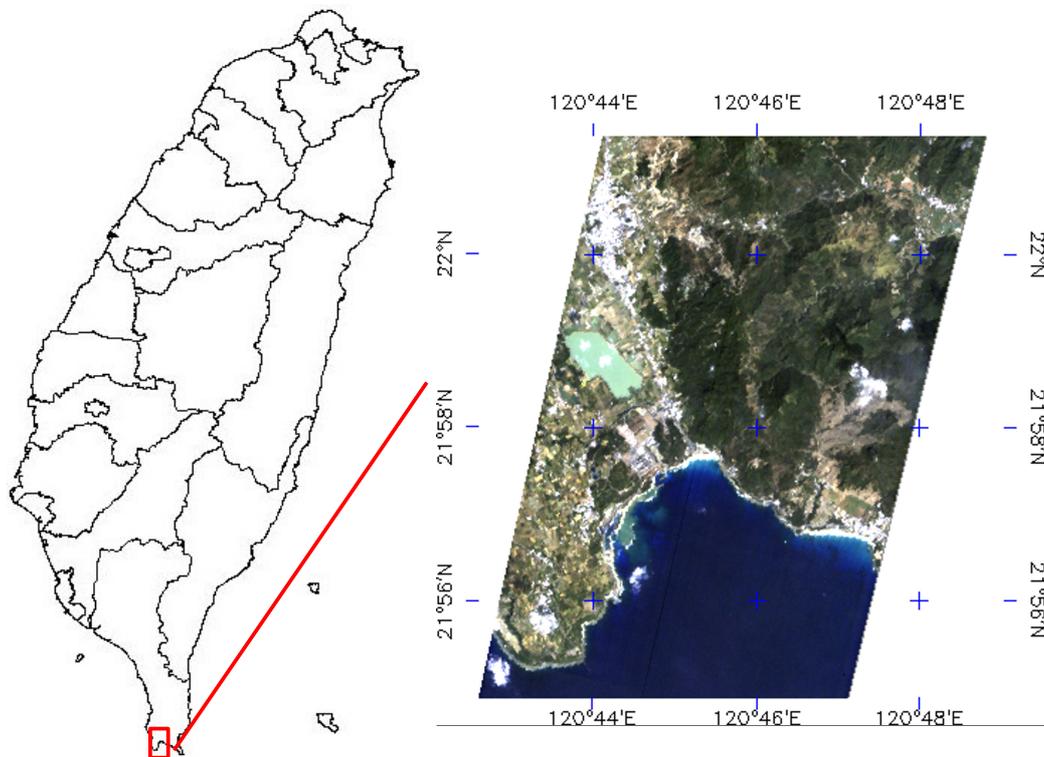


Figure 2: Study site.

The general procedure of spectrally segmented PCA for plant target detection was separated into several phases as illustrated in Figure-3. In preprocessing, the Hyperion image was registered to 1/5000 maps of the study area. The image was atmospherically corrected using the ENVI FLAASH® atmospheric correction module to minimize variations caused by atmosphere, with the built-in tropic atmospheric model and maritime aerosol model as suggested in the ENVI FLAASH User's Guide (Research System Inc., 2004) and EO-1 User Guide (Beck, 2003). In addition, because some of the Hyperion channels recorded no data or were too noisy, they were removed from the image cube. This resulted in a total of 95 bands remained (see Table-1) for analysis. The selection complied with the band selection principles presented by Datt et al. (2003), who proposed a 176-band Hyperion subset to exclude water vapor bands and a 155-band 'stable' subset to further avoid residual atmospheric noise for the application of agricultural indexes. In this study, more noisy bands were excluded, thus using a smaller subset. A sub-image (370 samples by 400 lines) of the geometrically registered and atmospherically corrected image covering the study area was then cropped out as the base image for classification.

After preprocessing, a vegetation index was calculated from selected suitable bands of the image. The index was used to filter out non-vegetation covered areas from the scene. The purpose of this masking process was to prevent subsequent PCA operations from being dominated by variances between plants and non-vegetation objects as explained in the previous section. In this study, band-30 (650 nm) and band-50 (854 nm) were selected to

compute NDVI (Normalized Difference Vegetation Index) and a threshold was set to filter out non-vegetation pixels of the image. If the threshold was too low, non-vegetation covered areas would not be masked, thus degrading the PCA performance in finding distinguishable features among different plants as demonstrated in Figure-1. On the other hand, a high threshold might mistakenly remove some vegetated areas and resulted in incomplete analysis. In order to determine the best threshold, masking results of different NDVI thresholds were compared with landuse maps and aerial photographs of the study area and 0.2 was found to be the most suitable cutoff for separating vegetation and non-vegetation areas in the Hyperion image used in this study.

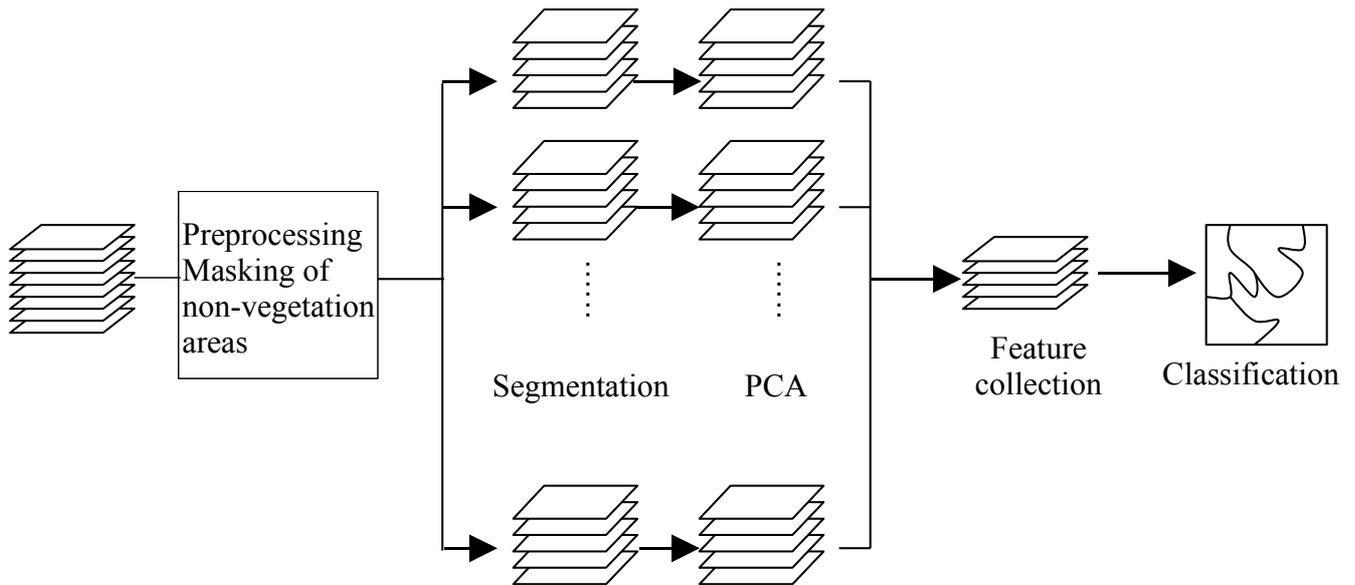


Figure 3: General procedure of spectrally segmented PCA.

Table 1: Selected 95 bands of Hyperion imagery.

<i>Channels</i>	<i>Wavelength region</i>
12 – 55	467.5 – 905.0 nm
83 – 93	927.3 – 1074.0 nm
95 – 96	1094.1 – 1104.2 nm
102 – 115	1164.7 – 1295.9 nm
117 – 118	1316.1 – 1326.1 nm
139 – 150	1537.9 – 1648.9 nm
156 – 160	1709.5 – 1749.8 nm
162 – 163	1770.0 – 1780.1 nm
207 – 209	2224.0 – 2244.2 nm

The ultimate goal of the analysis was to distinguish *Leucaena* from other vegetation types. To achieve this objective, the developed feature extraction function must be able to

discover subtle and helpful characteristics. As indicated earlier, despite all the efforts (atmospheric correction, masking of non-vegetation regions etc.) to minimize the impact of irrelevant variations in the Hyperion image, critical features helpful to separating the target plant species from other vegetation types might still be overlooked in conventional PCA of the 95 spectral bands. Therefore, the PCA procedure was modified in hopes of extracting variations among different plants more effectively.

Previous studies have discovered that biological factors of plant leaves affect the spectral reflectance of vegetation over different ranges of wavelength (Bell & Baranoski, 2004; Chelle et al., 1998). Accordingly, the remaining 95 Hyperion bands were divided into four groups based on spectral characteristics of vegetation over different wavelength sections. The four groups are listed in Table-2 and include: 1.) visible (VIS, 450 nm – 660 nm), where reflectance and transmittance are both relatively low because of high absorption in pigments of leaf tissues, especially chlorophyll (Devlin & Baker, 1971; Woolley, 1971; Zarco-Tejada et al., 2001); 2.) near infrared (NIR, 670 nm – 900 nm), where, as a result of multiple internal reflections in foliage structures, the reflectance of vegetation is strong (Tucker & Garrat, 1977); 3.) short-wavelength infrared-1 (SWIR-1, 970 nm – 1350 nm); and 4.) short-wavelength infrared-2 (SWIR-2, 1530 nm – 2240 nm), where the absorption is predominated by the water content (Gupta & Vijayan, 2001; Jacquemound & Baret, 1990).

Table 2: Four-group segmentation for spectrally segmented PCA.

<i>Group</i>	<i>Wavelength</i>	<i># of bands</i>
VIS	467 – 660 nm	20
NIR	671 – 905 nm	24
SWIR-1	972 – 1073 nm 1094 – 1104 nm 1164 – 1295 nm 1316 – 1326 nm	29
SWIR-2	1537 – 1648 nm 1709 – 1749 nm 1769 – 1780 nm 2224 – 2244 nm	22

For comparison, this study also performed a two-group segmentation (which divided Hyperion data into a visible and near-infrared (VNIR) group and a short-wave infrared (SWIR) group) and a SPCT analysis which resulted in a three-group segmentation. The two-group segmentation scheme is listed in Table-3 and details of the SPCT analysis are described in the next section.

After the spectral segmentation, PCA was applied to each spectral group independently for feature extraction. Based on the eigen analysis, the first few PC bands of each segmented group were collected to formulate a new feature image for a Maximum Likelihood (ML) classification. The four dominant vegetation types in the study site (Leucaena, Acacia, Vitex, Grass) were the four categories to classify. Pixels pertaining to the

four classes were selected from ground truth as training data for the classifier. Finally, the ML classification results were verified and evaluated qualitatively and quantitatively with ground truth and other reference data. The ground truth data were collected through field investigations in areas of homogeneous vegetation coverage located with aerial photograph interpretation and the existing Leucaena map. At least 300 pixels of each class (300 Leucaena, 320 Acacia, 300 Vitex, and 300 Grass pixels) were collected as training samples and the rest were used for validation.

Table 3: Two-group segmentation for spectrally segmented PCA.

<i>Group</i>	<i>Wavelength</i>	<i># of bands</i>
VNIR	467 – 660 nm	44
	671 – 905 nm	
SWIR	972 – 1073 nm	51
	1094 – 1104 nm	
	1164 – 1295 nm	
	1316 – 1326 nm	
	1537 – 1648 nm	
	1709 – 1749 nm	
	1769 – 1780 nm	
	2224 – 2244nm	

#### 4. Results and discussions

The classification result of the best collected PC features in segmented PCA (three PCs in VIS, two PCs in NIR, one PC in SWIR-1 and one PC in SWIR-2) is displayed in Figure-4. For comparison, Figure-5 displays a classification using the first seven PCs generated from conventional PCA. A quick visual comparison of the two figures indicates that conventional PCA features resulted in substantially more unclassified pixels. In addition, there were more pixels classified as Leucaena (the target) in Figure-5. A verification with aerial photographs and high resolution satellite images concluded that much of the chaste tree and grassland classes were misclassified as Leucaena by the conventional PCA analysis.

To further evaluate the performance of spectrally segmented PCA and different spectral segmentation schemes, a series of sixteen experiments were conducted on the same (95 bands) image and training data set but with one-, two-, three- (SPCT) and four-group segmentations and different numbers of PCs. The SPCT (Case16) segmentation scheme was determined by examining the correlation matrix of the 95-band base image. As displayed in the 'correlation image' (Figure-6), the 95 bands were approximately separated into three groups. They were: 1.) VIS (467 – 712 nm, 25 bands); 2.) NIR and SWIR-1 (722 – 1326 nm, 48 bands); and 3.) SWIR-2 (1537 – 2244 nm, 22 bands). Principal components used in SPCT were three PCs for the first group, three PCs for the second group and two PCs for the last group. The primary difference between the four-group segmentation (Table 2) and SPCT partitioning in this study was that correlation-based SPCT combined the NIR region with the

first SWIR into a single spectral segment.

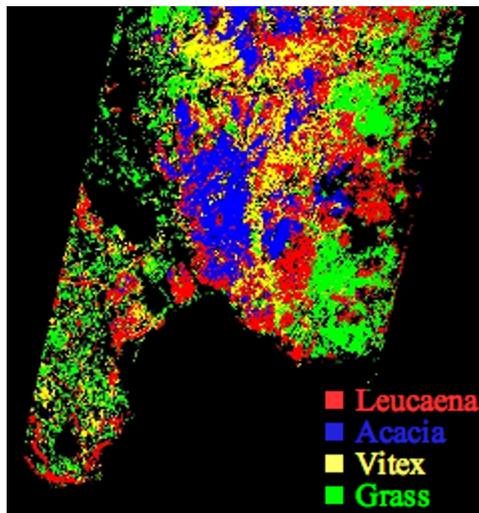


Figure 4: Classification result of a four-group segmented PCA.

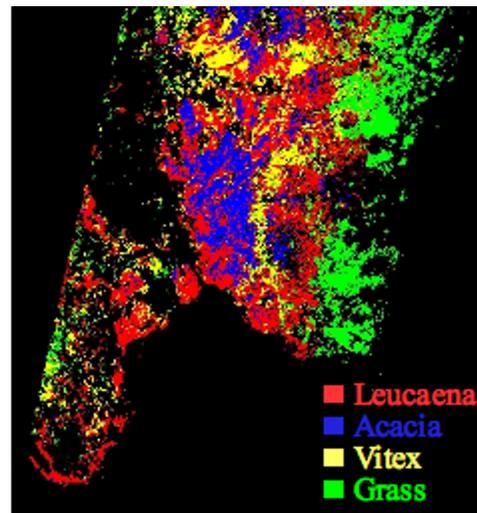


Figure 5: Classification result of a PCA with the first 7 PCs.

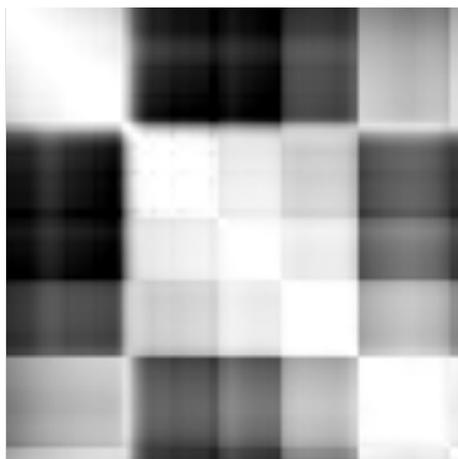


Figure 6: Correlations of the 95-band base image (black: no correlation; white: highest correlation).

The results of all experiments are displayed in Figure-7. It was noticed that in some of the cases, especially those with more PCs (Case3 to Case8, Case14, and Case15), many pixels were classified as Grass in the eastern portion of the study site, which did not conform to the ground truth. Part of this anomaly might be caused by sensor and atmospheric variations not being corrected properly. This could be observed from the first four PCs of the NIR group displayed in Figure-8. The fourth PC in Figure-8 emphasized the spectral 'smile' effect of Hyperion (Datt et al., 2003) instead of real reflectance variances of ground objects. Moreover, the signal-to-noise ratio (SNR) of Hyperion data is relatively low and decreases as wavelength increases. Therefore, when more PCs were included for classification, those

containing data in longer wavelength regions might become too noisy and caused the classifier to generate more false positives (commission errors) for Grass. As a result, Case15 seemed to exhibit the most serious symptom (highest commission error) because it took more PCs in SWIR-1 and SWIR-2 regions into account. Detailed preprocessing methods of Hyperion images were presented in Datt's paper to minimize noise and anomalies, but they were not performed in this study.

Evaluations of the above 16 cases are summarized in Table-4 and error matrices of the best results for each segmentation scheme (Case 8, 11, 13, 16) are listed in Table-5. As listed in Table-4, for conventional PCA (Case1 to Case8), as more PCs were selected, the classification accuracy also increased, but the performance improvement seemed to reach a barrier at 8 to 10 PCs with the best Overall Accuracy (OA) of 66% and 0.57 in the Kappa

value. With conventional PCA, a significant number of Leucaena and Grass pixels were unclassified, resulting in a high omission error. In addition, there seemed to be high misclassification between Leucaena and Acacia pixels during the classification. These also suggested that conventional PCA was not adequate for the differentiation among the four plants.

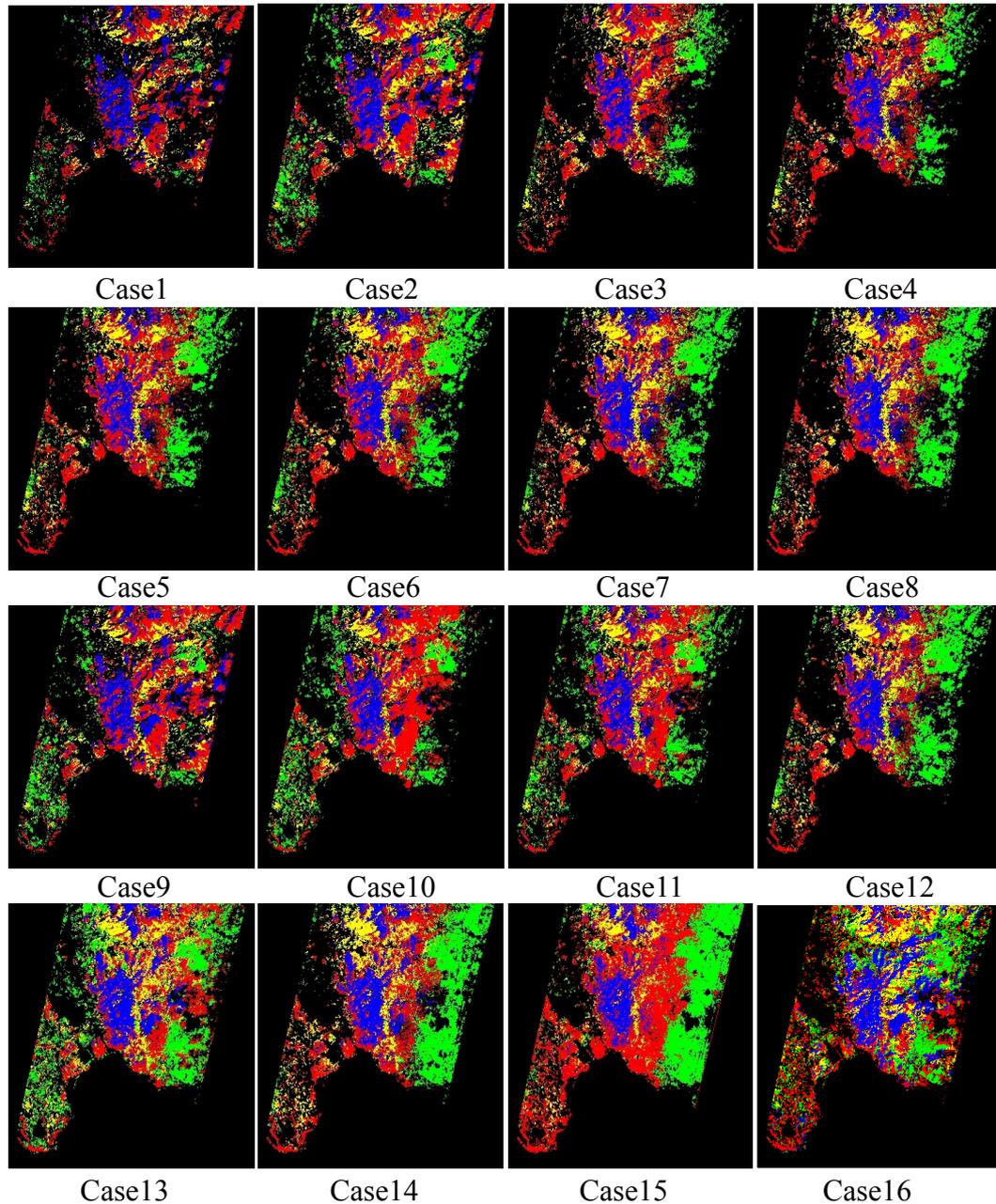


Figure 7: Classification results of all 16 test cases.

Table 4: Evaluations of 16 test cases.

<i>Case</i>	<i>Groups</i>	<i>PCs</i>	<i>Overall Acc.</i>	<i>Kappa</i>
Case1 (PCA)	1	3	42%	0.32
Case2 (PCA)	1	4	55%	0.45
Case3 (PCA)	1	5	55%	0.44
Case4 (PCA)	1	6	60%	0.5
Case5 (PCA)	1	7	62%	0.52
Case6 (PCA)	1	8	66%	0.56
Case7 (PCA)	1	9	65%	0.56
Case8 (PCA)	1	10	66%	0.57
Case9 (SSPCA)	2	2, 2	54%	0.44
Case10 (SSPCA)	2	3, 2	64%	0.55
Case11 (SSPCA)	2	3, 3	69%	0.61
Case12 (SSPCA)	2	4, 4	66%	0.57
Case13 (SSPCA)	4	3, 2, 1, 1	86%	0.81
Case14 (SSPCA)	4	4, 3, 2, 2	72%	0.64
Case15 (SSPCA)	4	4, 4, 4, 4	77%	0.69
Case16 (SPCT)	3	3, 3, 2	76%	0.69

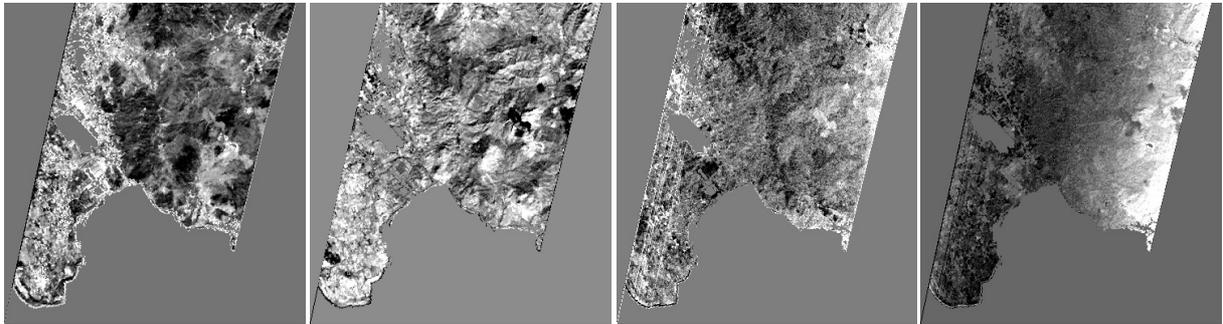


Figure 8: First four PCs (from left to right) in the NIR segment.

As for the segmented PCA, increasing the number of PCs for classification did not necessarily improve the accuracy. Rather, the number and placement of the segmentation appeared to be a more important factor. This could be observed from the results of Case9 to Case15 listed in Table-4. Regardless of the segmentation scheme, the best classification result did not occur in cases with the highest number of PCs (Case12 and Case15). However, in general, four-group segmentation (Case13 to Case15) produced better results than two-group segmentation (Case9 to Case12).

Table 5: Confusion matrices of the best classification results (Case 8, 11, 13, 16) from each segmentation scheme.

<b>Case8</b>		Ground truth					
		<b>Leucaena</b>	<b>Grass</b>	<b>Acacia</b>	<b>Vitex</b>	Total	UA
Classification Result	<b>Unclass.</b>	80	70	0	2	152	
	<b>Leucaena</b>	110	0	88	14	212	0.52
	<b>Grass</b>	31	109	0	0	140	0.78
	<b>Acacia</b>	7	0	224	4	235	0.95
	<b>Vitex</b>	4	29	7	219	259	0.85
	Total	232	208	319	239	998	
	PA	0.47	0.52	0.70	0.92		
Overall Accuracy = 0.66			Kappa = 0.57				

<b>Case11</b>		Ground truth					
		<b>Leucaena</b>	<b>Grass</b>	<b>Acacia</b>	<b>Vitex</b>	Total	UA
Classification Result	<b>Unclass.</b>	83	53	3	5	144	
	<b>Leucaena</b>	144	1	100	30	275	0.52
	<b>Grass</b>	2	136	0	7	145	0.94
	<b>Acacia</b>	2	0	214	2	218	0.98
	<b>Vitex</b>	1	18	2	195	216	0.90
	Total	232	208	319	239	998	
	PA	0.62	0.65	0.67	0.82		
Overall Accuracy = 0.69			Kappa = 0.61				

<b>Case13</b>		Ground truth					
		<b>Leucaena</b>	<b>Grass</b>	<b>Acacia</b>	<b>Vitex</b>	Total	UA
Classification Result	<b>Unclass.</b>	13	15	0	2	30	
	<b>Leucaena</b>	202	0	41	17	260	0.78
	<b>Grass</b>	8	176	0	8	192	0.92
	<b>Acacia</b>	8	0	272	7	287	0.95
	<b>Vitex</b>	1	17	6	205	229	0.90
	Total	232	208	319	239	998	
	PA	0.87	0.85	0.85	0.86		
Overall Accuracy = 0.86			Kappa = 0.81				

<b>Case16</b>		Ground truth					
		<b>Leucaena</b>	<b>Grass</b>	<b>Acacia</b>	<b>Vitex</b>	Total	UA
Classification Result	<b>Unclass.</b>	10	37	0	2	49	
	<b>Leucaena</b>	176	43	54	13	286	0.62
	<b>Grass</b>	31	120	0	11	162	0.74
	<b>Acacia</b>	13	0	257	6	276	0.93
	<b>Vitex</b>	2	8	8	207	225	0.92
	Total	232	208	319	239	998	
	PA	0.76	0.58	0.81	0.87		
Overall Accuracy = 0.76			Kappa = 0.69				

The classification results of two-group segmentation (Case9 to Case12) were similar to conventional PCA. In these cases, a substantial number of Leucaena and Grass pixels were

unclassified (high omission error). Also, there was still difficulty to correctly distinguish some of the Acacia and Leucaena pixels using the selected PCs. This implied that dividing the data into two groups (VNIR and SWIR) did not provide distinct features sufficient for effectively distinguishing the four vegetation classes from each other. On the other hand, in the four-group segmentation, the number of unclassified pixels was reduced significantly, suggesting that the four-group segmentation scheme had indeed provided more useful information. However, it was also noticed that omission errors in Leucaena and Grass increased slightly and some of the Leucaena pixels were misclassified as Grass, when more PC features were added in the NIR and SWIR regions. Part of the reason should have to do with the low SNR in the longer wavelength bands of Hyperion imagery as discussed previously in this section.

Another discovery was that there seemed to be a higher misclassification percentage of Acacia pixels (mainly being incorrectly classified as Leucaena) than the other three types, especially in PCA and two-group segmentation test cases. This suggested that the selected PC features might still not be adequate to clearly separate Acacia from Leucaena. A reasonable explanation for this deflection is that it was common to see (isolated) Acacia trees mixed within Leucaena forests in the study area. Given the broad ground resolution of Hyperion images, it was difficult to find pure pixels of Leucaena to train the classifier and thus caused the higher misclassification rate between these two categories. Four-group segmentation improved the situation. Using hyperspectral data with a better spatial resolution and selecting training data more carefully should be able to further mitigate the misclassification.

Comparing the best of four-group segmentation (Case13) with SPCT (Case16), it was noticed that SPCT seemed to have higher misclassification between Leucaena and Grass as well as between Leucaena and Acacia. The reason might have to do with the issue that SPCT put NIR bands and longer wavelength bands in the first SWIR region into a spectral group because they were highly correlated (as shown in Figure-6). Therefore, correlation-based SPCT may work well in general purpose classification applications. However, when differentiating different vegetation types, the four-group spectral partitioning scheme according to different interactions between plant leaves and solar radiation at different wavelength provided more differentiable features.

## 5. Conclusions

This paper presented a spectrally segmented principal component analysis of hyperspectral imagery for mapping an invasive plant species (*Leucaena leucocephala*) in a study site located in the Heng-Chun peninsula of southern Taiwan. The developed system spectrally segmented Hyperion hyperspectral imagery into different groups and then collected useful PC features in each group for distinguishing the target plant species from three other vegetation types.

Among the 16 cases tested in this study, Case13 produced the best result in detecting the target as well as the other three vegetation categories. This case employed the four-group spectral segmentation scheme and collected three PCs in VIS, two PCs in NIR, one PC in SWIR-1 and one PC in SWIR-2. The Overall Accuracy of Case13 was 86% and the Kappa value was 0.81. Both were better than using SPCT (Overall Accuracy: 76%; Kappa: 0.69) and conventional PCA (Overall Accuracy: 66%; Kappa: 0.57). In addition, the four-group segmentation analysis also generated the best User's Accuracy and Producer's Accuracy for

the classification of the target.

Therefore, in this study, four-group segmentation based on the characteristics of interactions between plant leaves and solar radiation over different wavelength regions performed better in identifying the target than two-group segmentation, correlation-based SPCT segmentation and conventional PCA. The results of this study demonstrated that spectrally segmented PCA with appropriate segmentation scheme and feature selection should be a viable approach in detecting specific plant species with hyperspectral imagery.

### Acknowledgments

The authors would like to thank Dr. Hsiang-Hua Wang and the staff of Heng-Chun Research Center, Taiwan Forestry Research Institute for providing the *Leucaena* map and their assistance in field work. We also thank Prof. Pei-Fen Lee of National Taiwan University for kindly supporting us with valuable images and other data as well as many useful suggestions. This study was supported in part by the National Science Council of Taiwan under project number NSC-94-2752-M-008-004-PAE.

### References

- Achalakul, T. and Taylor, S., 2000, A concurrent spectral-screening PCT algorithm for remote sensing applications, *Information Fusion*, 1(2), pp. 89-97.
- Acito, N., Gorsini, G. and Diani, M., 2002, An unsupervised algorithm for selection of endmembers in hyperspectral images. *Proc. 2002 IEEE Geoscience and Remote Sensing Symposium (IGARSS2002)*, vol. 3, pp. 1673-1675.
- Beck, R. (Ed.), 2003, *EO-1 User Guide*. Available online at: [eo1.usgs.gov/documents/EO1userguidev2pt320030715UC.pdf](http://eo1.usgs.gov/documents/EO1userguidev2pt320030715UC.pdf).
- Bell, I.E. and Baranoski, G.V.G., 2004, Reducing the dimensionality of plant spectral database. *IEEE Trans. Geoscience and Remote Sensing*, 42(3), pp. 570-576.
- Bruce, L.M., Koger, C.K. and Li, J., 2002, Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction. *IEEE Trans. Geoscience and Remote Sensing*, 40(10), pp. 2318-2338.
- Chelle, M., Andrieu, B. and Bouatouch, K., 1998, Nested radiosity for plant canopies. *Vis. Comput.*, 14(3), pp. 109-125.
- Cheriyadat, A. and Bruce, L.M., 2003, Why principal component analysis is not an appropriate feature extraction method for hyperspectral data. *Proc. 2003 IEEE Geoscience and Remote Sensing Symposium*, pp. 3420-3422.
- Chiang, M.-Y. and Hsu, L.-M., 2000, The wildization, impacts, and management of nonnative plants in Taiwan (in Chinese with English abstract), *Proc. 2000 Biodiversity and Preservation Symposium*, Taipei, Taiwan, pp. 399-411.
- Cochrane, M.A., 2000, Using vegetation reflectance variability for species level classification of hyperspectral data. *International J. Remote Sensing*, 21(10), pp. 2075-2087.
- Datt, B., McVicar, T.R., Van Niel, T.G., Jupp, D.L.B., and Pearlman, J.S., 2003, Preprocessing EO-1 Hyperion hyperspectral data to support the application of agricultural indexes. *IEEE Trans. Geoscience and Remote Sensing*, 41(6), pp. 1246-1259.
- Devlin, R.M. and Baker, A.V., 1971. *Photosynthesis* (New York: Van Nostrand Reinhold).
- Franklin, S.E., Hall, R.J., Moskal, L.M., Maudie, A.J., Lavigne, M.B., 2000, Incorporating texture into classification of forest species composition from airborne multispectral images. *International J. Remote Sensing*, 21(1), pp. 61-79.
- Green, A.A., Berman, M., Switzer, P., and Craig, M.D., 1988, A transformation for ordering

- multispectral data in terms of image quality with implications for noise removal. *IEEE Trans. Geoscience and Remote Sensing*, 26(1), pp. 65- 74.
- Gupta, R.K. and Vijayan, D., 2001, New hyperspectral vegetation characterization parameters. *Adv. Space Res.*, vol. 28, pp. 201-206.
- Hirano, A, Madden, M., Welch, R, 2003, Hyperspectral image data for mapping wetland vegetation. *Wetlands*, 23(2), pp. 436-448.
- Jacquemound, S. and Baret, F., 1990, PROSPECT: A model of leaf optical properties spectra. *Remote Sensing Environ.*, 34(2), pp. 75-92.
- Jia, X. and Richards, J.A., 1999, Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification. *IEEE Trans. Geoscience and Remote Sensing*, 37(1), pp. 538-542.
- Katoh, M., 2004, Classifying tree species in a northern mixed forest using high-resolution IKONOS data. *J. Forest Research*, 9(1), pp. 7-14.
- Laba, M., Tsai, F., Ogurcak, D., Smith, S. and Richmond, M.E., 2005, Field determination of optimal dates for the discrimination of invasive wetland plant species using derivative spectral analysis. *Photogrammetric Engineering & Remote Sensing*, 71(5), pp. 603-611.
- Lee, C. and Landgrebe, D.A., 1993, Analyzing high dimensional data. *IEEE Trans. Geoscience Remote Sensing*, 31(4), pp. 792-800.
- Liu, F.-Y. and Chen, M.-A., 2002, The impacts of alien plants to native vegetation ecosystems in the Kenting National Park – an example of *Leucaena leucocephala* (in Chinese), *Conservation Research Report No. 112*, Kenting National Park Administration Office, Construction and Planning Agency, Ministry of Interior, Taiwan (ROC).
- McCormick, C.M., 1999, Mapping exotic vegetation in the Everglades using large-scale aerial photographs. *Photogrammetric Engineering and Remote Sensing*, 65(2), pp. 179-184.
- Pearlman, J.S., Barry, P.S., Segal, C.C., Shepanski, J., Beiso, D. and Carman, S.L., 2003, Hyperion, a space-based imaging spectrometer. *IEEE Trans. Geoscience and Remote Sensing*, 41(6), pp. 1160-1173.
- Price, J.C., 1997, Spectral band selection for visible-near infrared remote sensing: Spectral-spatial resolution tradeoffs. *IEEE Trans. Geoscience Remote Sensing*, 35(5), 1277-1285.
- Research System Inc., 2004, *ENVI FLAASH User's Guide*, version 4.1.
- Richards, J.A. and Jia, X., 1999, *Remote Sensing Digital Image Analysis: An Introduction*, 3rd. Ed., Springer-Verlag.
- Staenz, K., 1996, Classification of a hyperspectral agricultural data set using band moments for reduction of the spectral dimensionality. *Canadian J. Remote Sensing*, 22(3), pp. 248-257.
- Tsai, F. and Philpot, W., 1998, Derivative analysis of hyperspectral data. *Remote Sensing of Environment*, 66(1), pp. 41-51.
- Tsai, F. and Philpot, W., 2002, A Derivative-aided hyperspectral image analysis system for land-cover classification. *IEEE Trans. on Geoscience and Remote Sensing*, 40(2), pp. 416-425.
- Tucker, C.J. and Garrat, M.W., 1977, Leaf optical system modeled as a stochastic process. *Appl. Opt.*, vol. 16, pp. 635-642.
- Underwood, E., Ustin, S. and DiPietro, D., 2003, Mapping nonnative plants using hyperspectral imagery. *Remote Sensing Environ.*, 86(2), pp. 150-161.
- Ungar, S.G., Pearlman, J.S., Mendenhall, J.A. and Reuter, D., 2003, Overview of the Earth

- Observing One (EO-1) mission. *IEEE Trans. Geoscience and Remote Sensing*, 41(6), pp. 1149-1159.
- Vidal, R., Ma, Y. and Sastry, S., 2003, Generalized principal component analysis (GPCA). *Proc. 2003 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'03)*, vol. 1, pp. 621-628.
- Vidal, R., Ma, Y. and Piazzi, J., 2004, A new GPCA algorithm for clustering subspaces by fitting, differentiating and dividing polynomials. *Proc. 2004 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'04)*, vol. 1, pp. 510-517.
- Woolley, J.T., 1971, Reflectance and transmittance of light by leaves. *Plant Physiol.*, vol. 47, pp. 656-662.
- Xiao, Q., Ustin, S.L. and McPherson, E.G., 2004, Using AVIRIS data and multiple-masking techniques to map urban forest tree species. *International J. Remote Sensing*, 25(24), pp. 5637-5654.
- Zarco-Tejada, P.J., Miller, J.R., Mohammed, G.H., Noland, T.L. and Sampson, P.H., 2001, Estimation of chlorophyll fluorescence under natural illumination from hyperspectral data. *International J. Applied Earth Observation and Geoinformation*, vol. 3, pp. 321-327.