Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
ISPRS Journal of Photogrammetry & Remote Sensing 62 (2007) 225–235
PHOTOGRAMMETRY & REMOTE SENSING
www.elsevier.com/locate/isprsjprs

A hyperspectral band selector for plant species discrimination

Chaichoke Vaiphasa [a,*], Andrew K. Skidmore [b], Willem F. de Boer [c], Tanasak Vaiphasa [d]

[a] *Department of Survey Engineering, Chulalongkorn University, Bangkok, Thailand*
[b] *Natural Resources Department, International Institute for Geo-Information Science and Earth Observation (ITC), Enschede, the Netherlands*
[c] *Tropical Nature Conservation and Vertebrate Ecology Group, Wageningen University, Wageningen, the Netherlands*
[d] *Environmental Education Consultant Office, Bangkok, Thailand*

Received 22 July 2006; received in revised form 8 May 2007; accepted 10 May 2007
Available online 14 June 2007

Abstract

The use of genetic search algorithms (GA) as spectral band selectors is popular in the field of remote sensing. Nevertheless, class information that has been used in the existing research for testing the performance of the GA-based band selector is broad (i.e. Anderson's level I or II). This means that each class possesses distinct spectral characteristics from one another, and it is relatively easy for the band selector to find spectral bands that maintain high spectral separability between classes. None of the existing studies has tested the band selector on class information that possesses very similar spectral characteristics (e.g. species-level data). A question therefore remains if the band selector can deal with such complexity. As a result, the key hypothesis of this research is that the GA-based band selector can be used for selecting a meaningful subset of spectral bands that maintains spectral separability between species classes. The testing data in use are very high-dimensional, spectrometer records that comprise 2151 bands of leaf spectra of 16 tropical mangrove species. The results turned out that the GA-based band selector was able to cope with spectral similarity at the species level. It meaningfully selected spectral bands that related to principal physio-chemical properties of plants, and, simultaneously, maintained the separability between species classes at a high level.

Keywords: Artificial_Intelligence; Classification; Hyper spectral; Mangrove; Remote sensing; Vegetation

1. Introduction

Since the first introduction by Holland (1975), many forms of genetic algorithms (GA) have been continuously proposed for remote sensing applications: (i) image segmentation and classification (Tseng and Lai, 1999; Pal et al., 2001; Harvey et al., 2002; Liu et al., 2004; Bandyopadhyay, 2005); (ii) sub-pixel classification (Mertens et al., 2003); (iii) model optimization (Jin and Wang, 2001; Chen, 2003; Fang et al., 2003); (iv) image registration (Jones et al., 2000; Chalermwat et al., 2001); (v) pixel aggregation (Lu and Eriksson, 2000); and (vi) image band selection (Siedlecki and Sklansky, 1989; Lofy and Sklansky, 2001; Kavzoglu and Mather, 2002; Yu et al., 2002; Fang et al., 2003; Kooistra et al., 2003; Luo et al., 2003; Ulfarsson et al., 2003; Cogdill and Rippke, 2004). Ranking by the number of publications, using GA as band selectors is the most popular.

In general, band selectors help alleviate the problem of high-dimensional complexity (Bellman, 1961; Kendall, 1961; Hughes, 1968; Fukunaga, 1990; Shahshahani and Landgrebe, 1994) that usually affects the

* Corresponding author. Tel.: +66 2 2186651; fax: +66 2 2186653.
E-mail addresses: chaichoke@hotmail.com, vaiphasa@itc.nl (C. Vaiphasa).

0924-2716/$ - see front matter © 2007 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.isprsjprs.2007.05.006

outcome of analyzing multiple band data (e.g. multi-sensors, multi-temporal, or hyperspectral images). In most cases, a large number of image bands (i.e. >20 bands) are too complex for familiar parametric tools (e.g. Jeffries–Matusita distance, Bhattacharyya distance, Maximum Likelihood classifier, etc.). Mathematically, the complexity of using such a large number of bands does not only decay the precision of class model estimation of these parametric tools (Bellman, 1961; Hughes, 1968), but it also causes the singularity of covariance matrix inversion (Fukunaga, 1990). Additionally, this high-dimensional complexity results in an excessive demand of field samples in which, in most cases, it is not feasible in practice due to the time and budget limitations (Shahshahani and Landgrebe, 1994).

By comparison, GA-based band selectors perform better than many other popular band selection algorithms (e.g. branch and bound search, exhaustive search, and sequential forward selection) (Siedlecki and Sklansky, 1989). The comparison has been rigorously done using a synthetic error model instead of real remotely sensed data so as to eliminate the variables (e.g. sample size, the number of spectral bands, the number of classes of interest, etc.) that could cause bias to the outcome of the comparison. In addition, mounting evidence of success of GA-based band selectors in real-life remote sensing applications are also found in recent literature: (i) selecting a subset of multiple sensor/date data for image classification (Lofy and Sklansky, 2001; Kavzoglu and Mather, 2002; Ulfarsson et al., 2003); (ii) selecting spectral bands that relate to physio-chemical characteristics of plants and soils (Fang et al., 2003; Kooistra et al., 2003; Cogdill and Rippke, 2004); (iii) selecting a spectral subset of hyperspectral data for image classification (Yu et al., 2002).

Nevertheless, class information that has been used in the aforementioned studies (Lofy and Sklansky, 2001; Kavzoglu and Mather, 2002; Yu et al., 2002; Ulfarsson et al., 2003) for testing the performance of the GA-based band selector is broad (i.e. USGS level I or II (Anderson et al., 1976)). This means that each class possesses distinct spectral characteristics from one another, and it is relatively easy for GA to find spectral bands that maintain high spectral separability between classes. None of the aforementioned studies has tested the band selector on class information that possesses very similar spectral characteristics (e.g. species-level data). A question therefore remains if GA can deal with such complexity. As a result, the key hypothesis of this research is that the GA-based band selector can be used for selecting a meaningful subset of spectral bands that maintains spectral separability between species classes. The testing data in use are very high-dimensional, spectrometer records that comprise 2151 bands of leaf spectra of 16 tropical mangrove species.

2. Species-level data and methods

2.1. Species-level hyperspectral data

2.1.1. Mangrove leaf preparation

Top-level canopies of 16 tropical mangrove species (Table 1) were collected using a line-transect method in the natural mangrove forest of Ao Sawi (Sawi Bay), the province of Chumporn, the south of Thailand (10° 15′N, 99° 7′E) on February, 6, 2001. This line-transect method enabled us to collect the mangrove canopies from pioneer, intermediate, and landward zones. The canopies were only sampled from fully-grown trees (i.e. >2.5 m tall). During the sampling campaign, species identification was carried out by the staffs of Royal Thai Forestry Department; taxonomy follows Tomlinson (1994) and Teeratanatorn (2000). At the laboratory, the leaves were then picked off the canopies for the spectral measurement.

2.1.2. Leaf spectral measurements

The leaves were randomly shuffled and separated evenly into 30 piles per mangrove species. Each pile of leaves (top side up) was placed on top of a black metal plate painted with ultra-flat black paint until the background metal plate could not be seen. Next, the spectral response of each leaf plate was recorded 20 times. Each plate was rotated 90° horizontally after every fifth record to compensate for the bi-directional reflectance

Table 1
Thirty spectra of mangrove leaves were collected per mangrove species, using a 2151-band spectroradiometer

Mangrove species	Species code	Number of spectra
Avicennia alba	1	30
Acrostichum aureum	2	30
Bruguiera cylindrica	3	30
Bruguiera gymnorrhiza	4	30
Bruguiera parviflora	5	30
Ceriops tagal	6	30
Excoecaria agallocha	7	30
Heritiera littoralis	8	30
Lumnitzera littorea	9	30
Lumnitzera racemosa	10	30
Nypa fruticans	11	30
Pluchea indica	12	30
Rhizophora apiculata	13	30
Rhizophora mucronata	14	30
Sonneratia ovata	15	30
Xylocarpus granatum	16	30

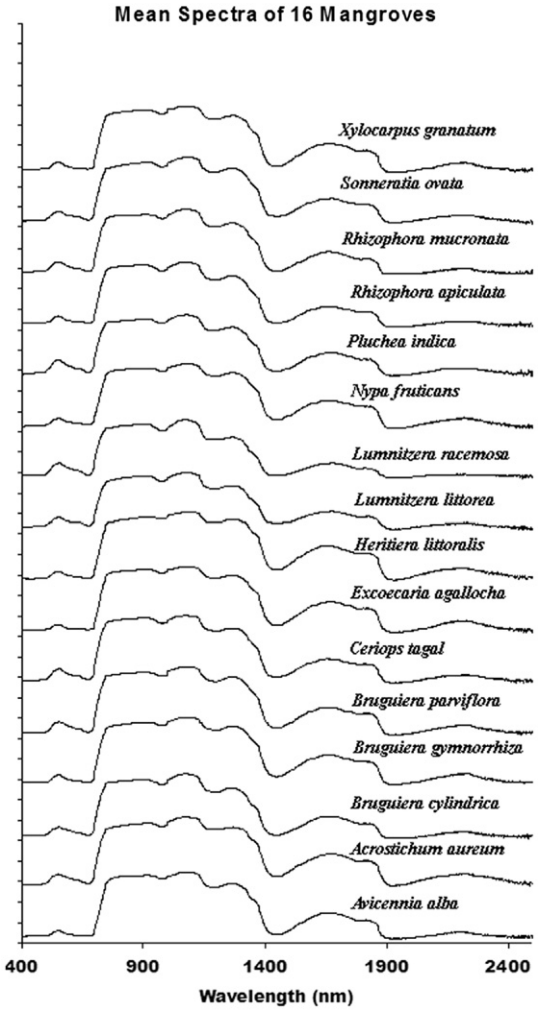

Fig. 1. The plot represents mean leaf spectra of 16 mangrove species stacked on top of one another. Each reflectance value tick mark along the *y*-axis is 0.2 reflectance value apart.

distribution function (BRDF). Then, the mean of the 20 records was calculated to construct a radiance curve. Finally, the radiance was converted to a reflectance curve by using a reference panel as well as the correction of the spectrometer internal current (dark current). The steps above were followed for all other leaf plates. As a result, we have 30 reflectance curves per each mangrove species (Table 1). The spectral mean of each species is illustrated in Fig. 1.

The spectral measurement was conducted under laboratory conditions by using a spectroradiometer (FieldSpec Pro FR, Analytical Spectral Device, Inc.). This spectroradiometer was equipped with three spectrometers (i.e. VNIR, SWIR1, and SWIR2), covering 350 nm to 2500 nm, with sampling intervals of 1.4 nm between 350 nm and 1000 nm, and 2 nm between 1000 nm and 2500 nm. The spectral resolution of the spectrometers was 3 nm for the wavelength interval 350 nm to 1000 nm, and 10 nm for the wavelength interval 1000 nm to 2500 nm. The sensor, equipped with a field of view of 25°, was mounted on a tripod and positioned 0.5 m above the leaf plate at the nadir position. A halogen lamp fixed on the tripod at the same position as the sensor of the spectrometer was used to illuminate the sample plate. The room was conditioned to be dark with 25 °C in order to avoid unwanted external energy sources.

The reader may note that leaf stacking experiments could cause some degrees of additional uncertainty to the resultant spectra. However, we did not try to remove such an effect from our spectral data. We included them in the measured signal and assumed that they are signal noise (i.e., Measured Mangrove Signal = Mean Mangrove Signal + Mangrove Spectral Variations + Signal Noise). Moreover, we have no intention to simulate any effect of background spectra (e.g., background spectra of understory plants and soils) in our measurement. We wanted to measure "pure" mangrove spectra from the leaf plate and use them for testing the proposed algorithm.

Statistical separability of the species-level data used in this study is also plotted in Fig. 2 (after Vaiphasa et al., 2005) so as to indicate where, along the whole wavelength region, the mangroves are likely to be spectrally separable (i.e. the locations where the black line (*p*-value trace) are below the lower statistical threshold ($\alpha = 0.01$)). Full details on this statistical visualization and its limitations as well as other aspect of statistical

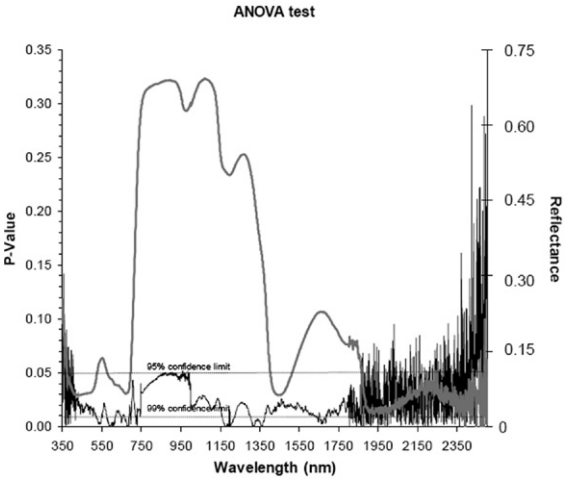

Fig. 2. After Vaiphasa et al. (2005), the plot shows *p*-values of the ANOVA test (black line) plotting against a laboratory reflectance of *Rhizophora apiculata* (gray line). The *p*-values indicate where, along the whole wavelength region, the mangroves are likely to be spectrally separable.

studies can be found in the previous work (Vaiphasa et al., 2005; Vaiphasa, 2006).

2.2. Genetic search algorithms (GA)

The theory of GA is first introduced by Holland (1975). The elaboration of its practical side including a basic computer source code can be found in Goldberg (1989). In this report, step-by-step guidelines of Goldberg (1989) are strictly followed (see Fig. 3). As a result, there is no attempt to make an exhaustive description of GA, but three major connections between the concept of GA and remote sensing applications are emphasized (i.e. gene encoding scheme, reproduction mechanism, and fitness criterion). Additionally, the code of GA used in this study has been developed in IDL language at the International Institute for Geo-Information Science and Earth Observation (ITC) (Vaiphasa, 2003).

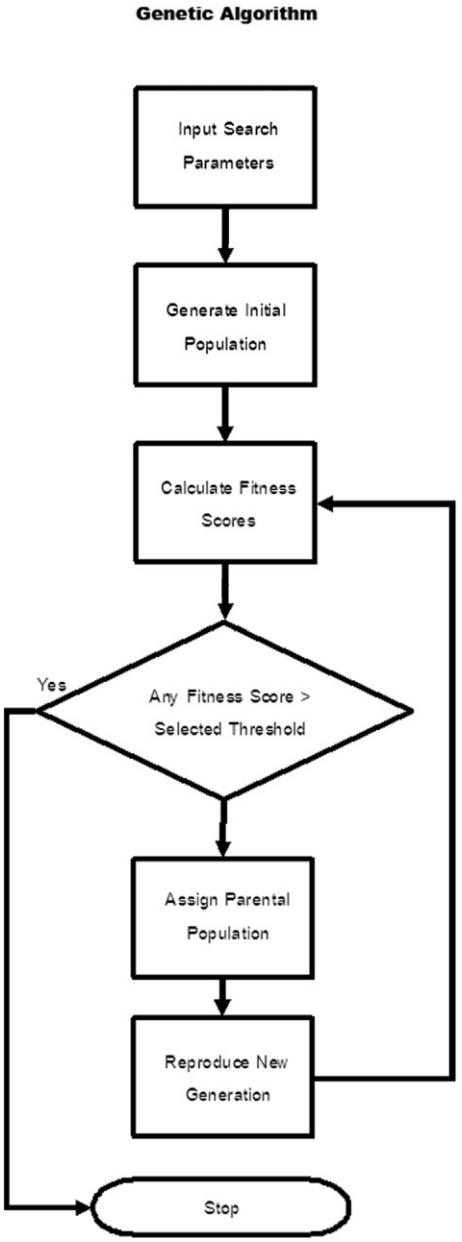

Fig. 3. A flowchart showing the process of the genetic algorithm used in this study.

2.2.1. Gene encoding

The gene encoding scheme in use is a direct method instead of binary encoding that is popularly used in related studies (Siedlecki and Sklansky, 1989; Kavzoglu and Mather, 2002; Yu et al., 2002). The key reasons for choosing direct encoding are that it is transparent for tracking the process of evolution as well as straightforward for reproducing the population (i.e. crossover and mutation) (Vaiphasa, 2003).

Fig. 4a illustrates an example of two chromosomes with chromosome size (number of gene inside the chromosome) equal to 6. Following the direct encoding scheme, the 1st chromosome comprises 6 different genes flagged by the letter A to F, and the 2nd one comprises gene G to L. Each gene can be assigned with a band label. For example, the 1st chromosome, {A, B, C, D, E, F}, in Fig. 4a can be assigned with an array of band names, {Band 2, Band 8, Band 37, Band 59, Band 97, Band 99}, and, similarly, the {G, H, I, J, K, L} can be set to {Band 3, Band 5, Band 38, Band 55, Band 83, Band 100}.

2.2.2. Reproduction mechanism

The mechanism of crossover and mutation is illustrated in Fig. 4. By mating the two chromosomes in Fig. 4a, the offspring that they produce share, in this example, half the characters of the first parent and the other half from the second parent. The two offspring are shown in Fig. 4b.

Occasionally, some of the genes in any newly produced chromosome are randomly altered by mutation. This phenomenon causes a change in the character of the offspring, independent from the chromosome composition of the parents. The illustration of the mutation effect is shown in Fig. 4c. The “J” gene is mutated to the “X” gene through random mutation. In remote sensing context, this is equal to a random flip of a band label inside a chromosome.

2.2.3. Fitness criterion

The fitness function chosen in this study is a well-known spectral angle mapper based nearest neighbor classifier (SAM). It is responsible for calculating fitness

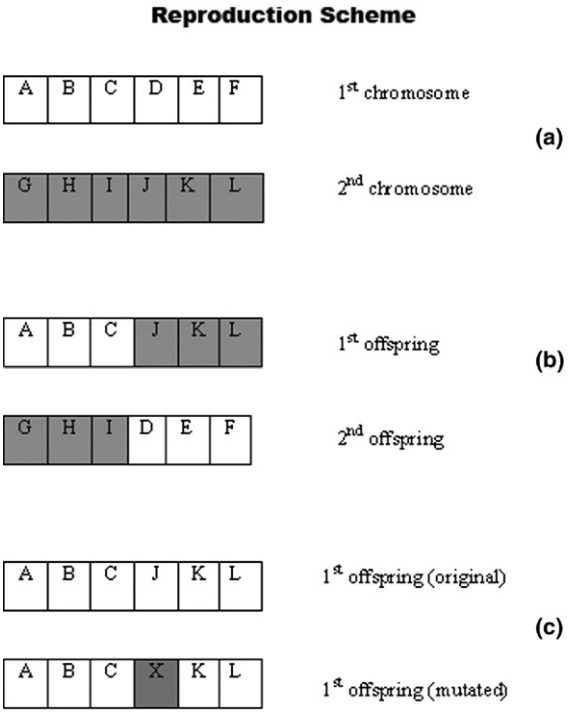

Fig. 4. The figures illustrate (a) Two parent chromosomes, (b) Two offspring chromosomes, and (c) An example of random mutation. Each block alphabet represents a spectral band label.

scores of the chromosome population during the evolution process (see Fig. 3). This means that the evolution is guided by the classification accuracy reported by SAM. Chromosomes (i.e. a subset of spectral bands) that possess higher classification accuracy are likely to have more chance to mate and produce young chromosomes that possess lower classification accuracy. The reader is recommended to consult Kruse et al. (1993) and Keshava (2004) for additional details on SAM.

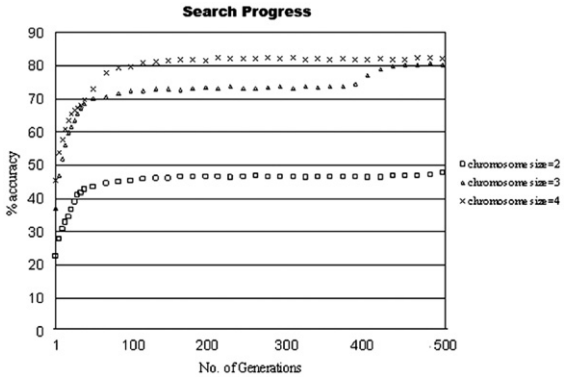

Fig. 5. A comparison between the performance of three different chromosome sizes.

3. Experiments and results

3.1. Initializing the genetic search algorithm

GA was initialized with the following parameters: population size = 1000, crossover rate = 100%, and mutation rate = 1%. The maximum number of generations was 500. The fitness function (i.e. SAM) was trained with half of the mangrove spectra of Table 1 (15 spectra per class), and the other half was used for calculating online fitness progress.

3.2. Choosing an appropriate chromosome size

Since the genetic algorithm in use was an unconstrained combinatorial optimization search (i.e. search without any constraint or penalty on the size of a chromosome), preliminary runs of GA had to be carried out to look for an appropriate chromosome size (i.e. chromosome size = the number of genes in a chromosome) that maintained high class separability (i.e. classification accuracy). The 80% level of classification accuracy was chosen as a threshold in this study as it was appropriate for separating very similar spectra of 16 mangrove species (USGS level III or IV (Anderson et al., 1976)). As a result, it was found that a minimum chromosome size that could maintain class separability above the chosen threshold was four. A comparison between the performance of three different chromosome sizes (i.e. chromosome size = 2, 3, and 4) was illustrated in Fig. 5.

3.3. Running the genetic search algorithm

GA with chromosome size four was repeatedly run 30 times to check the consistency of the results. The spectra were randomly rotated at the start of every run (i.e. data rotation) to avoid the bias. The real-time progress was plotted for each run in Fig. 6. The highest

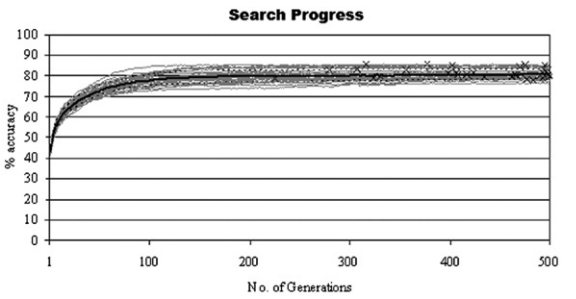

Fig. 6. The real-time progress of 30 runs (gray lines) with their peaks (crosses), mean (black line), and standard deviation limits (dashed lines).

fitness score of each run was marked with a cross. Overall, the genetic algorithm quickly reached an averaged fitness score level of 80% at about the 100th generation.

An example of the evolution process of a single run was illustrated in Fig. 7 to give an impression of how GA worked. The horizontal axis represented band labels (or genes) from B1 to B2151. The vertical axis of Fig. 7 is the frequency of gene types found in the population. In general, the gene distribution pattern converged from originally 2151 types of genes at the 1st generation (Fig. 7a) to only a few kinds of genes at the 500th generation (Fig. 7f). The convergence quickly happened as early as the 100th generation (Fig. 7b) where most genes were already extinct. This convergent evolution from Fig. 7a to b directly connected to Fig. 6 where the

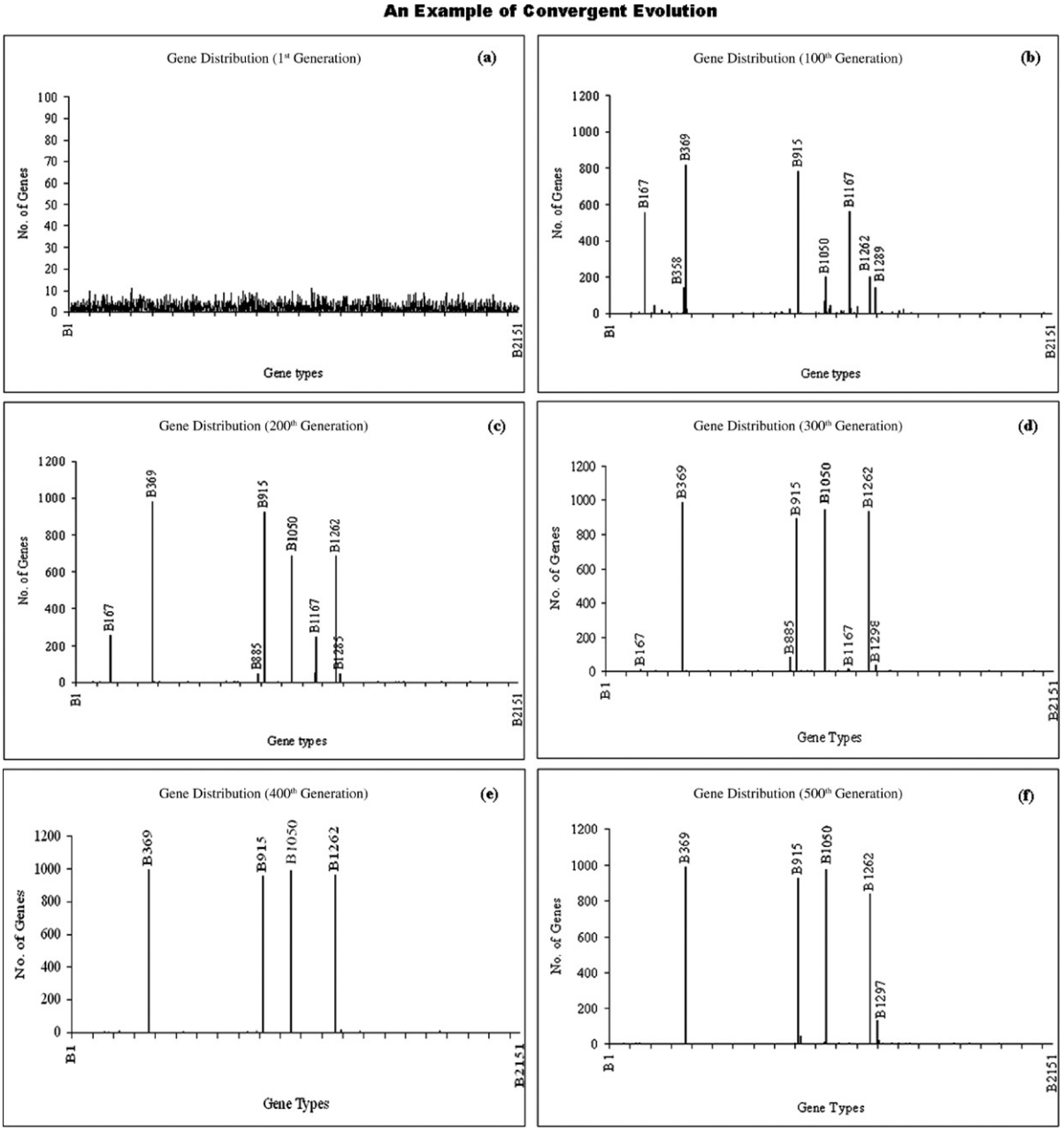

Fig. 7. This is an example of the convergence of gene distribution patterns from (a) the 1st generation to (f) the 500th generation. The horizontal axis represents band labels from B1 to B2151. The vertical axis of each plot is the frequency (count) of gene types found in the population.

majority of the progress lines leveled off at the 100th generation as the convergence happened. Genes that dominated the evolution were individually texted in the plots. In this example, at the last generation (the 500th generation), the gene pool was dominated by the following genes: B369, B915, B1050, B1262, and B1297.

The winning chromosomes from every run were reported in Table 2 along with their fitness scores (i.e. SAM classification accuracy). The best of all were chromosome No.2 and No.10. Both possessed an 86% level of classification accuracy. Then, all the genes of the 30 winning chromosomes (i.e. 120 genes in total) are grouped by minimizing their variance. The results were illustrated in a plot against a mangrove reflectance (Fig. 8). It was found that the genes (spectral bands) can be grouped at 6 different spectral positions (mean± standard deviation): visible area (21 genes at 513± 19 nm); red edge (15 genes at 717±16 nm); near-infrared region (9 genes at 1263±23 nm); infrared slope (44 genes at 1385±27 nm); mid-infrared absorption pitch (5 genes at 1489±21 nm), and mid-infrared peak (26 genes at 1669±25 nm).

Table 2
30 winning chromosomes with their encapsulated genes

Chromosome no.	Genes (nanometer)				Fitness scores (%)
1	523	1358	1385	1710	83
2	518	1381	1393	1685	86
3	517	708	1436	1639	82
4	523	1372	1418	1671	81
5	524	1375	1496	1681	82
6	521	1378	1398	1665	79
7	549	1333	1390	1681	81
8	679	1316	1389	1673	79
9	535	1384	1506	1667	85
10	716	1246	1409	1607	86
11	722	758	1392	1436	80
12	534	1364	1385	1685	84
13	528	1363	1408	1661	81
14	725	1264	1402	1682	79
15	533	1369	1507	1660	79
16	548	1335	1458	1644	80
17	546	714	1403	1626	82
18	515	1380	1409	1674	78
19	593	1388	1480	1667	80
20	711	1234	1381	1699	83
21	536	717	1230	1397	85
22	526	725	1253	1395	83
23	495	705	1355	1398	81
24	726	1282	1381	1692	83
25	532	1350	1435	1676	79
26	523	1368	1418	1671	85
27	717	1290	1389	1668	80
28	717	1276	1405	1721	82
29	713	1289	1393	1658	78
30	528	1337	1386	1628	81

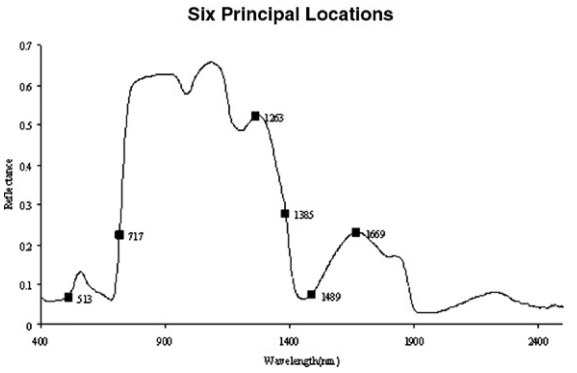

Fig. 8. After all the genes of the 30 winning chromosomes reported in Table 2 are grouped by minimizing their variance, it was found that the genes (spectral bands) can be grouped at 6 different spectral positions.

3.4. Testing the key hypothesis

The key hypothesis of this study was tested to see whether the results of band selection done by GA were meaningful. Specifically, the results of GA were statistically compared against the results of random selection using *t*-tests. The Jeffries–Matusita (J–M) distance was chosen as an evaluation tool (Richards, 1993). For each of the 30 winning chromosomes in Table 2, its 4 encapsulated spectral bands were used for calculating J–M distances between all mangrove classes. The averaged J–M distances of the 30 winning chromosomes were demonstrated in Table 3a. Next, the J–M distances were calculated for 30 sets of randomly generated band combinations, and their averaged J–M distances were reported in Table 3b. Subsequently, the *t*-test results between the two cases were demonstrated in Table 3c in terms of *p*-values. It was clear that the class separability of band combinations selected by the genetic algorithm was significantly higher than the class separability of randomly selected band combinations with a 95% level of confidence ($\alpha = 0.05$), as most of the *p*-values ($94/120 \approx 78\%$) in Table 3c were ≤ 0.05.

3.5. Discussion and conclusion

In this study, a form of GA-based band selectors was challenged to select spectral subsets of very high-dimensional, species-level data. Unlike the broad-level data (i.e. Anderson's level I or II (Anderson et al., 1976)) used in the existing studies (Lofy and Sklansky, 2001; Kavzoglu and Mather, 2002; Yu et al., 2002; Ulfarsson et al., 2003), spectral profiles of the species-level data were very similar (Figs. 1 and 2). Despite that, the results in Table 2 and Fig. 6 demonstrated that the GA-based band selector overcame spectral similarity of

Table 3
A statistical comparison between the class separability of band combinations selected by the genetic algorithm and the class separability of band combinations selected by chance (please see Table 1 for the class information): (a) genetic search result; (b) random search result; and (c) *p*-values

Class	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
(a)																
1																
2	2.00															
3	2.00	2.00														
4	1.99	1.74	2.00													
5	1.98	2.00	2.00	1.87												
6	2.00	2.00	2.00	2.00	1.84											
7	1.98	2.00	2.00	1.98	2.00	2.00										
8	2.00	1.95	2.00	1.90	2.00	2.00	1.99									
9	2.00	2.00	2.00	2.00	2.00	1.99	2.00	2.00								
10	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00							
11	2.00	1.97	2.00	1.87	1.95	2.00	1.97	1.99	2.00	2.00						
12	2.00	2.00	1.99	2.00	2.00	1.99	2.00	2.00	2.00	2.00	2.00					
13	1.97	2.00	1.94	2.00	1.99	2.00	2.00	2.00	2.00	2.00	2.00	2.00				
14	1.99	2.00	1.96	2.00	1.97	1.96	2.00	2.00	2.00	2.00	2.00	1.98	1.97			
15	1.98	2.00	2.00	1.99	1.82	1.97	2.00	2.00	2.00	2.00	1.98	2.00	1.97	1.98		
16	1.91	2.00	2.00	1.96	1.99	2.00	1.98	2.00	2.00	2.00	2.00	2.00	1.99	2.00	1.98	
(b)																
1																
2	1.99															
3	1.84	2.00														
4	1.89	1.41	1.98													
5	1.73	1.98	1.97	1.77												
6	1.56	1.99	1.91	1.95	1.59											
7	1.88	1.86	1.99	1.55	1.64	1.88										
8	1.99	1.64	2.00	1.73	1.98	1.98	1.88									
9	1.96	1.99	1.98	1.96	1.95	1.92	1.98	1.97								
10	1.95	2.00	1.95	1.99	1.98	1.98	1.98	2.00	1.94							
11	1.81	1.97	1.97	1.68	1.67	1.92	1.67	1.98	1.95	1.96						
12	1.82	1.98	1.95	1.91	1.90	1.82	1.88	1.97	1.92	1.92	1.86					
13	1.75	2.00	1.22	1.98	1.97	1.83	1.99	2.00	1.97	1.94	1.96	1.96				
14	1.51	1.99	1.68	1.97	1.79	1.46	1.93	2.00	1.94	1.91	1.93	1.81	1.55			
15	1.79	2.00	1.98	1.94	1.75	1.77	1.97	2.00	1.93	1.99	1.81	1.86	1.96	1.82		
16	1.58	1.82	1.95	1.36	1.47	1.78	1.38	1.90	1.97	1.97	1.51	1.77	1.94	1.83	1.77	
(c)																
1																
2	0.01															
3	0.00	0.11														
4	0.00	0.00	0.05													
5	0.00	0.03	0.03	0.02												
6	0.00	0.06	0.00	0.02	0.00											
7	0.00	0.00	0.02	0.00	0.00	0.00										
8	0.06	0.00	0.16	0.00	0.12	0.13	0.01									
9	0.00	0.07	0.03	0.05	0.05	0.01	0.10	0.04								
10	0.04	0.10	0.04	0.06	0.08	0.02	0.10	0.15	0.01							
11	0.00	0.50	0.03	0.00	0.00	0.00	0.00	0.06	0.02	0.06						
12	0.00	0.09	0.02	0.04	0.02	0.00	0.01	0.14	0.00	0.01	0.01					
13	0.00	0.07	0.00	0.05	0.03	0.00	0.04	0.08	0.10	0.02	0.02	0.04				
14	0.00	0.11	0.00	0.02	0.00	0.00	0.02	0.13	0.00	0.02	0.00	0.00	0.00			
15	0.00	0.03	0.02	0.01	0.12	0.00	0.01	0.03	0.01	0.12	0.00	0.00	0.22	0.00		
16	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.01	0.00	0.00	

the species-level data as it was able to select spectral subsets that maintained class separability at an acceptable level (i.e. ≈an 80% level of classification accuracy).

Additionally, the results of hypothesis testing in Table 3 also confirmed that band selection done by the GA-based band selector was meaningful. By majority, spectral separability between 16 mangrove species when the spectral bands were selected by chance was significantly lower than when the spectral bands were selected by the GA-based band selector (i.e. with a 95% level of confidence).

The success of the GA-based band selector may be explained by the chosen spectral locations (Fig. 8), as each location directly related to principal physio-chemical properties of plants that helped distinguish between the species. The details of the relationships between these spectral locations and plants can be found in the following literature. In brief, the band selected from the visible area was needed for discriminating between mangroves that possessed different leaf pigments and different sensitivity levels to the visible light source (Elvidge, 1987, 1990; Curran, 1989; Menon and Neelakantan, 1992; Basak et al., 1996; Kumar et al., 2001; Das et al., 2002). The band on the red-edge slope was for separating mangrove species that contained different leaf pigments, internal leaf structure and water (Elvidge, 1987, 1990; Curran, 1989; Kumar et al., 2001; Williams and Norris, 2001). Similar to the red edge band, the near-infrared band helped sort different plants according to the dissimilarity of their leaf internal structure such as the size of intercellular volume (Elvidge, 1987, 1990; Curran, 1989; Kumar et al., 2001; Williams and Norris, 2001). Finally, the spectral information of the mid-infrared region (i.e. the infrared slope, the mid-infrared absorption pitch, and the mid-infrared peak) was necessary for dissolving the internal structure variables and foliar biochemical contents other than the leaf pigments (Himmelsbach et al., 1988; Curran, 1989; Kumar et al., 2001).

The reader may note that the form of GA and its parameters used in this study were not the only options available. To tackle the problem at hand, it was possible to alter the encoding scheme, population size, crossover rate, and mutation rate. Additionally, the fitness function could be replaced with any popular pattern classifier other than SAM. Using other decision criteria for assigning parental chromosomes instead of the biased roulette wheel, suggested by Goldberg (1989), is also possible. Even though the alteration may affect the evolution process depicted in Fig. 3, it was expected that the robustness of the evolutionary search could still produce a similar outcome (see "freedom of choice" in Goldberg (1989), page 80). In other words, GA was likely to find meaningful spectral bands that possess high spectral separability despite the alteration. It is, however, beyond the scope of this study to compare different designs of GA and the use of different search parameters.

The optimism gained from the results of this laboratory-level study (i.e. using laboratory spectra) encouraged further investigation into the potential of the GA-based band selector for vegetation discrimination when hyperspectral images taken by airborne or satellite sensors above mangrove canopies are in use. This will surely increase the complexity of measured spectral signals as a number of additional factors are involved (e.g. the fluctuation of light source energy, the change of daily atmospheric states, the effect of canopy formations, the cost of accessibility, the coarser spatial and spectral resolutions of on-board hyperspectral sensors, the effect of seasonal changes, the effect of background soils and water, the difference between the energy of artificial lamps used in the laboratory and the sun, etc.). The reader may consult Ramsey and Jensen (1996) on this issue. They have illustrated the differences between leaf-level and canopy-level spectra measured from Florida mangroves. Furthermore, it was also anticipated that the use of the GA-based band selector was not limited to the application of vegetation discrimination. The GA-based band selector is now being tested by the author to detect spectral bands that show strong vegetation responses to different physio-chemical treatments (e.g. nitrogen, illumination, etc.) in both laboratory and field scenarios. It is hoped that the GA-based band selector could be used as an alternative to traditional methods such as statistical and derivative analyses that are normally used for detecting vegetation responses to external influences (Tsai and Philpot, 1998; Mutanga et al., 2003).

In conclusion, this study strengthened the confidence of using GA as band selection tools. The results confirmed that the GA-based band selector was able to cope with spectral similarity at the species level. It meaningfully selected spectral bands that related to principal physio-chemical properties of plants, and, simultaneously, maintained the separability between species classes at a high level. Additionally, the application of the GA-based band selector other than vegetation discrimination such as the investigation into vegetation spectra in response to different physio-chemical treatments was also anticipated.

Acknowledgements

Thanks to Dr. Suwit Ongsomwang for his great support on the previous publication that inspired us to

write this manuscript. Thanks to Henk van Oosten for his help on part of the IDL program. This manuscript is financially supported by Ratchadaphiseksomphot Endowment Fund, Chulalongkorn University, Bangkok, Thailand.

References

Anderson, J.R., Hardy, E.E., Roach, J.T., Witmer, R.E., 1976. A land use and land cover classification system for use with remote sensor data. U.S. Geological Survey Professional Paper 964, 1–41.

Bandyopadhyay, S., 2005. Satellite image classification using genetically guided fuzzy clustering with spatial information. International Journal of Remote Sensing 26 (3), 579–593.

Basak, U.C., Das, A.B., Das, P., 1996. Chlorophyll, carotenoids, proteins and secondary metabolites in leaves of 14 species of mangroves. Bulletin of Marine Science 58 (3), 645–659.

Bellman, R.E., 1961. Adaptive Control Processes. Princeton University Press, Princeton.

Chalermwat, P., El-Ghazawi, T., LeMoigne, J., 2001. 2-phase GA-based image registration on parallel clusters. Future Generations Computer Systems 17 (4), 467–476.

Chen, L., 2003. A study of applying genetic programming to reservoir trophic state evaluation using remote sensor data. International Journal of Remote Sensing 24 (11), 2265–2275.

Cogdill, R.P., Hurburgh Jr., C.R., Rippke, G.R., 2004. Single-kernel maize analysis by near-infrared hyperspectral imaging. Transactions of the American Society of Agricultural Engineers 47 (1), 311–320.

Curran, P.J., 1989. Remote sensing of foliar chemistry. Remote Sensing of Environment 30 (3), 271–278.

Das, A.B., Parida, A., Basak, U.C., Das, P., 2002. Studies on pigments, proteins and photosynthetic rates in some mangroves and mangrove associates from Bhitarkanika, Orissa. Marine Biology 141 (3), 415–422.

Elvidge, C.D., 1987. Reflectance characteristics of dry plant materials. Proc. the 21st International Symposium on Remote Sensing of Environment, Ann Arbor, pp. 721–733.

Elvidge, C.D., 1990. Visible and near infrared reflectance characteristics of dry plant materials. International Journal of Remote Sensing 11 (10), 1775–1795.

Fang, H., Liang, S., Kuusk, A., 2003. Retrieving leaf area index using a genetic algorithm with a canopy radiative transfer model. Remote Sensing of Environment 85 (3), 257–270.

Fukunaga, K., 1990. Introduction to Statistical Pattern Recognition. Academic Press, Orlando.

Goldberg, D., 1989. Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, Reading.

Harvey, N.R., Theiler, J., Brumby, S.P., Perkins, S., Szymanski, J.J., Bloch, J.J., Porter, R.B., Galassi, M., Young, A.C., 2002. Comparison of GENIE and conventional supervised classifiers for multispectral image feature extraction. IEEE Transactions on Geoscience and Remote Sensing 40 (2), 393–404.

Himmelsbach, D.S., Boer, J.D., Akin, D.E., Barton, E.E., 1988. Solid-state carbon-13 NMR, FTIR, and NIR spectroscopic studies of ruminant silage digestion. In: Creaser, C.S., Davies, A.M.C. (Eds.), Analytical Applications of Spectroscopy. Royal Society of Chemistry, London, pp. 410–413.

Holland, J.H., 1975. Adaptation in Natural and Artificial Systems. University of Michigan, Ann Arbor.

Hughes, G.F., 1968. On the mean accuracy of statistical pattern recognizers. IEEE Transactions on Information Theory 14 (1), 55–63.

Jin, Y.Q., Wang, Y.A., 2001. Genetic algorithm to simultaneously retrieve land surface roughness and soil wetness. International Journal of Remote Sensing 22 (16), 3093–3099.

Jones, G.A., Greenhill, D., Orwell, J., Forte, P., 2000. Coastline registration: efficient optimization in large dimensions using genetic algorithms. Geographical and Environmental Modelling 4 (1), 21–41.

Kavzoglu, T., Mather, P.M., 2002. The role of feature selection in artificial neural network applications. International Journal of Remote Sensing 23 (15), 2919–2937.

Kendall, M.G., 1961. A Course in the Geometry of n Dimensions. Dover Publications, Inc., New York.

Keshava, N., 2004. Distance matrices and band selection in hyperspectral processing with applications to material identification and spectral libraries. IEEE Transactions on Geoscience and Remote Sensing 42 (7), 1552–1565.

Kooistra, L., Wanders, J., Epema, G.F., Leuven, R.S.E.W., Wehrens, R., Buydens, L.M.C., 2003. The potential of field spectroscopy for the assessment of sediment properties in river floodplains. Analytica Chimica Acta 484 (2), 189–200.

Kruse, F.A., Lefkoff, A.B., Boardman, J.W., Heidebrecht, K.B., Shapiro, A.T., Barloon, P.J., Goetz, A.F.H., 1993. The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data. Remote Sensing of Environment 44 (2–3), 145–163.

Kumar, L., Schmidt, K.S., Dury, S., Skidmore, A.K., 2001. Review of hyperspectral remote sensing and vegetation science. In: van der Meer, F.D., de Jong, S.M. (Eds.), Imaging Spectrometry: Basic Principles and Prospective Applications. Kluwer Academic Press, Dordrecht, pp. 111–155.

Liu, Z., Liu, A., Wang, C., Niu, Z., 2004. Evolving neural network using real coded genetic algorithm (GA) for multispectral image classification. Future Generations Computer Systems 20 (7), 1119–1129.

Lofy, B., Sklansky, J., 2001. Segmenting multisensor aerial images in class-scale space. Pattern Recognition 34 (9), 1825–1839.

Lu, F., Eriksson, L.O., 2000. Formation of harvest units with genetic algorithms. Forest Ecology and Management 130 (1–3), 57–67.

Luo, J.C., Zheng, J., Leung, Y., Zhou, C.H., 2003. A knowledge-integrated stepwise optimization model for feature mining in remotely sensed images. International Journal of Remote Sensing 24 (23), 4661–4680.

Menon, G.G., Neelakantan, B., 1992. Chlorophyll and light attenuation from the leaves of mangrove species of Kali estuary. Indian Journal of Marine Sciences 21 (1), 13–16.

Mertens, K.C., Verbeke, L.P.C., Ducheyne, E.I., De Wulf, R.R., 2003. Using genetic algorithms in sub-pixel mapping. International Journal of Remote Sensing 24 (10), 4241–4247.

Mutanga, O., Skidmore, A.K., Van Wieren, S., 2003. Discriminating tropical grass (*Cenchrus ciliaris*) canopies grown under different nitrogen treatments using spectroradiometry. ISPRS Journal of Photogrammetry and Remote Sensing 57 (4), 263–272.

Pal, S.K., Bandyopadhyay, S., Murthy, C.A., 2001. Genetic classifiers for remotely sensed images: comparison with standard methods. International Journal of Remote Sensing 22 (13), 2545–2569.

Ramsey, E.W., Jensen, J.R., 1996. Remote sensing of mangrove wetlands: relating canopy spectra to site-specific data. Photogrammetric Engineering and Remote Sensing 62 (8), 939–948.

Richards, J.A., 1993. Remote Sensing Digital Image Analysis: An Introduction. Springer-Verlag, Berlin.

Shahshahani, B.M., Landgrebe, D.A., 1994. The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon. IEEE Transactions on Geoscience and Remote Sensing 32 (5), 1087–1095.

Siedlecki, W., Sklansky, J., 1989. A note on genetic algorithms for large-scale feature selection. Pattern Recognition Letters 10 (5), 335–347.

Teeratanatorn, W., 2000. Mangroves of Pak Phanang Bay (in Thai). Royal Forest Department, Bangkok.

Tomlinson, P.B., 1994. The Botany of Mangroves. Cambridge University Press, Cambridge.

Tsai, F., Philpot, W., 1998. Derivative analysis of hyperspectral data. Remote Sensing of Environment 66 (1), 41–51.

Tseng, D.C., Lai, C.C., 1999. A genetic algorithm for MRF-based segmentation of multi-spectral textured images. Pattern Recognition Letters 20 (14), 1499–1510.

Ulfarsson, M.O., Benediktsson, J.A., Sveinsson, J.R., 2003. Data fusion and feature extraction in the wavelet domain. International Journal of Remote Sensing 24 (20), 3933–3945.

Vaiphasa, C., 2003. Innovative genetic algorithm for hyperspectral image classification. Proc. MAP ASIA 2003 Conference, Kuala Lumpur, 13–15 October 2003, p. 45.

Vaiphasa, C., 2006. Consideration of smoothing techniques for hyperspectral remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 60 (2), 91–99.

Vaiphasa, C., Ongsomwang, S., Vaiphasa, T., Skidmore, A.K., 2005. Tropical mangrove species discrimination using hyperspectral data: a laboratory study. Estuarine, Coastal and Shelf Science 65 (1–2), 371–379.

Williams, P., Norris, K., 2001. Near-Infrared Technology in the Agricultural and Food Industries. American Association of Cereal Chemists, St. Paul, Minnesota.

Yu, S., De Backer, S., Scheunders, P., 2002. Genetic feature selection combined with composite fuzzy nearest neighbor classifiers for hyperspectral satellite imagery. Pattern Recognition Letters 23 (1–3), 183–190.