EXPERT CLASSIFICATION TECHNIQUE FOR MAPPING TEAK PLANTATION AREAS IN THAILAND

Siripun Taweesuk  
Prasong Thammapala  
Faculty of Science and Technology  
Thammasat University, Rangsit Campus  
99 Moo 18, Paholyotin Rd., Pathumtani, 12120, Thailand

ABSTRACT

Thailand is one of the ten countries with the largest plantation development programs, which occupying three percent of the total global plantation area. Teak specie (Tectona grandis) makes up more than half of all the forest plantation areas (FAO, 2003). Thus, deriving the accurate and up-to-date teak plantation areas are necessary steps in forest plantation management as these data are significant to the decision-making processes. In this paper, we focus on the use of remote sensing analysis techniques with the Landsat data in order to discriminate teak plantation areas. Our main objective is to explore whether the expert classification technique can improve the accuracy of teak plantation classification when compare to the typical classification technique, such as supervised classification. In our experiment, expert and supervised classification techniques are applied to the Landsat data sets, which cover area of Muang district, Petchabun province Thailand. The training and reference data were acquired from the field survey that took place roughly at the same time as the acquisition of the satellite data in 2003. The overall accuracy, kappa statistics, producer’s and user’s accuracy were computed from the error matrices in order to evaluate and compare the accuracy of the classification. The analysis of pairwise and Z-test were also employed for statistical interpretation. Our results indicated that using an expert classification technique yielded a significantly higher discrimination accuracy of teak plantation. However, the complication of an analytical procedure, the processing time and cost of analysis may overcome the benefits. The knowledge and experience of the data analysts, as well as other factors should be taken into consideration when selecting an appropriate classification technique.

INTRODUCTION

Maps that accurately discriminate teak plantation areas from other land cover types are highly desirable for efficient and sustainable management of forest plantation in Thailand. Especially, in the areas that are dominated by monoculture of teak plantations. In recent literatures, the potential of various classification techniques based on remote sensing for teak plantation mapping are rarely explored. The remotely sensed data, however, are extensively applied in other types of land cover classification. In a digital image classification process, there is a limitation in deriving land cover information from the inter-band correlation of multispectral images as the spectral signatures are often similar, particularly in broad-leaves deciduous species (Jensen, 2000; Lillesand and Kiefer, 2000; Schowengerdt, 1997; Schriever & Congalton, 1995). Supervised classification technique with the traditional maximum likelihood decision rule was thus a primary algorithm that has been widely applied in deriving such information from the Landsat data. As the method is based on statistical models, non-unique spectral signatures can degrade the accuracy of the classifier. Many studies have attempted to incorporate ancillary data, such as digital elevation model (DEM), slope, principal components analysis (PCA) and other geographic variables in order to improve the accuracy of the image classification (Eiumnoh and Shrestha, 2000; Palacio-Prieto and Luna-González, 1996).

An expert classification system as define by Leica Geosystems (2003) is “a hierarchy of rules, or a decision tree that describes the conditions under which a set of low level constituent information (user-defined variables, raster imagery, vector coverages, spatial models, external program and simple scalars) gets abstracted into a set of high level information classes”. Many research papers have revealed the potential of expert system (knowledge-based system) incorporated with the GIS data in order to improve the accuracy of image classification. Gao, et al. (2004) applied the spatial knowledge in GIS to the classification process of SPOT data, and concluded that it considerably
improved the accuracy of mapping mangrove forest in New Zealand. Another expert system for image classification was implemented by combining airborne hyperspectral imagery with terrain data. This system also increased the accuracy of coastal vegetation map in salt marshes in the Netherlands (Schmidt, et al., 2004). Huang and Jensen (1997) built the knowledge bases by using a machine learning approach for an integration of image analysis expert system, SPOT multispectral imagery, and GIS data to classify wetland areas in South Carolina, USA. The result yielded the highest accuracy and was significantly different from the traditional unsupervised and supervised classification.

This research aims at developing digital image classification technique for discrimination of forest plantation in tropical landscape, specifically for teak plantation areas in Thailand. Our work investigated the available integration of multi-sources, spectral and non-spectral data applying in expert classification technique. The results are then compared with the traditional supervised classification technique.

**METHODS**

**Study Area**

In 1994, the government subsidies, provided to the private land owner practicing forest plantation, has declared teak as the significantly precious specie. Our research thus emphasized the classification of teak plantation areas in Thailand. The study area covers about 2,200 sq. km. in parts of Pasak watershed, administratively in Muang District, Petchabun Province, in the northern part (Figure 1). The terrain elevation of the area ranges from 100 to 1,340 meters. The western and the eastern parts are mostly mountainous with slopes higher than 35 percent, and covered by primary and secondary of mixed deciduous and dry dipterocarp forests. Most of the areas in the central part lies from north to south are under agricultural practices with the major crops being paddy, sugarcane, maize, orchards (mostly are tamarind) and other plantation crops. Forest plantation, especially teak plantation areas are distributed over Petchabun province. Sixty percent of teak plantation areas were at seven to ten years old with an average diameter at breast height (DBH) of more than 9.0 centimeters and the total height between 7.0 - 14.0 meters (Thammapala and Taweesuk, 2004).

![Figure 1](image1.png)

**Figure 1** The study area in Muang District, Petchabun Province, Thailand

**Image Acquisition and Preprocessing**

The subset of two Landsat scenes (band 1, 2, 3, 4, 5 and 7) covered path 129 and row 49 with cloud cover less than five percent were obtained for this experiment. The two scenes (Landsat 7 ETM+ and Landsat 5 TM) were acquired on January 11 (dry season) and November 19, 2003 (wet season) respectively. Then, the Landsat 7 image was geometrically corrected to the UTM Zone 47 coordinate system using forty ground control points (GCPs), obtained from the field survey using the GPS device. The Landsat 5 image was also registered to the Landsat 7 image.
The third order polynomial transformation and the nearest neighborhood resampling technique were then applied to both images. All the RMS errors occurred from the geometric correction process were acceptable at less than one pixel. Figure 2 describes the steps of Landsat data preprocessing.

![Flowchart showing the steps of Landsat data preprocessing](image)

**Field Data Gathering**

Field survey for training and independent reference data were gathered roughly at the same time as the Landsat data acquisition. The UTM zone 47 coordinate system data were recorded by using the GPS device in the field. Selection of the independent reference or the ground truth data was based on a stratified random sampling.

**Database for Knowledge-based Construction**

To identify and map teak plantation using expert classification, we carried out various pre-processing steps of spatial data from spectral and non-spectral database. The data were also incorporated with data from field survey for each of land cover type. The data can be described as follow:

- **Spectral data:** First, Landsat 5 and 7 DN values were analyzed to obtain the range of specific spectral signature for each land cover type. Second, the normalize difference vegetation index (NDVI) and normalize difference moisture index (NDMI) from Landsat 5 and 7 were calculated and used as a criteria in the process of expert classification. Third, the principle component analysis (PCA) on the stacked image of Landsat 5 and 7 (band1-5 and 7) were performed in order to obtain the components that can aid in land cover classification. Taweesuk and Thammapala (2004) stated that the use of principal component analysis (PCA) in the classification process could improve the discrimination of Teak plantation.

- **Non-spectral data:** We generated and used five GIS layers in spatial analysis to build the rule-based information used in the process of expert classification. These spatial data inputs were digital elevation model (DEM) and slope images, which were derived from digital contour lines coverage, and other three raster images of conservative and non-conservative forest and municipal boundaries. These inputs were generated from the vector format.

**Image Processing and Classification**

To meet the study objectives, an expert classification for mapping teak plantation were developed and compared with the conventional supervised classification: maximum likelihood algorithms. We performed all analysis using ERDAS IMAGINE Version 8.7 and ArcGIS 9: ArcInfo software. Six land cover types: forest and semi-natural areas,
teak plantation, agricultural areas, urban and built-up areas, orchards and water bodies were classified. The details of each class are described in Table 1. The steps of applying both image classification techniques are shown in Figure 3.

Table 1 Land cover class name definition using in Landsat data classification schemes

<table>
<thead>
<tr>
<th>Class</th>
<th>Class Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Forest and Semi-natural Areas</td>
<td>Vegetation formation composed predominantly of evergreen and deciduous trees including shrubs, bush under-stories, forest areas in conservative forest and secondary forest in other areas</td>
</tr>
<tr>
<td>2</td>
<td>Teak Plantation</td>
<td>Areas covered with monoculture of Teak plantation with the average DBH &gt; 9.0 cm.</td>
</tr>
<tr>
<td>3</td>
<td>Agricultural Areas</td>
<td>Paddy, field crops (maize, sugarcane, tobacco, etc.), and bare land (exposed soil surface covered with shrubs or trees &lt; 10 percent)</td>
</tr>
<tr>
<td>4</td>
<td>Urban and Built-up Areas</td>
<td>A combination of buildings, urban material and vegetation mixed with urban/built-up areas, residential and other man-made structures such as industrial, commercial, and services, and other utilities</td>
</tr>
<tr>
<td>5</td>
<td>Orchards</td>
<td>Monoculture of plantation crops such as: tamarinds, mangoes, and others</td>
</tr>
<tr>
<td>6</td>
<td>Water Bodies</td>
<td>Natural or artificial open water surfaces such as: lakes, reservoirs, and etc.</td>
</tr>
</tbody>
</table>

**Supervised Classification** A one-pixel-to-one class method or the supervised classification with maximum likelihood classifier, was conducted to classify the Landsat 7 image using the spectral signatures from the training areas. At least five training areas from field survey were selected for each land cover class in order to generate the spectral signatures using in the process.

**Expert Classification** The expert system for mapping teak plantation was developed using the IMAGINE Expert Classifier™. We applied the expert knowledge gained from the spatial analysis of spectral and non-spectral data in the study area. The DNs values from each band and the values of NDVI and NDMI were generated from image transformation. These values were our spectral data. The PCA1 and PCA3 were also used as our data inputs. Moreover, we incorporated the geographical background knowledge of unique class, such as the teak plantation class, based on terrain elevation and slope that was derived from DEM. Figure 4 shows the rules and data inputs used in the expert classification process.

**Post-classification** The classified images were smoothen twice with a 3 x 3- majority rule filter in order to eliminate salt-and-pepper from the classification process (post classification filtering).

**Accuracy Assessment**

An accuracy assessment determines the quality and the reliability of information extracted from the satellite imagery. The classification accuracy on two classified images from both supervised with maximum likelihood and expert classification techniques were evaluated against 350 independent reference data acquired from the field survey with the stratified random sampling. A minimum of fifty reference samples for each land cover categories were applied in the error matrix (Congalton and Green, 1999). An overall accuracy, a producer’s accuracy and a user’s accuracy from an error matrix together with the kappa statistic were then calculated.

**Statistical Analysis**

In order to compare the two classification techniques, the analysis of pairwise significant and Z-test were analyzed at the ninety five percent confidence level. The analysis may show the possibility of one technique producing statistically higher accuracy for teak plantation mapping than the other.
Figure 3 Flowchart showing the steps of data analysis for teak plantation mapping using supervised classification with maximum likelihood classifier and expert classification techniques.
RESULTS AND DISCUSSION

The results from our study demonstrated that the expert classification technique produced an overall accuracy of 81.1 percent with the kappa value of 0.7717, which was higher than 70.6 percent with the kappa of 0.6395 using the maximum likelihood classifier. The 84.8 percent of teak plantation pixels were correctly classified using expert classifier. This number was also higher than that of the traditional supervised classifier (64.4 percent). Furthermore, the expert classification error matrix revealed higher producer’s accuracy from “forest and semi-natural areas”, “agricultural areas” and “urban and built-up areas” classes with the values of 79.7, 87.7 and 78.8 percent, respectively.
The supervised classification error matrix, on the other hand, showed the values of 65.6, 78.1 and 51.9 percent of accuracy. Tables 2 and 3 show in details the accuracy of teak plantation and other land cover types classification for both techniques.

**Table 2** The classification error matrix shows the accuracy of teak plantation and other land cover types using supervised classification: maximum likelihood algorithm

<table>
<thead>
<tr>
<th>Supervised Classification</th>
<th>Reference Data</th>
<th>User’s Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FS</td>
<td>FS 42 TP 11 AG 6 UB 0 O 3 W 1 Total 63</td>
<td>66.7</td>
</tr>
<tr>
<td>TP</td>
<td>TP 6 TP 38 AG 1 UB 0 O 0 W 0 Total 45</td>
<td>84.4</td>
</tr>
<tr>
<td>AG</td>
<td>AG 10 TP 5 AG 57 UB 24 O 16 W 0 Total 126</td>
<td>50.9</td>
</tr>
<tr>
<td>UB</td>
<td>UB 0 TP 0 AG 1 UB 27 O 0 W 0 Total 28</td>
<td>96.4</td>
</tr>
<tr>
<td>O</td>
<td>O 5 TP 5 AG 8 UB 1 O 33 W 0 Total 52</td>
<td>63.5</td>
</tr>
<tr>
<td>W</td>
<td>W 1 TP 0 AG 0 UB 0 O 0 W 49 Total 50</td>
<td>98.0</td>
</tr>
<tr>
<td>Total</td>
<td>Total 64 TP 59 AG 73 UB 52 O 52 W 50 Total 246</td>
<td>65.6</td>
</tr>
</tbody>
</table>

| Producer’s Accuracy (%) | 65.6 64.4 78.1 51.9 63.5 68.0 Overall Accuracy = 70.3 % Overall Kappa = 0.6395 |

Note: FS = Forest & Semi-natural Areas TP = Teak Plantation AG = Agricultural Areas UB = Urban & Built-up Areas O = Orchards W = Water Bodies

**Table 3** The classification error matrix shows the accuracy of teak plantation and other land cover types using expert classification

<table>
<thead>
<tr>
<th>Expert Classification</th>
<th>Reference Data</th>
<th>User’s Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FS</td>
<td>FS 51 TP 0 AG 2 UB 0 O 7 W 3 Total 63</td>
<td>80.9</td>
</tr>
<tr>
<td>TP</td>
<td>TP 2 TP 50 AG 4 UB 0 O 4 W 0 Total 60</td>
<td>83.3</td>
</tr>
<tr>
<td>AG</td>
<td>AG 5 TP 6 AG 64 UB 11 O 9 W 1 Total 96</td>
<td>66.7</td>
</tr>
<tr>
<td>UB</td>
<td>UB 0 TP 1 AG 1 UB 41 O 0 W 0 Total 43</td>
<td>95.4</td>
</tr>
<tr>
<td>O</td>
<td>O 5 TP 2 AG 2 UB 0 O 32 W 0 Total 41</td>
<td>78.1</td>
</tr>
<tr>
<td>W</td>
<td>W 1 TP 0 AG 0 UB 0 O 0 W 46 Total 47</td>
<td>97.9</td>
</tr>
<tr>
<td>Total</td>
<td>Total 64 TP 59 AG 73 UB 52 O 52 W 50 Total 284</td>
<td>79.7</td>
</tr>
</tbody>
</table>

| Producer’s Accuracy (%) | 79.7 84.8 87.7 78.9 61.5 92.0 Overall Accuracy = 81.1 % Overall Kappa = 0.7717 |

Note: FS = Forest & Semi-natural Areas TP = Teak Plantation AG = Agricultural Areas UB = Urban & Built-up Areas O = Orchards W = Water Bodies

The confusion of class discrimination between teak plantation and forest including semi-natural areas using the maximum likelihood classifier obviously caused the lower accuracy in teak plantation discrimination. Performing such classification technique could not effectively discriminate the teak plantation, due to the similarities of their spectral signatures. Similarly, these also occurred to the discrimination between agricultural, orchards and teak plantation classes. It is cleared that using only the spectral information and one statistical model-derived teak plantation class could not yield an acceptable accuracy. To reduce the misclassification, the decision trees and knowledge based construction from spatial analysis can be applied. In our experiment, the rule-based expert classifier revealed a significant improvement of teak plantation classification (from 64.4 to 84.8 percent).

It is noticeable that the incorporation of GIS and spectral data play significant roles to the accuracy improvement. For example, the DEM value, the slope and the boundary of conservative forest areas were applied to separate the known boundary of natural forest, which did not have teak plantation areas in the image. The process aids in
reducing the misclassification caused the spectral similarity of teak plantation and the natural forest classes. An example of the rules that was applied in the expert classifier is shown below:

Rule: If a pixel value is equal to conservative forest, then class = forest and semi-natural areas
   If a pixel value is not equal to conservative forest, then other combination of rules will be further applied.

Figure 5 shows an example of teak plantation classified image comparing the supervised with maximum likelihood classifier and expert classifier techniques. Observe that the classified image from the expert classification displays a smoother and a more homogenous area of the classified teak plantation. However, the classified teak plantation image using expert classifier still has some considerable misclassification pixels from orchards and agricultural classes, which should be further investigated in order to develop additional rules for the expert classifier technique.

The results of the kappa analysis that compared two different classification techniques are showed to be significantly different at the 95 percent confidence level. The $Z$ statistic value was greater than 1.96 (Table 4). The expert classification technique is efficient and, therefore should be considered as an acceptable classification technique for teak plantation mapping in the future. However, we believe that, the accuracy of teak plantation classified by this method can be improved by applying the multi-date Landsat data in the spatial analysis to construct knowledge-based for an expert classifier.

Figure 5 Comparisons between teak plantation identification from different classification techniques (the rectangle shows known teak plantation boundary obtained from field survey)
Table 4 Pairwise comparisons between supervised: maximum likelihood classifier and expert classification techniques for teak plantation mapping

<table>
<thead>
<tr>
<th>Pairwise Comparison</th>
<th>Z Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised: Maximum likelihood vs. Expert Classification</td>
<td>3.3342*</td>
</tr>
</tbody>
</table>

Note: * Z Statistic > 1.96: The two classification techniques are significantly different.

CONCLUSION

Our results indicated that using the expert classification technique with the incorporation of spectral data from Landsat TM and ETM+, and non-spectral data form GIS database, yielded satisfactory teak plantation classification accuracy of 84.8 percent, compared to 64.4 percent gained from the maximum likelihood classifier. The accuracy generated by expert method was also significantly different from the supervised with maximum likelihood classifier. Based on our study, the results encourage the use and the development of the effective algorithms based on expert system technique. With some fine-tuning the technique can further improve the accuracy of teak plantation mapping procedures. However, the complication of an analytical procedure, the processing time and cost of analysis may overcome the benefits. The knowledge and experience of the data analysts, as well as other factors should be taken into consideration when selecting an appropriate classification technique.

ACKNOWLEDGMENT

This research work is a part of ongoing research project “Analysis of Forest Ecosystems in Pasak Watershed Area Using Remote Sensing and Digital Image Analysis Techniques”, supported by Thammasat University and the National Research Council of Thailand.

REFERENCES


