LAND COVER MAPPING USING A NOVEL COMBINATION MODEL OF SATELLITE IMAGERIES: CASE STUDY OF A PART OF THE CAMERON HIGHLANDS, PAHANG, MALAYSIA

MOHAMMADI, A. 1* – BAHARIN, B. A. 1 – SHAHABI, H. 2

1Faculty of Built Environment and Surveying, Universiti Teknologi Malaysia (UTM)
81310 Johor Bahru, Malaysia
(e-mail: baharinahmad60@gmail.com)

2Department of Geomorphology, Faculty of Natural Resources, University of Kurdistan
Sanandaj, Iran
(e-mail: himanshahabi@gmail.com)

*Corresponding author
(e-mail: ayubmohammadi1990@gmail.com)

(Received 19th Oct 2018; accepted 22nd Dec 2018)

Abstract. Information about land covers is essential for a variety of purposes, such as environmental studies, sustainable development, and regional managements. This study aims to use a novel combination model to generate a land cover map in a part of the Cameron Highlands, Malaysia, where there are different kind of land covers, including tea plantation, florification and forest. Because of the high similarity in land covers of the study area, only through satellite imageries with a high spatial and spectral resolution the land covers can be differentiated. We have combined satellite imageries of Sentinel-1 (S1A, GRD, IW) and Landsat-8 (Operational land imager) for the year 2017 as well as different algorithms of Maximum Likelihood (ML), Minimum Distance (MD), Support Vector Machine (SVM), Spectral Angle Mapper (SAM) and Artificial Neural Network (ANN). The results showed that the combination model is an applicable technique for extracting land covers in areas with high similarities in land covers. The overall accuracy of the confusion matrix and the Kappa Coefficient are 98.1984% and 0.9579, respectively, which indicate that it is a robust model for extracting land covers in areas like the Cameron Highlands. The obtained results can be useful for different purposes, including urban and environmental management, change detection, agriculture and many more purposes.

Keywords: land use, satellite data, remote sensing, geographic information system, Cameron Highlands

Introduction

Information about Land Covers (LC) and changes affecting them is essential for many purposes, including urban planning, vegetation and environmental issues (Shahabi et al., 2012; Chen et al., 2015; Gómez et al., 2016). Combination model refers to the ability of using two or more satellite imageries, in which we can improve the spectral resolution to detect the land covers more precisely. In recent years, database and methodology for creating a land cover map have undergone considerable changes (Tehrany et al., 2014; Megahed et al., 2015). Advancements in Remote Sensing (RS) technology and also a demand for using them for different issues have made it a the most used technology for gaining information about earth surface objects since 2 decades ago (Walker and Blaschke, 2008; Belward and Skøien, 2015; Ma et al., 2017).

Land cover has been undergoing considerable changes in either local or global scales (Megahed et al., 2015; Miettinen et al., 2016). Land cover mapping is not an easy issue and needs considerable attention regarding both data collection and model selection (Megahed et al., 2015; Khatami et al., 2016). Selecting data sets should be based on the
study area and the kinds, shapes and color of the land covers, which help us to select appropriate satellite imagery (Moser et al., 2013; Tehrany et al., 2014; Hasmadi et al., 2017). Since the study area is one of the touristic destinations for both locals and foreigners, so its preservation for Malaysian government is vital.

One of the highlighted issues when generating a land cover map from satellite imagery is the spectral and the spatial resolutions (Griffiths et al., 2013; Poursanidis et al., 2015). However, the precise information about land cover is highly essential. Until about 15 years ago land cover data bases were in form of pre-existing maps, gathered from field survey and previous maps (Griffiths et al., 2013). From two decades ago onwards, RS and Geographical Information System (GIS) have drawn researcher’s attention toward using satellite imagery and machine learning algorithms (Loveland et al., 2000; Tehrany et al., 2014).

Many studies have been conducted with the land cover mapping using different models, techniques and satellite imageries, including global land cover map using Landsat satellite imageries and SVM algorithm (Gong et al., 2013); object-based and pixel-based approaches using SPOT-5 satellite imagery (Tehrany et al., 2014); airborne LiDAR data (Yan et al., 2015); different algorithms of Savitzky-Golay, asymmetric Gaussian, double-logistic, Whittaker smoother and discrete Fourier using MODIS imagery for Great Lake, USA (Shao et al., 2016); supervised and unsupervised algorithms using SPOT-5 satellite imagery (Hasmadi et al., 2017); hyperion data using SVMs classifier (Lamine et al., 2018).

The main objective of the current study was to use a novel combination model for extracting land covers in a part of the Cameron Highlands, Pahang, Malaysia. However, this study dealt with two significant combinations of satellite imageries (Landsat-8 and Sentinel-1), and also combination of different algorithms of ML, MD, SVM, SAM, and ANN by using a Decision Tree (DT) model.

Materials and methods

Description of the study area

The study area is a part of the Cameron Highlands, Pahang, Malaysia between longitudes 101° 20’ 00” E to 101° 27’ 10” E and latitudes 4° 23’ 30” N to 4° 31’ 10” N (Fig. 1). This scope with an area of 81.249 km² is located approximately in the south-western part of the Cameron Highlands (Tien Bui et al., 2018). The lowest and the highest area of the study area is 912 and 1960 m above the sea level, respectively. However, the average rainfall fluctuates between 1800 mm to 3000 mm annually (Jebur et al., 2015; Tien Bui et al., 2018).

Data collection

Sentinel-1 is a part of the Copernicus Programme conducted by the European Space Agency (ESA). This product consists of the two satellite platforms of 1A and 1B. It has four sensor modes (Strip Map (SM), Interferometric Wide Swath (IW), Wave (WV), and Extra Wide Swath (EW)), and three product types namely, Ground Range Detected (GRD), Single Look Complex (SLC), and Ocean (OCN) (Potin et al., 2012). In this research, we have acquired Landsat-8 and Sentinel-1 satellite imageries for the year 2017 online for free from www.earthexplorer.usgs.gov and www.scihub.copernicus.eu,
respectively. Table 1 and Figure 2 illustrate the technical characteristics of satellite data and geographical position of the study area on the satellite imageries, respectively.

**Research methodology**

Using Sentinel Application Platform (SNAP) software and Environment for Visualizing Image (ENVI) software, either images radiometrically, spectrally and geometrically were corrected, co-registered and stacked. Using pan-sharpening command in ENVI software, the spatial resolution of Landsat-8 improved to 15 m by its panchromatic band, then via co-registration process the spatial resolution of Landsat-8 (Slave image) enhanced to 10 m by Sentinel-1 as the master image.

![Figure 1. Geographical position of the study area](image)

<table>
<thead>
<tr>
<th>No.</th>
<th>Sentinel-1</th>
<th>Landsat-8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensor</td>
<td>Synthetic Aperture Radar (SAR)</td>
<td>Operational Land Imager (OLI)</td>
</tr>
<tr>
<td>Spatial resolution</td>
<td>10 m</td>
<td>-30 m (Multispectral) -15 m (Panchromatic)</td>
</tr>
<tr>
<td>Temporal resolution</td>
<td>12 days</td>
<td>16 days</td>
</tr>
<tr>
<td>Spectral resolution</td>
<td>Single band (C)</td>
<td>9 bands</td>
</tr>
<tr>
<td>Date</td>
<td>06/09/2017</td>
<td>23/08/2017</td>
</tr>
<tr>
<td>More information</td>
<td>- Interferometry Wide swath (IW) - Ground Range Detected (GRD)</td>
<td></td>
</tr>
</tbody>
</table>
After picking up the Region of Interests (ROIs) from co-registered image, different algorithms of maximum likelihood, minimum distance, support vector machine, spectral angle mapper and artificial neural network, were applied for classification. Each algorithm has shown a 1 or 2 land covers better, therefore in order to use the information of all algorithms, the DT method was used. Finally, after the post classification process, to evaluate the overall accuracy some Ground Control Points (GCPs) were extracted using Google Earth (For water body) and field checks for the other land covers (Fig. 3). Using ArcGIS the raster map was converted into a vector file and then was dissolved as the number of land covers. Moreover, after calculating the area for each land cover, the attribute table was imported into Microsoft Excel for calculating the total area and creating the final output table.

**Description of the used algorithms**

**Maximum likelihood (ML)**

Maximum likelihood model assumes that the statistics for each class are normally distributed and measures the probability that a given-pixel belongs to a certain class (Stamatakis, 2006). Each pixel is assigned to a class, which has the highest probability (Richards and Richards, 1999; Stamatakis, 2006). Once the highest probability has been smaller than the threshold, then the pixel remained unclassified (Richards and Richards, 1999).

**Minimum distance (MD)**

The minimum distance classification utilize the mean vector of every endmember and measures the Euclidean distance from unknown pixels to the mean vector for each class (Abtan et al., 2017; Kolesár, 2018). Based on the specified standard deviation, number of classes and a distance threshold, all pixels will be classified to the nearest class, however, some pixels may remain unclassified if they do not meet the selected criteria.
Neural network (NN)

The neural network technique implies standard back-propagation for a supervised learning (Yoshida and Omatu, 1994). You can choose a number of hidden layers to use and also you can select between hyperbolic or logistic activation functions (Zhang and Foody, 2001). The NN algorithm can be used to perform non-linear classification (Scott et al., 2017). In this model, the error is back-propagated through the network and weight adjustment is made by using a recursive method (Scott et al., 2017; Chakraborty and Roy, 2018).

Support vector machine (SVM)

The SVM is a fast and accurate model for the land cover/land use studies (Sukawattanavijit et al., 2017). This model provides a good result from noisy and
complex data (Gu and Sheng, 2017). It tries to separate the created classes with a decision surface, which maximizes the margin between the selected classes (Mohammadi et al., 2018). Despite this fact that the SVM is a binary classifier, it works as a multiclass classifier as well, provided that several binary SVM classifiers are combined (Gu and Sheng, 2017; Mohammadi et al., 2018). The classifier performs training at a lower resolution level, because retraining at each level provides higher accuracy for the resampled image, however, it considers the rule image values to identify those that exceed the probability threshold (Sukawattanavijit et al., 2017; Mohammadi et al., 2018).

**Spectral angle mapper (SAM)**

The SAM classifier is a physically-based classification that utilize an n-D angle to match pixels to reference spectra (Renza et al., 2017). The model determines the spectral similarity between two different spectra by calculating the angle among the spectra (Kumar et al., 2015). End-member spectra used by SAM model can come from ASCII files, spectral libraries, and you can extract them from Region of Interests (ROIs) (Renza et al., 2017; Yan and Roy, 2018). SAM compares the angles between the end-member spectrum vector and each pixel vector in n-D space (Yan and Roy, 2018). However, smaller angles represent a closer match to the reference spectrum.

**Decision tree (DT)**

A DT is a type of multistage classifier, which can be applied to a single imagery or a stack of images (Chasmer et al., 2014). It comprises of a series of binary decisions, which are applied to determine the correct group for each pixel (Hua et al., 2017). It can be based on the available characteristics of the dataset (Lu et al., 2014). As an example, you may have an elevation dataset and two multispectral imageries that collected at different times, and any of those images can contribute to decision within the same tree.

**Validation**

The confusion matrix is important to validate statistical classification (Stehman, 1997). Overall accuracy is calculated by measuring the number of corrected classified pixels then dividing them by the total number of pixels (Eq. 1) (Jensen and Lulla, 1987). Kappa coefficient is measured through multiplying the total pixel classes by the sum of the confusion matrix, then subtracting the sum of ground truth pixels in a class times the sum of classified pixels in that class summed over all classes (Eq. 2).

\[
OA = \frac{1}{N} \sum P_{ii}
\]  
(Eq.1)

where: OA = total accuracy, N = total number of test pixels, and \( \sum P_{ii} \) = total pixels that are correctly classified.

Equation for the Kappa coefficient is written as below:

\[
K = (OA - \frac{1}{q})(1 - \frac{1}{q})
\]  
(Eq.2)

where: K = kappa coefficient and q = unclassified pixels.
Results and discussion

One of the most important steps in extracting ROIs is the separability assessment of them (Laurin et al., 2013; Megahed et al., 2015; Levin, 2016). However, the value for measuring the ROIs separability is between 1 and 2, the values less than 1.8 are not acceptable (Laurin et al., 2013). The higher the value the more precise the ROIs are (Levin, 2016). Table 2 illustrates the least to most separability values of extracted ROIs for this study.

Table 2. Pair separation of ROIs (least to most)

| Vegetation/florification and township = 1.99 | Water body and cleared forest = 1.99 |
| Vegetation/florification and tea plantation = 1.99 | Township and tea plantation = 1.99 |
| Road network & vegetation/florification = 1.99 | Cleared forest & vegetation/florification = 2 |
| Road network and township = 1.99 | Forest and water body = 2 |
| Forest and township = 1.99 | Forest and road network = 2 |
| Cleared forest and township = 1.99 | Forest and cleared forest = 2 |
| Water body and township = 1.99 | Water body and tea plantation = 2 |
| Water body and road network = 1.99 | Cleared forest and tea plantation = 2 |
| Forest and vegetation/florification = 1.99 | Forest and tea plantation = 2 |
| Water body & vegetation/florification = 1.99 | Cleared forest and road network = 2 |
| | Road network and tea plantation = 2 |

It is worth mentioning that the pair separation of the extracted ROIs from the combination of Sentinel-1 and Landsat-8 satellite imageries showed that this model is a robust way in extracting ROIs for generating land cover maps (Fig. 4).

Figure 4. Different land cover’s tones using combination model
As a matter of fact, the land covers were extracted based on the ROIs extraction, where every land cover has its own reflection in the satellite imagery. If we do not extract the ROIs for one single land cover with a different reflection, it will remain unclassified. In this study, because of the different reflection in the satellite imagery, the tea plantation fields are separated from the vegetation coverage, spatially and statistically. Overall, a total number of 6 land covers were extracted for the study area (Table 3 and Fig. 5). Using ArcGIS and Microsoft Excel the area of extracted land covers were calculated and finalized. The vegetation and florification fields as well as the forest areas have the biggest area among all of 39.13176 km² and 33.90341 km², respectively. However, water body has the smallest area of only 0.364556 km².

### Table 3. The area of extracted land covers

<table>
<thead>
<tr>
<th>No</th>
<th>Name</th>
<th>Area km²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Forest</td>
<td>33.90341</td>
</tr>
<tr>
<td>2</td>
<td>Water body</td>
<td>0.364556</td>
</tr>
<tr>
<td>3</td>
<td>Tea plantation</td>
<td>3.024364</td>
</tr>
<tr>
<td>4</td>
<td>Township</td>
<td>4.139532</td>
</tr>
<tr>
<td>5</td>
<td>Vegetation and florification</td>
<td>39.13176</td>
</tr>
<tr>
<td>6</td>
<td>Cleared forest</td>
<td>0.436157</td>
</tr>
</tbody>
</table>

**Figure 5. Land cover map of the study area**
Validation

Validation of land cover map is important to show the quality of used methods and techniques (Chen et al., 2017; Ma et al., 2017). Using a handheld Geographic Positioning System (GPS), a total number of 20 Ground Control Points (GCPs) were picked up for each land cover (Except the water body, where the Google Earth was employed for extracting GCPs for) by Ayub Mohammadi on Thursday 19 July, 2018 (Fig. 6). Because of the hilly situation of the Cameron Highlands, the GCPs were extracted near the main road and the accessible areas only. However, these GCPs were applied for validation purposes using the overall accuracy and Kappa statistic. Table 4 lists the result of the validation.

Comparison of the combination model with a single model

Using Enhanced Thematic Mapper plus (ETM+) satellite imagery for the date 06/09/2017 and Support Vector Machine (SVM) model, the land covers of the study area were extracted to compare with the combination model. Based on the results there are not too much differences between the two models statistically. The most considerable change has occurred inside the tea plantation area, where in the single model about 0.600 km$^2$ was added to the tea fields. At the same time, roughly 1.80 km$^2$ was decreased from the vegetation and florification areas (Table 5).

![Figure 6. Shows the geographical positions of the extracted GPS points](image)

**Table 4. Highlights the confusion matrix for this study**

<table>
<thead>
<tr>
<th></th>
<th>Overall accuracy</th>
<th>Kappa coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>98.1984%</td>
<td>0.9579</td>
</tr>
</tbody>
</table>
Table 5. Represents the area for extracted land covers from ETM+ and SVM model

<table>
<thead>
<tr>
<th>No.</th>
<th>Name</th>
<th>Area km²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cleared forest</td>
<td>0.429154</td>
</tr>
<tr>
<td>2</td>
<td>Forest</td>
<td>34.1433</td>
</tr>
<tr>
<td>3</td>
<td>Tea plantation</td>
<td>3.623</td>
</tr>
<tr>
<td>4</td>
<td>Township</td>
<td>4.63918</td>
</tr>
<tr>
<td>5</td>
<td>Vegetation and florification</td>
<td>38.0521</td>
</tr>
<tr>
<td>6</td>
<td>Water body</td>
<td>0.364556</td>
</tr>
</tbody>
</table>

As a matter of fact, compared to the Google Earth images (Fig. 7), the combination model has shown the final map more realistic. In the output map of the single model the land covers are seen more unrealistic (especially for the vegetation and florification fields; Fig. 8).

Conclusion

Information about land covers is essential for environmental issues. Combination model refers to using the ability of two or more satellite imagery or even different algorithms, in which we can improve the spectral resolution of them to detect the land covers more precisely. In this study, a novel combination model of satellite imageries of Sentinel-1 and Landsat-8 as well as the combination of different algorithms namely: maximum likelihood, minimum distance, support vector machine, spectral angle mapper, and artificial neural network were used for extracting land covers in a part of the Cameron Highlands, Pahang, Malaysia.

Figure 7. Shows the study area on the Google Earth image
However, a total number of 6 land covers were mapped, which using the Google earth and field survey all land covers were validated. Eventually, by a set of ground control points (GCPs), overall accuracy and Kappa statistic were measured of 98.1984% and 0.9579, respectively. It is noticed that the outputs of this study have promising results for multi-temporal monitoring of land covers in the Cameron Highlands for long term analysis. The obtained results showed that dual Sentinel-1 (S1A, GRD, IW) and Landsat-8 Operational Land Imager (OLI) are suitable satellite imagers for rapid land cover extraction in the geographical situation with high similarity in land covers.

Acknowledgements. The authors wish to express their sincere thanks to Universiti Teknologi Malaysia (UTM) based on Research University Grant (Q.J130000.2527.17H84) for their financial supports.

REFERENCES


