THE USE OF RANDOM FOREST AND OBJECT-BASED IMAGE CLASSIFICATION FOR THEMATIC MAP CREATION: A CASE STUDY OF KELANTAN, MALAYSIA

Bahareh Kalantar*[1], Ahmad A. A. Moneir[1], Shattri Bin Mansor[1], Biswajeet Pradhan[1]

[1] Department of Civil Engineering, Faculty of Engineering, University Putra Malaysia, 43400, Serdang, Selangor Darul Ehsan, Malaysia,

Email: Bahare_kgh@yahoo.com

KEY WORDS: Random forest, object-based image analysis, thematic map, classification.

ABSTRACT: This paper presented results of creation of thematic map for Kelantan city using an integration of object-based analysis and random forest algorithm. The thematic map was generated based on SPOT image with a spatial resolution of 5 m. Object-based analysis was used for its advantages over per-pixel based methods. On the other hand, random forest algorithm was selected because it was suggested in very recent literature and it has advantages of good generalization. A thematic map with six classes was generated with an overall accuracy and kappa index of 92% and 0.90, respectively using the proposed method. Other details are included in the paper.

1. INTRODUCTION

Image classification is one of the important tasks in remote sensing providing information for several geospatial applications such as landuse planning, hazard and risk assessments of natural disasters, transportations, and land monitoring. In general, there are two types of image classification, object-based and per-pixel methods. In object-based method, first the input image is segmented into image objects by grouping image pixels using spectral and spatial information contained in the image. On the other hand, per-pixel based techniques are grouping image pixels either automatically by computer algorithms or by supervised methods. Literature suggests that object-based methods are better than per-pixel methods for high-resolution images as spatial information could be utilized. Other advantages of object-based methods include less affected by noise, final products are GIS-ready files, and results of classification are high quality.

In object-based image classification, there are two possibilities to generate thematic maps, rule-based and example-based approaches. Rule-based methods are suggested for specific applications while example-based methods are good for general landcover mapping. In supervised classification, several algorithms can be utilized such as k nearest neighbor, support vector machines, decision trees, and random forest. The last one has advantages of good generalization and producing relatively high quality classification results compared to other techniques. In this study, random forest is used with object-based image analysis for producing thematic maps for Kelantan city located in west of Malaysia.

Numerous studies investigated landcover mapping using object-based image analysis. For example, Guan et al. (2013) explored the integration of orthoimagery and lidar data for object-based urban thematic mapping using random forests. The results of their study presented that object-based classification method, compared with the pixel-based classification, improves by 0.02 and 0.05 in kappa statistics, and by 3.9% and 4.5% in overall accuracy, respectively. Duro et al. (2012) used object-based image analysis and feature selection of multi-sensor earth observation imagery for landcover mapping using random forests. Results indicated that object-based approaches performed well and an overall accuracy of >85% was achieved. The study also discussed that the method increased interpretability of classification models due to the feature selection process and the use of variable importance scores generated by the RF algorithm. In another study, Li and Shao (2011) demonstrated an object-oriented method for detailed urban vegetation delineation by using 1 m resolution, four-band digital aerial photography as the only input data. A hierarchical classification scheme was developed to discriminate vegetation types at both coarse and fine scales. More recently, Li et al. (2016) investigated a systematic comparison of different object-based classification techniques using high spatial resolution imagery in agricultural environments. Results indicated that random forest and support vector machines are highly suitable for object-based classifications in agricultural areas and confirm the expected general tendency, namely that the overall accuracies decline with increasing segmentation scale.

2. STUDY AREA

Kelantan is positioned in the northeast of Peninsular Malaysia (Figure 1). It is bordered by Narathiwat Province of Thailand to the north, Terengganu to the southeast, Perak to the west, and Pahang to the south. To the northeast of Kelantan is the South China Sea. Kelantan is located in the northeastern corner of the peninsula. Kelantan, which is said to translate as the "Land of Lightning" (see alternate theories below), is an agrarian state with lush paddy fields, rustic fishing villages and casuarina-lined beaches. Kelantan is home to some of the most ancient archaeological discoveries in Malaysia, including several prehistoric aboriginal settlements. Due to Kelantan's relative isolation and largely rural lifestyle, Kelantan's culture differs somewhat from Malay culture in the rest of the peninsula; this is reflected in the cuisine, arts, and the peculiar Kelantan's Malay language, which is unintelligible even for some speakers of standard Malay.

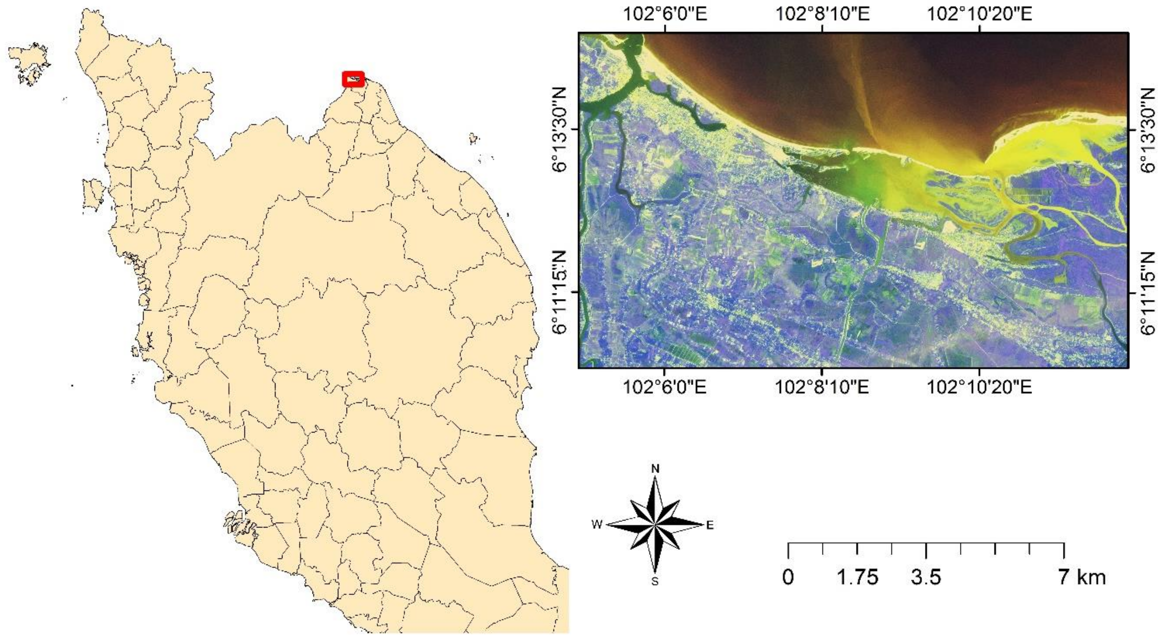

Figure 1. Location of study area used in the current research.

3. METHOD

The overall workflow of classification conducted in this study is presented in Figure 2. In the first step, the SPOT image of Kelantan was pre-processed by two main steps, radiometric calibration, and geometric correction. Then, classification scheme was designed based on Google maps. After that, the processed image was segmented by using multiresolution segmentation algorithm. Once image objects were created, an adequate numbers of segments were selected for training and testing of random forest algorithm. Next, random forest algorithm was used to classify the image objects and thematic map of the study area was produced and exported into GIS. Finally, accuracy assessment of the final image classification was calculated and thematic map was produced as its final version for end users.

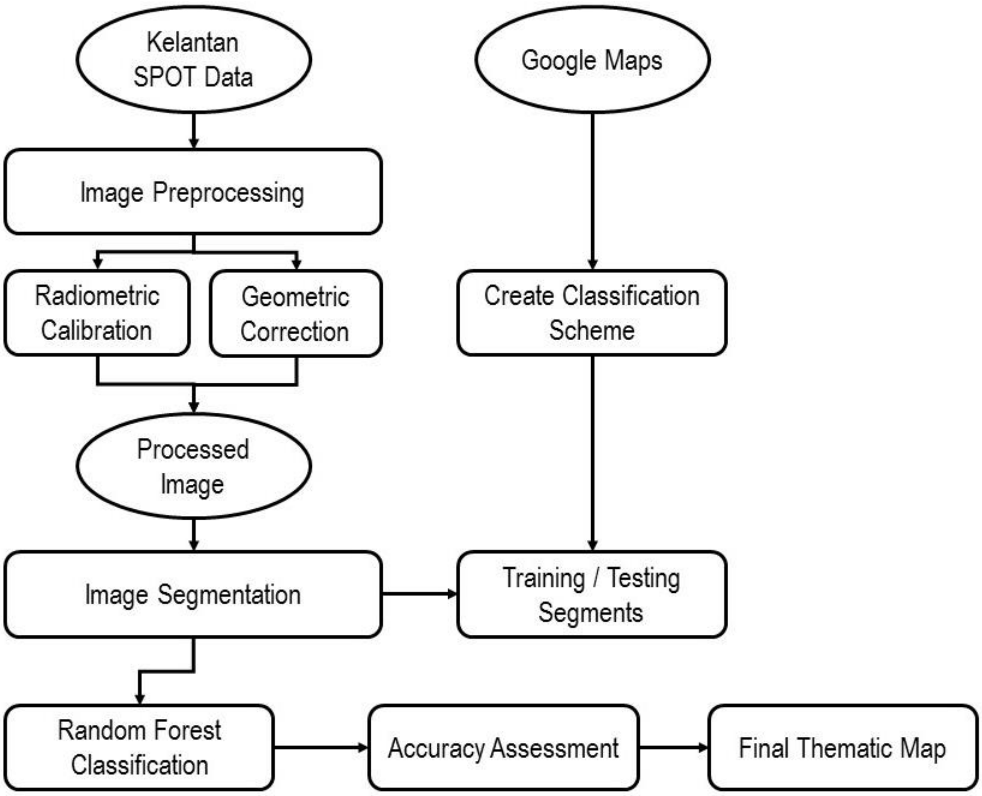

Figure 2. Overall workflow of image classification used to produce a thematic map for Kelantan city.

3.1 Image Pre-processing

Satellite images are usually containing noise due to malfunction and other errors. These errors must be removed before performing any actual analysis so that better results could be obtained. First, the SPOT image was radiometrically calibrated in a process of converting the digital numbers of the original image into a radiance image and then into reflectance image. Once the image was corrected and light, environment, and atmosphere effects were reduced, it was exported into GIS. Based on Google maps, the image was geometrically corrected using 25 ground control points selected in clearly identifiable locations in the study area. Results of geometric correction indicated an RMSE of less than of 1.00 so that the image was prepared for actual processing. The processed data is shown in Figure 3.

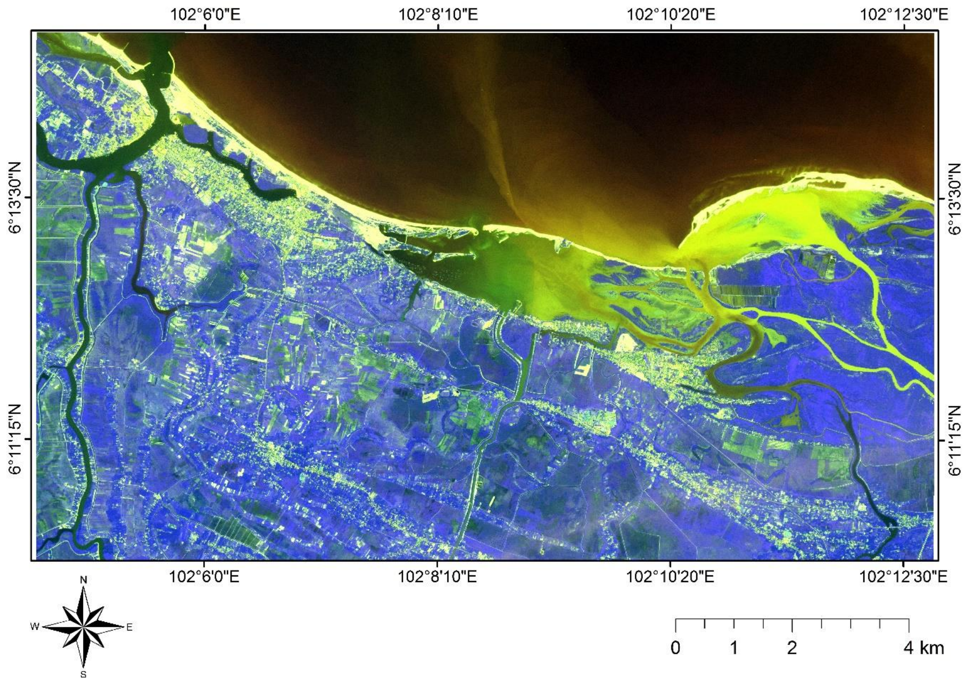

Figure 3. The processed SPOT data.

3.2 Image Segmentation

In object-based image analysis, the first step is the image segmentation. In this study, the processed SPOT image was segmented in eCognition software using a well-known multiresolution segmentation algorithm. Parameters of this algorithm were selected by trial and error method. Parameters that were selected are as follow: scale=250, shape=0.2 and compactness= 0.5. After the selection of user defined parameters, image objects were created and attributes were calculated. Image objects with their attributes were ready for classification.

3.3 Image Classification

Once image objects were created, 25 samples for each class were selected as training data. These training objects were used to train the random forest algorithm. The classification algorithm was applied to the whole study area and all image objects were then classified into six classes, ocean, inland water, urban, wet land, cropland, and coastal areas. These classes were selected based on the analysis of classification scheme conducted using Google maps.

4. EXPERIMENTAL RESULTS

4.1 Results of Image Segmentation

Results of image segmentation are shown in Figure 4. Results indicate good segmentation as image objects well defining each object in reality. For example, the ocean was created as one object as clearly can have been from the results presented in Figure 4. Creation of accurate segmentation is very important for accurate classification results. This is because attributes including spectral and spatial then could be used to better classify these image objects. Some undersegmentation and oversegmentation were occurred in the results, mostly, in urban and wet land classes.

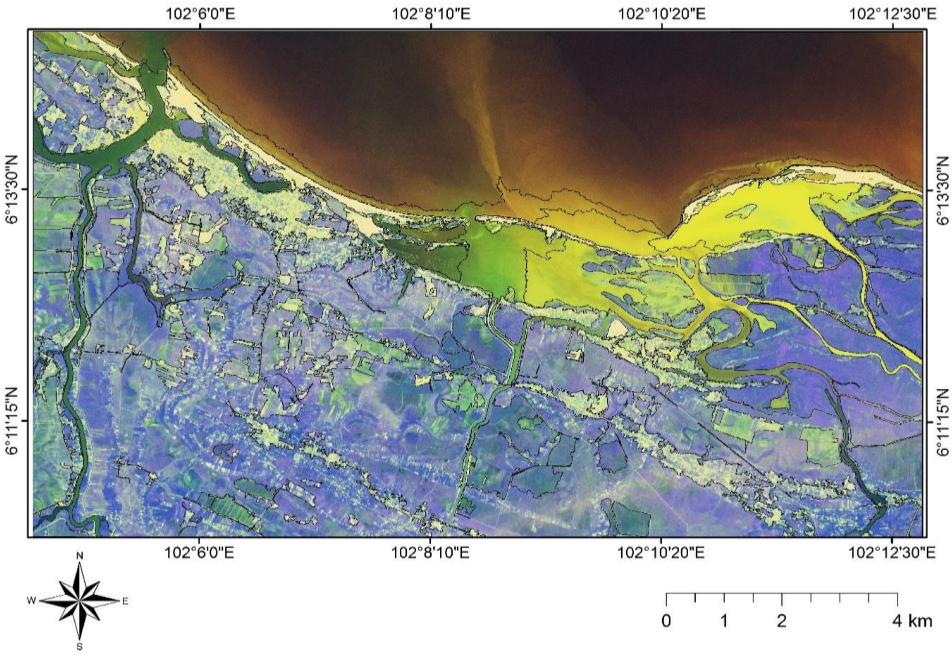

Figure 4. Results of image segmentation.

4.2 Results of Image Classification

Results of image classification are shown in Figure 5. In this result, we can see the accurate thematic map as generated for the study area. The first indication could be made is that the results show smooth classification results and boundaries between classes were clearly detected and identified. Some misclassifications were occurred in wet land and urban classes, as they were difficult to identify them in the SPOT image.

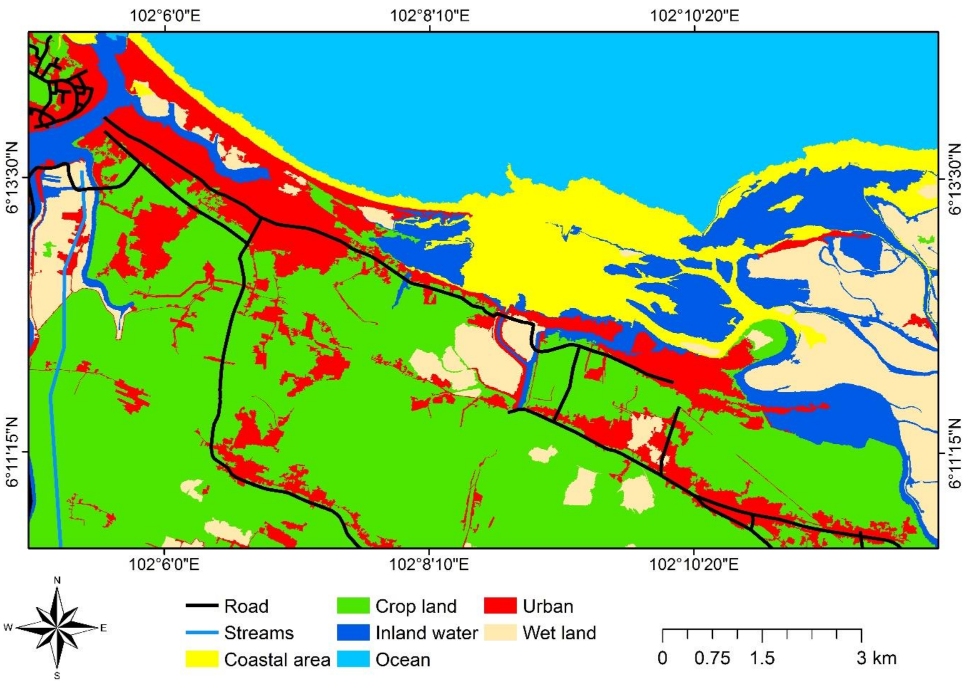

Figure 5. Thematic map of Kelantan city produced with integration of object-based analysis and random forest.

4.3 Accuracy Assessment

Results of accuracy assessment are shown in table 1. Overall accuracy of 92% and kappa index of 0.92 were obtained by the use of the proposed method. The highest and lowest user's accuracy of 100% and 85% were achieved for ocean and inland water classes, respectively.

Table 1 Accuracy assessment of the produced thematic map.

User class/ Sample	Inland water	Urban	Wet land	Crop land	Ocean	Coastal area	Sum
Confusion matrix							
Inland water	23	0	3	0	0	1	27
Urban	0	41	0	3	0	0	44
Wet land	0	1	27	0	0	1	29
Crop land	0	0	1	17	0	0	18
Ocean	0	0	0	0	1	0	1
Coastal area	1	0	0	0	0	21	22
Unclassified	0	0	0	0	0	0	0
Sum	24	42	31	20	1	23	
Accuracy							
Producer	0.95	0.97	0.87	0.85	1	0.913	
User	0.85	0.93	0.93	0.94	1	0.95	
Hellden	0.90	0.95	0.9	0.89	1	0.933	
Short	0.82	0.911	0.818	0.809	1	0.87	
KIA per Class	0.94	0.96	0.83	0.828	1	0.89	
Totals							
Overall Accuracy	0.922						
KIA	0.90						

4. CONCLUSION

This study proposed an object-based classification workflow for Worldview-2 data. First, the image was pre-processed, and supplementary data from Google maps and Open Street Maps were gathered. These supplementary data were used to enhance the results of thematic map of the study area and they used for selecting training and testing samples. After that, the image was segmented and classified by SVM method. Finally, a thematic map was produced for the study area in ArcGIS software. The thematic map was also validated and results indicated the robustness of the method as the overall accuracy was over 85%. Findings suggest that OBIA is a suitable method for producing thematic maps using very high-resolution images. Future works should focus on optimization of parameters and enhancing results of segmentation.

REFERENCES

Kavzoglu, T., Colkesen, I., & Yomralioglu, T. (2015). Object-based classification with rotation forest ensemble learning algorithm using very-high-resolution WorldView-2 image. Remote Sensing Letters, 6(11), 834-843.

Benarchid, O., & Raissouni, N. (2013). Support Vector Machines for Object Based Building Extraction in Suburban Area using Very High Resolution Satellite Images, a Case Study: Tetuan, Morocco. IAES International Journal of Artificial Intelligence, 2(1).

Aguilar, M. A., Saldaña, M. M., & Aguilar, F. J. (2013). GeoEye-1 and WorldView-2 pan-sharpened imagery for object-based classification in urban environments. International Journal of Remote Sensing, 34(7), 2583-2606.

Petropoulos, G. P., Kalaitzidis, C., & Vadrevu, K. P. (2012). Support vector machines and object-based classification for obtaining land-use/cover cartography from Hyperion hyperspectral imagery. Computers & Geosciences, 41, 99-107.