Comparison of two classification methods (MLC and SVM) to extract land use and land cover in Johor Malaysia

B Rokni Deilmai[1], B Bin Ahmad[1], H Zabihi[2]
[1]Department of Remote sensing, Faculty of Geoinformation and Real Estate, Universiti Teknologi Malaysia, 81310 Johor, Malaysia
[2] Department of Geoinformation, Faculty of Geoinformation and Real Estate, Universiti Teknologi Malaysia, 81310 Johor, Malaysia

E-mail: bashirrokni@yahoo.com

Abstract. Mapping is essential for the analysis of the land use and land cover, which influence many environmental processes and properties. For the purpose of the creation of land cover maps, it is important to minimize error. These errors will propagate into later analyses based on these land cover maps. The reliability of land cover maps derived from remotely sensed data depends on an accurate classification. In this study, we have analyzed multispectral data using two different classifiers including Maximum Likelihood Classifier (MLC) and Support Vector Machine (SVM). To pursue this aim, Landsat Thematic Mapper data and identical field-based training sample datasets in Johor Malaysia used for each classification method, which results indicate in five land cover classes forest, oil palm, urban area, water, rubber. Classification results indicate that SVM was more accurate than MLC. With demonstrated capability to produce reliable cover results, the SVM methods should be especially useful for land cover classification.

1. Introduction

Land use and land cover are the basic assumption to identify the global ecology or environmental changes.Land use/ land cover mapping is a vital component, which has various parameters that are integrated on the basis of requirement. Land cover refers to a water body, cultivated land, built-up, natural vegetation, fallow land, glacial, rock/soil, artificial cover and others observed on the land. Global Information monitoring of land use and land cover is possible due to remote sensing technology in the form of spatial, spectral and temporal resolution. Remote sensing technology has many important roles, like reduction of survey time, latest map availability, more economic, digital image classification (pixels), spectral information etc. Image classification is an important part in many remote sensing applications. Classification of land cover is one of the most important task and one of the primary objectives in the analysis of remotely sensed data [16].
In this study, two different classification methods were compared for evaluating land cover features in the Johor Malaysia.

A Maximum Likelihood classifier, representing a conventional classification method and Support Vector Machines were applied. Maximum Likelihood is one of the most popular supervised

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

classification method used with remote sensing image data. The Maximum Likelihood classification method is well known for the analysis of satellite images [5]. So far, satellite image interpretation using maximum likelihood approach was mostly applied for land cover classification [6] and monitoring of land use changes [17] showing overall high accuracies (mostly over 80%). MLC classification is based on parametric approach that involves assumption of the selected classes of signature in the normal distribution [1].

Improving land use/cover classification accuracy is an important issue in remote sensing literature. Moreover, advanced classification algorithm including Neural Network and Support Vector Machine approach instead of conventional classification method have been developed recently.

Support vector machines (SVMs) are a group of supervised classification algorithms that have been used recently in the remote sensing field. The classification accuracy produced by SVMs may show variation depending on the choice of the kernel function and its parameters [7].

Szuster et al. (2011) examined the performance of the SVM classification technique in the tropical coastal zone. He also compared this technique with the MLC and Artificial Neural Network techniques and has concluded that the SVM is better classifier for challenges like separating human infrastructures, for instance buildings from rocks and sandy beaches as they possess similar spectral signatures.

Yu et al. (2012) has applied the support vector machine (SVM) algorithm for the automated lithological classification in a part of northwestern India using ASTER imagery. He has also showed that SVM gives higher accuracy in comparison to MLC. The SVM algorithm has been used widely for pattern recognition applications. Many researchers in this field have found that a higher level of accuracy can be achieved by SVM than other processes of classification like MLC, artificial neural network (NN) etc. [11, 12, 14 and 15].

2. Data and methodology

2.1Study area

The study area considered in this study is the entire state of Johor. Johor is one of the developed states in Peninsular Malaysia In term of the area and population, Johor is the fifth largest and second populous state in Malaysia, With a total area of 19,210 km2 and population of 3,233,434 in 2010 [22]. The largest land uses in Johor are oil palm and forest (Figure 1).

Figure1.Studyarea showing the state of Johor [23]

2.2 Data used

Landsat TM satellite imagery has been selected for the classification of Land use by MLC and SVM algorithms. Date of acquisition of satellite imagery with spatial resolution of 30m has been 10

May, 2009 available from the Earth Explorer website [24]. It is corrected for geometric and topographic errors (level 1T).

2.3 Methods

Image classification is the stage of image analysis in which the multivariate quantitative measurement associated with each pixel is translated into a label from a pre-defined set (e.g., land use categories). Recall that the intent of the classification process is to assign pixels in an image to several classes of data based on their pixel value. Remote sensing classification is a complex process and requires consideration of many factors. The major steps of image classification may include determination of a suitable classification system, selection of training samples, image pre-processing, feature extraction, selection of suitable classification approaches, post-classification processing and accuracy assessment. Before classification, the land cover types in the study area were defined with the help of a land use map produced by the department of Agriculture Malaysia (year 2008). The main land cover types are forest, oil palm, urban area, rubber and water bodies.

Two types of classification techniques have been applied for analyzing the accuracy of land use map by using corrected satellite imagery. Training areas were selected according to the land use map of years 2006. Training stage is the most important part in supervised classification because it could influence the final classification results.

Maximum Likelihood classifier is the most common method used in the applications of remote sensing as a parametric statistical method where the analyst supervises the classification by identifying representative areas, called training zones. MLC is a process for determination of known class distribution as the maximum for a given statistic. This classification is a standard pixel based technique which is based on a multivariate probability density function of classes [9].

Early development of SVM started in 1970s and the popularity of SVM for pattern recognition and classification is actually surged in the late 1990s [19, 20]. The primary objective of the SVM method is the generation of a hyperplane that represents the optimal separation of linearly-separable classes in decision boundary space. In remote sensing, SVM was primarily used for the hyperspectral image classification and object detection [3, 10], although researchers have recently expanding its application for multispectral remote sensing data [2, 5 and 15]. The primary advantage of SVM was good generalization capability with limited training samples. The authors acknowledged SVM's limitations in parameter selection and computational requirements. However, SVM provided superior performance compared to most other image classification algorithms for both real-world remote sensing data and simulated experiments.

In general, the SVM represents a novel approach compared to conventional MLC classifiers. Additional SVM applications for regional scale land cover classification need to be conducted to better understand performance.

After classification, the accuracy of the classified images was assessed using reference data (land use map of year 2006). A total of 250 random points were selected from the images generating via a stratified random sampling method.The accuracy was assessed using error matrices (overall, user's and producer's accuracies and Kappa statistics.

3. Results and Analyisis

3.3 Landuse/landcover classification

The final LULC (Land use and land cover) maps presented in figure 2 show that the major classes are forest, oil palm, rubber, city and water. An evaluation of accuracy of the classified images (table 1) shows that the overall kappa and overall accuracy for Support Vector Machine (SVM) classification is 0.86 and 91.67%, and it is higher than the pixel based classification results with overall accuracy and Kappa 78.33% and 0.65.

A B

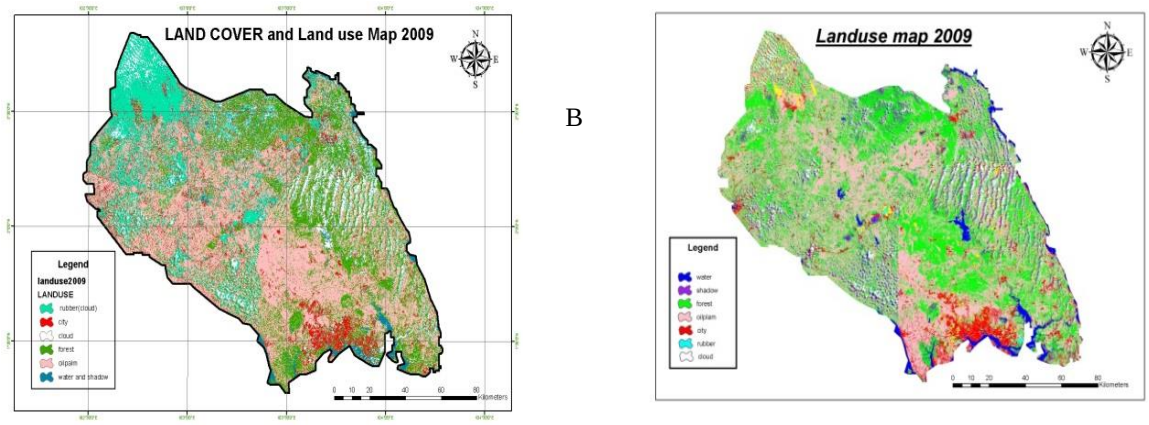

Figure 2. Land use/land cover maps produced using (A) maximum likelihood and (B) support vector machine for year 2009

Table 1. Accuracy assessment of maximum likelihood and support vector machine classifier

	Maximum likelihood		Support vector machine	
Land use - land cover	Producer's accuracy (%)	User's accuracy (%)	Producer's accuracy (%)	User's accuracy (%)
forest	65.5	71.43	87.5	87.5
oil palm	86.21	78.13	93.65	90.77
city	83.33	83.33	83.33	83.33
water	71.43	78.95	91.89	97.14
rubber	66.67	80.00	83.33	83.33
Overall accuracy	78.33%		91,67 (%)	
Kappa	0.65		0.86	

4. Conclusion

In this study, SVM and MLC classification techniques have been applied to identify the improved and accurate method for land use classification. Results of classifications from both the methods have been verified by land cover map for year 2006.Level of accuracy has also been measured through different proven parameters. It has been concluded that the accuracy level is comparatively better in SVM method than MLC method. The statistical Maximum Likelihood classifier cannot handle complex images so that many pixels cannot be classified correctly, but SVM also could overcome the mixed pixel problem in the image. According to SVM accuracy, it is better we use SVM for hyperspectral images.

References

[1] Al-Ahmadi, F. and Hames, A. (2009). Comparison of four classification methods to extract land use and land cover from raw satellite images for some remote arid areas, kingdom of Saudi Arabia. *Earth,* 20**,** 167-191.

[2] Foody, G. M. and Mathur, A. (2004). A relative evaluation of multiclass image classification by support vector machines. *Geoscience and Remote Sensing, IEEE Transactions on,* 42**,** 1335-1343.

[3] Gualtieri, J. A. and Cromp, R. F. (Year). Support vector machines for hyperspectral remote sensing classification. *The 27th AIPR Workshop: Advances in Computer-Assisted Recognition,* 1999. International Society for Optics and Photonics, 221-232.

[4] Hagner, O. and Reese, H. (2007). A method for calibrated maximum likelihood classification of forest types. *Remote Sensing of Environment,* 110**,** 438-444.

[5] Huang, C., Davis, L. and Townshend, J. (2002). An assessment of support vector machines for land cover classification. *International journal of remote sensing,* 23**,** 725-749.

[6] Kanellopoulos, I., Varfis, A., Wilkinson, G. and Megier, J. (1992). Land-cover discrimination in SPOT HRV imagery using an artificial neural network—a 20-class experiment. *International journal of remote sensing,* 13**,** 917-924.

[7] Kavzoglu, T. and Colkesen, I. (2009). A kernel functions analysis for support vector machines for land cover classification. *International Journal of Applied Earth Observation and Geoinformation,* 11**,** 352-359.

[8] Lewis, D., Corr, D., Gent, C. and Sheppard, C. (1992). Semi-supervised artificial neural networks for classification of remotely sensed images. *Remote Sensing from Research to Operation. Remote Sensing Society, Nottingham***,** 489-497.

[9] Lillesand, T. M., Kiefer, R. W. and Chipman, J. W. (2008). *Remote sensing and image interpretation 6th edition (New York:John Wiley & Sons Ltd).* John Wiley & Sons Ltd.

[10] Melgani, F. and Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *Geoscience and Remote Sensing, IEEE Transactions on,* 42**,** 1778-1790.

[11] Mondal, A., Kundu, S., Chandniha, S. K., Shukla, R. and Mishra, P. (2012). Comparison of Support Vector Machine and Maximum Likelihood Classification Technique using Satellite Imagery.

[12] Naguib, A., Farag, M., Yahia, M., Ramadan, H. and Abd Elwahab, M. (2009). Comparative study between support vector machines and neural networks for lithological discrimination using hyperspectral data. *Egypt Journal of Remote Sensing and Space Science,* 12**,** 27-42.

[13] Nasrabadi, N. M. and Choo, C. Y. (1992). Hopfield network for stereo vision correspondence. *Neural Networks, IEEE Transactions on,* 3**,** 5-13.

[14] Oommen, T., Misra, D., Twarakavi, N. K., Prakash, A., Sahoo, B. and Bandopadhyay, S. (2008). An objective analysis of support vector machine based classification for remote sensing. *Mathematical Geosciences,* 40**,** 409-424.

[15] Pal, M. and Mather, P. (2005). Support vector machines for classification in remote sensing. *International journal of remote sensing,* 26**,** 1007-1011.

[16] Roy, P. and Giriraj, A. (2008). Land use and land cover analysis in Indian context. *Journal of Applied sciences,* 8**,** 1346-1353.

[17] Shalaby, A. and Tateishi, R. (2007). Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt. *Applied Geography*, 27**,** 28-41.
[18] Szuster, B. W., Chen, Q. and Borger, M. (2011). A comparison of classification techniques to support land cover and land use analysis in tropical coastal zones. *Applied Geography*, 31**,** 525-532.
[19] Vapnik, V. (2000). *The nature of statistical learning theory*. Springer.
[20] Vapnik, V. N. (1998). Statistical learning theory.
[21] Yu, L., Porwal, A., Holden, E.-J. and Dentith, M. C. (2012). Towards automatic lithological classification from remote sensing data using support vector machines. *Computers & Geosciences*, 45**,** 229-239.

[22] http://en.wikipedia.org/wiki/Johor_Bahru
[23] https://maps.google.com/
[24] http://earthexplorer.usgs.gov/