Land cover classification of Landsat 8 satellite data based on Fuzzy Logic approach

Afirah Taufik¹ and Sharifah Sakinah Syed Ahmad²

Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia¹. Department of Industrial Computing, Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia².

Email: afirahtaufik@gmail.com

Abstract. The aim of this paper is to propose a method to classify the land covers of a satellite image based on fuzzy rule-based system approach. The study uses bands in Landsat 8 and other indices, such as Normalized Difference Water Index (NDWI), Normalized difference built-up index (NDBI) and Normalized Difference Vegetation Index (NDVI) as input for the fuzzy inference system. The selected three indices represent our main three classes called water, built-up land, and vegetation. The combination of the original multispectral bands and selected indices provide more information about the image. The parameter selection of fuzzy membership is performed by using a supervised method known as ANFIS (Adaptive neuro fuzzy inference system) training. The fuzzy system is tested for the classification on the land cover image that covers Klang Valley area. The results showed that the fuzzy system approach is effective and can be explored and implemented for other areas of Landsat data.

1. Introduction

The classification of Landsat 8 satellite data using fuzzy logic is a method to extract features of a land cover into certain classes. Landsat 8 contains multispectral bands which are invisible and infrared bands. The combination of certain bands is able to classify the features of the classes such as water, vegetation and built-up land. The indices that are used are input for fuzzy inference system are Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI) and Normalized Difference Built-up Index (NDBI).

The aim of this paper is to propose a method to classify the land covers of a satellite image based on fuzzy rule-based system approach. The purpose of the image classification system is to assign each input of the indices to certain classes according to fuzzy logic approach. Fuzzy logic has been used widely to make a decision with incomplete or uncertain data. The combination of the neural networks and fuzzy theoretic approaches, are formed into Adaptive neuro fuzzy inference system (ANFIS) by attaching the fuzzy inference system to the framework of adaptive networks. The parameter selection of fuzzy membership function using supervised method ANFIS training is important to achieve the best performance for the fuzzy inference system. The fuzzy system is tested for the classification on the land cover image that covers the Klang Valley area. The results are based on the fuzzy rules-based used as detection for a land cover classes and based on the accuracy assessment for the classified pixels.
2. Related works

Vegetation index can be used as an indicator to quantify the greenness of plants within the satellite data. There are several vegetation indices, but the most frequently used index is the Normalized Difference Vegetation Index (NDVI) [1]. NDWI is sensitive to changes in liquid water content of vegetation canopies. NDWI is less sensitive to atmospheric effects than NDVI, NDWI does not remove the background soil reflectance effects completely as NDVI [2]. Urban spatial areas have expanded in an accelerated speed during the last five decades, and rates of urban population growth are higher than the overall growth in most countries because urban areas are the locus of economic activity and transportation nodes [3]. The paper proposes a technique to extract urban built-up land features from Landsat Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM) imagery taking two cities in the South Eastern China as examples. The study selected three indices, Normalized Difference Built-up Index (NDBI), Modified Normalized Difference Water Index (MNDWI), and Soil Adjusted Vegetation Index (SAVI) to represent three major urban land-use classes, built-up land, open water body, and vegetation, respectively [4]. Fuzzy set theory [5] provides useful concepts and tools to deal with imprecise information. Partial membership allows the information about more complex situations, such as cover mixture or intermediate conditions, can be better represented and utilized. However, works in remote sensing image analysis using fuzzy sets are rather scarce. The proposed ANFIS can construct an input-output mapping based on both human knowledge (in the form of fuzzy if-then rules) and stipulated input-output data pairs. In the simulation, the ANFIS architecture is employed to model nonlinear functions, identify nonlinear components on line in a control system, and predict a chaotic time series, all yielding remarkable results [6]. In this study, object-based image analysis is used for image classification. The main problems in high resolution image classification are the uncertainties in the position of object borders in satellite images and also multiplex resemblance of the segments to different classes. In order to solve these problems, fuzzy logic is used for image classification, since it provides the possibility of image analysis using multiple parameters without requiring inclusion of certain thresholds in the class of assignment process [7].

3. Materials and methodology

Figure 1 shows a framework of the land cover classification study. The details of materials and methodology are explained in the next section.

3.1. Data

As the initial stage, the study was conducted with information collection of satellite image at Klang, Selangor. This study area is situated between longitudes 101° 16' to 102° 00' E and scope 3° 05' to 3° 22' N. The satellite information was procured from USGS Landsat 8 information that was utilized as a part of this study. The information was chosen by considering the high calibre of information gained by the Landsat satellite. The multispectral bands Landsat 8 data contains visible and infrared wavelength. The data were acquired on 24th March 2014. By utilizing an unearthly subset as a part of ENVI programming, the chosen region of interest (ROI) was picked as training pixels. The DAT record contains 7 groups with 349 pixels × 329 pixels.

3.2. Pre-processing

The Landsat 8 digital data of study area was read into MATLAB. Image processing was carried out to produce better CIR (color-infrared) composite using a decorrelation stretch. Band combinations were performed many times/several times in order to obtain visual of interpretation of classes and other information. The image processing system can display three band combinations by assigning three different colors to each band. The images before and after processed are shown in Figure 2. The shown images adopt band 5, 4 and 3.
Figure 1. The framework of classification experiment.

Figure 2. The process before (a) and after (b) pre-processing.
3.3. Classification

In this study, the layer of the multispectral bands consist of NIR (near-infrared), SWIR (short-infrared), Red and Green are used. The layer of the bands 5-4-3 (NIR, Red and Green) are extracted and mapped into red, green and blue (RGB color). The RGB color is a standard color of infrared (CIR) image. The NDVI, NDWI and NDBI were determined by utilizing the formula (1), (2) and (3):

\[
NDVI = \frac{NIR - Red}{NIR + Red}
\]

(1)

\[
NDWI = \frac{Green - NIR}{Green + NIR}
\]

(2)

\[
NDBI = \frac{NIR - SWIR}{NIR + SWIR}
\]

(3)

3.3.1. ANFIS modelling

Fuzzy inference systems are the fuzzy rule based systems which consists of a fuzzification interface, rule base, database, decision making unit, and defuzzification interface [8]. ANFIS is the supervised training method to determine the selection of membership function for the three inputs based on the output (ground truth information). Figure 3 shows the fuzzy modelling by using ANFIS training to determine the selection of parameter of membership function of the inputs for the fuzzy inference system. The data are divided into training and testing data. The three inputs are NDVI, NDWI and NDBI. The output data is also a vital part where the ground truth information is needed to classify the satellite image data. Output data consists of three classes based on the ground truth information, water, built-up land and vegetation. The left-nodes are input and the right-nodes are output. Each node performs a particular function from input node as well as asset of parameters to the node. There are 27 rules are used for the land cover classification by using fuzzy inference system. The fuzzy if-then rules for three classes are discussed in the next section.

---

**Figure 3.** ANFIS model structure.
3.4 Accuracy assessment

The classified image was compared with information of ground truth data in order to check the accuracy. The user, producer and overall accuracy were carried out to measure the classification accuracy. Producer’s accuracy is where the individual class accuracy can be acquired by dividing the sum of correctly classified pixels. Meanwhile, the user’s accuracy is a measurement where the individual class acquired from the classified pixels in same group [8]. The number of misclassified pixels into another class were recorded. The overall accuracy (4) was calculated according to the confusion matrix that obtained for user’s accuracy and producer’s accuracy. The description of overall accuracy is shown below:

\[
\text{Overall accuracy} = \frac{\text{Total number of correct classified}}{\text{Total number of pixels}} \times 100
\]  

(4)

Kappa coefficient (5) is another measurement to measure the training pixels with the ground truth data information. The Kappa values are +1.0 to -1.0. The positive values show high accuracy and the value of zero shows no correlation in the classification.

\[
K = \frac{n \sum_{i=1}^{p} x_{ii} - \sum_{i=1}^{p} (x_{i+} \times x_{+i})}{n^2 - \sum_{i=1}^{p} (x_{i+} \times x_{+i})}
\]

(5)

Where,
- \(n\) = total number of training pixels
- \(p\) = number of classes
- \(\sum x_{ii}\) = total number elements of confusion matrix
- \(\sum x_{i+}\) = sum of row \(i\)
- \(\sum x_{+i}\) = sum of column \(i\)

The results of accuracy assessment for a classification based on the confusion matrices were obtained and recorded in the next section.

4. Results and discussion

4.1 Accuracy assessment results.

The results of this study are acquired based on the accuracy assessment and the fuzzy rule based. The classification process to satellite data must be quantitatively determined their accuracy of classification. By using the ANFIS training to determine the selection of parameter membership function in fuzzy inference system, the three classes output were classified. The pixels that have been classified using fuzzy classification were compared with the ground truth information. In order to evaluate the fuzzy inference system for land cover classification, an area (349 pixels × 329 pixels) of image is classified based on three input indices: NDVI, NDWI and NDBI. The results of classification showed the color mapping and labelling of 3 classes. As we can observe in Figure 3, the images are grouped into 3 categories, water, built-up land and vegetation. Table 1 shows the related confusion matrices assessment and classification results (overall accuracy and kappa coefficient). Based on the results, the total pixels are correctly classified is very high rather than misclassified pixels for every category class. The misclassified pixels for the vegetation class are 228 and 9 pixels, built-up land and water, respectively.
The misclassified pixels for the built-up land class are in the other two classes; vegetation and water, 387 and 168, respectively. The misclassified pixels for water class into the other class; built-up land, is 60 pixels. The user’s accuracy and producer’s accuracy were obtained from the submission of every row and column from the confusion matrix table. The correctly classified pixels of every row and column are divided with the sum of total pixels for every row and column. The percentage for overall accuracy that obtained from formula (4) is 97.6%, where the result is highly accurate. Kappa coefficient is 0.956 and it is noteworthy that the value is the positive range. This result is classified as high accuracy for kappa coefficient. The final accuracy is very high and exceeds 90 % for every classes.

![Figure 4. Color mapping and labelling.](image)

| Class         | Vegetation | Built-up land | Water | Total  | User’s Accuracy (%) |
|---------------|------------|---------------|-------|--------|---------------------|
| Vegetation    | 22,559     | 228           | 9     | 22,796 | 98.9%               |
| Built-up land | 387        | 5,875         | 168   | 6,430  | 91.3%               |
| Water         | 0          | 60            | 8,928 | 8,988  | 99.3%               |
| Total         | 22,946     | 6,163         | 9,105 | 38,274 |                     |
| Producer’s Accuracy (%) | 98.3% | 95.3%         | 98%   |         |                     |

Overall accuracy = 97.6%
Kappa coefficient = 0.956

Defuzzification of the fuzzy output function into the crisp output is one of the processes in fuzzy logic. Based on the fuzzy if-then rules, there are three rules are used for a class detection from Takagi-Sugeno. The three classes are vegetation, water and built-up land. The rules for vegetation shows that the membership function for NDVI is high, and another two membership functions for two inputs are low and medium; NDWI and NDBI, respectively.
The rules for water shows that the membership function for NDWI is high, and another two membership functions for two inputs are medium and low; NDVI and NDBI, respectively. The rules for built-up land shows that the membership function for NDBI is high, and another two membership functions for two inputs are low for NDVI and NDWI.

If NDVI is high and NDWI is low and NDBI is medium then class is Vegetation.
If NDVI is medium and NDWI is high and NDBI is low then class is Water.
If NDVI is low and NDWI is low and NDBI is high then class is Urban.

5. Conclusion

The results that presented in this study demonstrate the effectiveness of fuzzy approach for a land cover classification by using ANFIS modelling. The parameter selection of fuzzy membership function using supervised method ANFIS training is important to achieve the best performance for the fuzzy inference system. The fuzzy rule-based are used for class detection. The supervised Adaptive neuro that is formed with fuzzy inference system is effective and can be explored and implemented for other areas of the Landsat data.

6. References

[1] J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering, “Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation,” 1973.
[2] B. C. Gao, “NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space,” Remote Sens. Environ., vol. 58, no. 3, pp. 257–266, 1996.
[3] “The Dynamics of Global Urban Expansion 31,” pp. 31–48, 2000.
[4] H. Xu, “Extraction of urban built-up land features from Landsat imagery using a thematic-oriented index combination technique,” Photogramm. Eng. Remote Sensing, vol. 73, no. 12, pp. 1381–1391, 2007.
[5] L. A. Zadeh, “Fuzzy Sets-Information and Control.”, vol. 8, pp. 338-353, 1965.
[6] J. S. R. Jang, “ANFIS: Adaptive-Network-Based Fuzzy Inference System,” IEEE Trans. Syst. Man Cybern., vol. 23, no. 3, pp. 665–685, 1993.
[7] S. Jabari and Y. Zhang, “Very High Resolution Satellite Image Classification Using Fuzzy Rule-Based Systems,” Algorithms, vol. 6, no. 4, pp. 762–781, Nov. 2013.
[8] S. N. Sivanandam, S. Sumathi, and S. N. Deepa, Introduction to Fuzzy Logic using MATLAB. 2007.
[9] R. G. Congalton, “A review of assessing the accuracy of classifications of remotely sensed data,” Remote Sens. Environ., vol. 37, no. 1, pp. 35–46, 1991.

Acknowledgements

This research is supported by Universiti Teknikal Malaysia Melaka and the Ministry of Education Malaysia under grant no PJP/2015/FTMK (4B)/S01433.