IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

Segmentation and classification models validation area mapping of peat lands as initial value of Fuzzy Kohonen Clustering Network

To cite this article: Erwin *et al* 2017 *IOP Conf. Ser.: Mater. Sci. Eng.* **190** 012037

IOP Conf. Series: Materials Science and Engineering **190** (2017) 012037 doi:10.1088/1757-899X/190/1/012037

Segmentation and classification models validation area mapping of peat lands as initial value of Fuzzy Kohonen Clustering Network

Erwin , Saparudin , Muhammad Fachrurrozi

Informatics Engineering, Computer Science Faculty, Universitas Sriwijaya, Indonesia

erwin@unsri.ac.id

Abstract. Ogan Komering Ilir (OKI) is located at the eastern of South Sumatra Province, $2^0$30'-$4^0$15' latitude and $104^0$20'-$106^0$00' longitude. Digital image of land was captured from Landsat 8 satellite path 124/row 062. Landsat 8 is new generation satellite which has two sensors, Operation Land Manager (OLI) and Thermal Infra-Red Sensor (TIRS). In pre-processing step, there are a geometric correction, radiometric correction, and cropping of the digital images which resulting coordinated geography. Classification uses maximum likelihood estimator algorithm. In segmentation process and classification, grey value spread is into evenly after applying histogram technique. The results of entropy value are7.42 which is the highest of result image classification, then the smallest entropy value in the result of correction mapping are 6.39. The three of them prove that they have enough high entropy value. Then the result of peatlands classification is given overall accuracy value = = 94.0012% and *overall kappa value* = 0.9230 so the result of classification can be considered to be right.

1. Introduction

Image segmentation is one of the complicated problems in image processing. The target which has to be achieved from image processing is partitioning the image into multiple areas which mean according to some features, so there are some features in one area and others area are not same. Image segmentation is a feature based on pixel classification procedure, then the clustering analysis is a viable way to apply the image segmentation [1]. Image segmentation techniques automatically classify adjacent pixels to contiguous region according to similarity criteria of property pixels. The object can be better than a pixel, in terms of knowing their neighbors along with spatial connection and spectral between pixels.

Some of the features affect the segmentation result depending on several things: parameters scale, shape, Smoothness, and harmony. Parameters scale is a measure that specifies the maximum value of allowed heterogeneity in generating image objects where for heterogeneous data, the objects that are generated will be smaller than more homogeneous data and by modifying parameters scale value can be made the diverse size of image objects. An indirect form can specify color criteria, stating what percentage of spectral values in the layer image that will contribute to the overall criteria of their homogeneity.

Image segmentation includes part of image processing [2]. Image segmentation aims to separate the region object with the background region so that the object in the image is easy to be analyzed in order to identify objects [3].

The utilization of neuro fuzzy systems is for automatic image segmentation and detect the edge is one of new breakthrough that will certainly add to the advanced digital world particularly image [4].

The whole procedure of image classification of remote sensing aims to group all the pixels in the image into thematic cover class and land utilization. Conventional classification technique uses the unsupervised technique as well as supervised technique while taking decision method can be used minimum distance method, parallelepiped and maximum likelihood. However, because of the clustering process, the partition is very sensitive in the steps of clustering center initialization so that causes tend to stick on local optimization, therefore it needs validation on the model.

Several methods can be used to find a model segmentation optimization and classification such as Fuzzy Kohonen Clustering Network for Hazard zonation with Algorithm Kringing for Interpolation Data, identifying flood area using Classifier which is known as Fuzzy Kohonen Local Information C-Means (FKLICM) with the assistance or the comparison using HKFCM-r Classifier [5], comparing learning methods in clustering on pattern recognition , extraction, VQ, Image Segmentation, and data mining [6], Analysis Entropy using spatial information by adopting general characteristics Hopfield Neural Network (HNN) and multi-synapse neural network (MSNN) [7], Comparing a Segmentation of different color images [8]. Test the effectiveness of the maximum iteration t and initialization maximum value 0 m on the behavior of FKCN[9]. Obtaining and Analyzing information about object or earth, region or symptom[8]. The concept of fuzzy gives the set to handling a vague and inaccurate data [10]. It can be divided into two main categories : Supervised and unsupervised [11]. The algorithm used is a modified version of the Kohonen Algorithm which found in network nerve. So, this algorithm can also refer as Fuzzy Clustering Kohonen Network (FKCN) [12].

2. Statistic Characteristic

One of the Characteristic of digital image textures which can be calculated from statistics is the intensity value of the gray scale in the image that is mean, standard deviation, histograms, and entropy value. In this research, all those three parameters are calculated and used to validate the model. Mean indicates the average value of the matrix which formed from the value of the pixels of an image while the deviation standard indicates the spread of data. Mathematically the calculation of mean and standard deviation using equations (1)

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$std = \sqrt{\frac{1}{n}\left(\sum_{i=1}^{n} x_i - \bar{x}\right)^2} \quad (1)$$

2.1. Histogram

The histogram is a graph that shows the spread of the values of the pixels of the image, then it can be known the value of relative frequency. In addition, the histogram can indicate the level of brightness and contrast of an image. Histogram calculation mathematically can be calculated using equation (2)

$$I_t = \frac{f_t}{f}, \quad t = 0,1,\dots,K-1 \quad (2)$$

where I_t is histogram value, f is sum of all pixel

The distribution of I_t and f_t provide information about the appearance of the image. The image histogram is distributed evenly on the whole level of grayscale which means having a good level of contrast when histogram accumulates in the dark area means having Dim image and histograms which accumulatein the light area (high intensity) means showing a bright image.

2.2. Histogram
Image Entropy is statistic to measure the randomness of distribution value pixels. The image is called as perfect if have zero entropy. Entropy can be used as a measure of statistics on the characterization of the texture of the greyscale image input. Entropy formulated by equation (3).

$$E = -\sum_t I_t log_2 I_t \quad (3)$$

3. Materials and Method
The subject of the research is an area of peatlands distribution in the province of Ogan Komering IIir (OKI) South Sumatera which is geographically situated between 104°20 ' and 106°00 ' East longitude and 2°30 ' to 4°15 ' South latitude. The data used in this study i.e. Landsat 8 image downloaded from the USGS website with path/row 124/062 with photo shoot on 17-07-2015 Landsat7 Image and ETM orthorectification in 2010 which will be used for geometric correction of reference. Steps Image data processing on Landsat 8 which done in this research are described in section 3.1 until 3.3.

3.1. Radiometric and Geometric Correction
Image data of the Landsat 8 is done using the image of Landsat 7 ETM orthorectification (the image has been processed and corrected geometric). While radiometric corrections were done to fix the value of the pixels to match with the histogram adjustment technique. While the geometric correction is done for corrected between the coordinates of the pixel digital image field controls and topographic map coordinates.

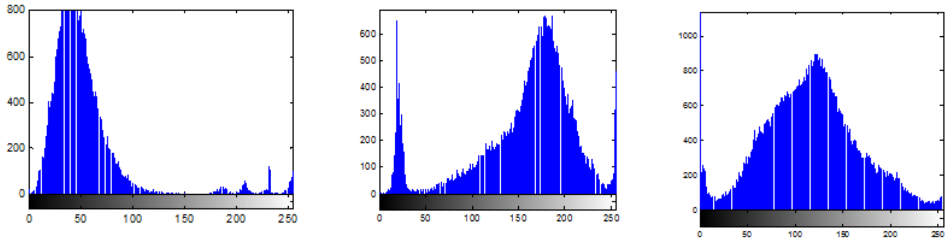
3.2. Cropping
The image is done to get the research areas with a view to being able to do image data processing that is more focused, detailed and optimal. With expectation generates the image of a representative and continuous. The cutting image has value to other utilities, namely reducing the areas that will be examined in accordance with the area of interest. The image cuts can be carried out in accordance with the desired polygon shape such as restrictions on the territory of the County, district or village. So, image cropping (cutting images) can be useful to facilitate the performance of a person while being observing the image, especially in restricting certain regions.

3.3. Classification of land cover
Land cover classification is used the guidance classification by using the maximum likelihood algorithm and based on land cover types on the map. In this algorithm the pixels classified as a particular object according to the shape, size and orientation of the samples in the feature space. While the accuracy of the algorithm can be computed using the confusion matrix with the given tolerance limit i.e. ≥ 80%.

Maximum Likelihood Algorithm
1. Supposed random variable $x_1, x_2, \dots, x_n$ has joint probability distribution $f(x_1, x_2, \dots, x_n; \theta)$
2. Form of Likelihood function as $L(\theta) = f(x_1, x_2, \dots, x_n; \theta)$
3. Maximum Likelihood function (Max $L(\theta)$) with method:
   $\frac{dL(\theta)}{d\theta} = 0$ or $\frac{dLog\ L(\theta)}{d\theta} = 0$
4. If more than one parameter:
   $$L(\theta_1, \theta_2, \dots, \theta_n) = \prod_i^n f(x_i; \theta_1, \theta_2, \dots, \theta_n)$$
   $\frac{dL(\theta_1, \theta_2, \dots, \theta_n \theta)}{d\theta_1} = 0$, ---, $\frac{dL(\theta_1, \theta_2, \dots, \theta_n \theta)}{d\theta_n} = 0.$

4. Result and Discussions

Histogram for 3 (three) mapping image is the map of correction result (radiometric and geometric) and map of classification result are counted using Matlab 2009a. Those are presented in figure 1 as :



a b c

Figure 1. Histogram a. Map of correction result, b. Map of segmentation result and c. Map of classification result

From histogram result which can be seen in figure 1. Map image of correction result originally has heap histogram in a low area (0-100). The heap histogram in this area is caused the map is too dark. After the segmentation process has done, the histogram is seen to start spreading and in the classification process, the histogram is seen spreading in all grey area (0-255). Spread histogram evenly means the image result into better quality and better level brightness. Calculation value of statistic value RGB (red, green and blue) is in the form of entropy, mean, and deviation standard for three type of image is seen in table 1 below:

Table 1. The Statistic of Correction Geometry, Segmentation and Classification Results

Statistics	Corrected Map	Segmentation	Classification
Matrix Dimension	193x223x3	227x226x3	309x304x3
Entropy of Image	6.39	7.69	7.42
Red			
Entropy	6.50	7.36	7.60
Mean	53.32	157.71	115.51
Standard Deviation	39.04	54.34	51.07
Green			
Entropy	6.30	6.80	7.34
Mean	43.69	134.50	108.89
Standard Deviation	19.98	47.42	47.10
Blue			
Entropy	6.28	6.87	5.48
Mean	43.80	76.49	18.95
Standard Deviation	19.99	32.28	17.55

Based on table 1, it is shown that the highest entropy value in classification image result is 7.42, while the smallest entropy value in the map of correction result is 6.39.Statistic of blue color shows the smallest entropy value for corrected map and map of classification result are 6.28 and 5.48, while in the map of segmentation result shows green color has the smallest entropy value is 6.89. The three of image prove that those have enough high entropy value. It means that the image is still far from excellence. The image can be said excellence if it has zero in entropy value. Zero entropy value can be earned if image histogram evenly in all of the parts.

In this research, the image of red color shows peatlands classification result and generate overall accuracy value = 94.0012% and overall kappa value = 0.9230. With that result, classification can be considered to be right. Contour plot for the three map show in figure 2, as:

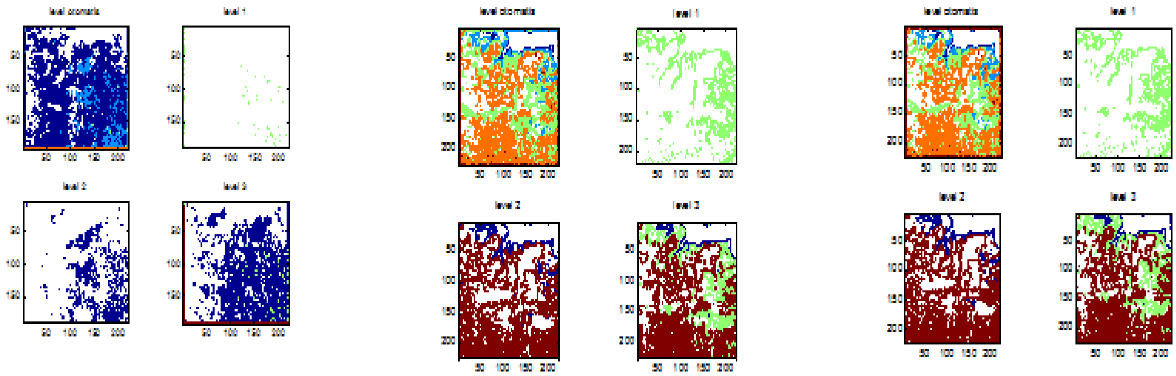


Figure 2. Contour Plot with level automatically, 1, 2 and 3, a. corrected map b. map of segmentation result and c. map of the classification result

5. Conclusion

Utilization histogram technique can be applied in segmentation process and classification in the image of peatlands mapping which causes the spread of grey value evenly. The highest entropy value in classification image result is 7.42, while the smallest entropy value in correction mapping result is 6.39. The three of image prove that those have enough high entropy value. The result of peatlands classification is given overall accuracy value = 94.0012% and overall kappa value = 0.9230 so the classification result is considered correct.

Acknowledgements

Thank you so much to reviewers for all review result which has completed this journal. This journal is powered by grants research Unggulan Kompetitif Universitas Sriwijaya scheme.

References

[1]. L. Bosheng, W. Yuke, and L. Jiangping, "A noise-resistant fuzzy kohonen clustering network algorithm for color image segmentation," *Proc. 2009 4th Int. Conf. Comput. Sci. Educ. ICCSE 2009*, pp. 44–48, 2009.
[2]. N. I. Jabbar and M. Mehrotra, "Application of Fuzzy Neural Network for Image Tumor Description," *Proc. World Acad. Sci.*, vol. 1, no. 8, pp. 575–577, 2008.
[3]. A. Z. Arifin and A. Asano, "Review Paper Image segmentation by histogram thresholding using hierarchical cluster analysis," vol. 1, pp. 1–4, 2006.
[4]. Shweta Lawanya Rao and C.L.Chandrakar, "An Adaptive Neuro-Fuzzy System for Automatic Image Segmentation and Edge Detection," *Int. J. Data Warehous.*, vol. 4, no. 1, pp. 33–40, 2012.
[5]. K. K. Singh and A. Singh, "Identification of flooded area from satellite images using Hybrid Kohonen Fuzzy C-Means sigma classifier," *Egypt. J. Remote Sens. Sp. Sci.*, 2016.
[6]. K. L. Du, "Clustering: A neural network approach," *Neural Networks*, vol. 23, no. 1, pp. 89–107, 2010.
[7]. B. Huo and F. Yin, "Image Segmentation using Neural Network and Modified Entropy," vol. 8, no. 3, pp. 251–260, 2015.
[8]. N. I. Jabbar and S. I. Ahson, "Modified fuzzy Kohonen clustering network for image segmentation," *2010 Int. Conf. Financ. Theory Eng. ICFTE 2010*, vol. 3, pp. 176–179, 2010.
[9]. N. I. Jabbar, "Estimation Clusters Based Preprocessing in Fuzzy Kohonen Clustering Network for Image Segmentation," *Int. Conf. Artif. Intell. Image Process.*, pp. 207–210, 2012.
[10]. A. Hamdy, "Fuzzy Logic for Enhancing the Sensitivity of COCOMO Cost Model," vol. 3, no. 9,

pp. 1292–1297, 2012.
[11]. V. Jecheva and E. Nikolova, “Some clustering-based methodology applications to anomaly intrusion detection systems,” *Int. J. Secur. its Appl.*, vol. 10, no. 1, pp. 215–228, 2016.
[12]. D. V. Fernandes, T. Thomas, N. Joseph, and J. Joseph, “Image Segmentation using Fuzzy - Kohonen Algorithm,” no. Icmlc, pp. 176–179, 2011.