

The Application of Fuzzy Clustering to Satellite Images Data

HIZIR SOFYAN¹, MD. AZLIN MD. SAID², MUZAILIN AFFAN¹, KHALED BAWAHIDI²

¹ Mathematics Department
Syiah Kuala University
Darussalam, Banda Aceh 23111
INDONESIA
<http://www.hizir.net>

² School of Aerospace Engineering,
Universiti Sains Malaysia, Transkrian Campus
14300 Nibong Tebal, Penang
MALAYSIA

Abstract: - In this study, a fuzzy clustering method that we implemented in statistical software package (XploRe), GIS (GRASS) and remote sensing (ERDAS) were used to classify and preset multi-spectral Landsat TM images. Comparatively, the result of this investigation shows that the underlying structures and patterns from satellite images data can be classified more precisely than the conventional ones.

Key-Words: - *Fuzzy Clustering, Satellite Images Data, Remote Sensing, XploRe*

1 Introduction

In the scope of the development of aerospace technology, many satellite images data are available for information processing. Using satellite images, a large number of environmental questions can be answered. Land use is one of the factors that influences the environmental changing. Most of the land use changes are caused by human activities, such as deforestation for expanding agricultural land or for urban use. This will be major aspects of attention we relation to the city such as Ipoh in Malaysia with the reduction in the rain forest stand, air pollution, extremely rapid extension of its urban area and many other environmental changes [4].

Clustering can be used to obtain some initial information from the satellite images data. This method constructs groups in such a way that the profiles of objects in the same groups are relatively homogenous whereas the profiles of objects in different groups are relatively heterogeneous. By conventional clustering methods, class is either assigned to or not assigned to a defined group. Fuzzy clustering which apply the concept of fuzzy sets [7] to cluster analysis may pertaining to group at each pixel of the images data by a membership function which associated to each cluster ranging between 0 and 1.

The purpose of this study is to classify the satellite image of Ipoh City, Malaysia into definite themes. Several bands are segmented and analyzed using

fuzzy clustering which will result a new layer containing a certain theme or a classes (clusters). These resulted classes can be then manually compared with their ground truth.

The rest of this paper is organised as follows. In section 2 we present review of remote sensing. Section 3 performs fuzzy clustering. The data will be described in Section 4. Section 5 presents results and discussions. The conclusions of this paper is presented in the last section.

2 Remote Sensing

Remote sensing is a form of measuring physical characteristics of remote object without being in contact with them [4]. In this case, the earth surface is using as the target of observation. The remote sensing system is a repeatedly consistent activity toward the earth surface which is supposed to monitor the earth system and its effect on the human activities. Here, the measurement is focused on images, two dimensional spatial grid of spectral and spatial information.

Using satellite images, a large number of environmental questions can be answered such as land-use and land-change. These images consist of pixels which are called raster cell. A pixel is the smallest unit of an image. After a basic image was rectified into a coordinate system, bands of a satellite image can be overlaid.

Satellite image consists of several bands. These bands give information about the spectral reflection from the earth surface. Each band will have different pixel value which depends on the measuring frequency and has a value between 0-255 representing the spectral distribution of these bands. In this paper we just concentrate on images as data without involving the process of how the data have been physically absorbed.

3 Fuzzy Clustering

One approach to fuzzy clustering, probably the best and most commonly used, is the Fuzzy C-Means (FCM) by [3]. He generalized the fuzzy objective function by introducing the weighting exponent m , $1 \leq m < \infty$;

$$J_m(U, P; X) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m d^2(x_k, v_i),$$

where U is a partition of X in c part, $P = v = (v_1, v_2, \dots, v_c)$ are the cluster centers in R^p . A is any $(p \times p)$ symmetric positive definite matrix defined as the following:

$d^2(x_k, v_i) = \|x_k - v_i\|_A = \sqrt{(x_k - v_i)^T A (x_k - v_i)}$, where $d^2(x_k, v_i)$ is an inner product induced norm on R^p , u_{ik} is referred to as the grade of membership of x_k to the cluster i . This grade of membership satisfies $0 \leq u_{ik} \leq 1$ with some constraints.

Steps of the FCM algorithm, according to [5] are the following:

1. Given a data set $X = \{x_1, x_2, \dots, x_n\}$, select the number of clusters $2 \leq c < N$, the maximum number of iterations T , the distance norm $\|\bullet\|_A$, the fuzziness parameter m , and the termination condition $\varepsilon > 0$.

2. Give an initial value $U_0 \in M_{fcn}$.

3. For $t = 1, 2, \dots, T$

(a) Calculate the c cluster centers $\{v_{i,t}\}$, $i = 1, \dots, c$

$$v_{i,t} = \frac{\sum_{k=1}^n u_{ik,t-1}^m x_k}{\sum_{k=1}^n u_{ik,t-1}^m}$$

(b) Update the membership matrix. Check the occurrence of singularities

$$(d_{ik,t} = \|x_k - v_{i,t}\|_A = 0).$$

Let $Y = \{1, \dots, c\}$, $Y_{k,t} = \{i \in Y \mid d_{ik,t} = 0\}$, and $Y_{k,t} = Y \setminus Y_{k,t}$. Then calculate

$$u_{ik,t} = \sum_{j=1}^c \left(\frac{d_{ik,t}}{d_{jk,t}} \right)^{\frac{2}{m-1}}, \quad \text{if } Y_{k,t} = \emptyset$$

Choose $a_{ik,t} = 1/\#(Y_{k,t})$, $\forall i \in Y$; $\#(\cdot)$ denotes the ordinal number.

4. If $E_t = \|U_{t-1} - U_t\| \leq \varepsilon$ then stop otherwise return to step 3.

This procedure converges to a local minimum or a saddle point of J_m . The FCM algorithm computes the partition matrix U and the clusters' prototypes in order to derive the fuzzy models from these matrices.

4 Data and Methods

The data set consists of 7 channels from a Landsat TM scene (256x256 pixels) around Ipoh city, Malaysia. The data were obtained from the School of Aerospace Engineering, University of Sains Malaysia. The image was rectified to Malaysian RSO projection, with the RMS error less than 0.5 pixel.

The chosen area study is located in the Ipoh city. Some reasons of this choice are:

- This location lies in the Kinta Valley, a mining area, which is currently undergoing changes to its land use..
- There is an availability of digital data since 1991.

The site and its surroundings include the urban and suburban area where the major land cover comprises the mixture of building, vegetation, bare soil, and water. The land classes appeared in Figure 1 show a fairly complex mixture that obviously will create a big challenge to determine the right class.

The methodology that we used are unsupervised classification (fuzzy clustering) and the ISODATA unsupervised classification method as comparison. More information of ISODATA unsupervised classification algorithm can be found at [1].

Firstly, we manipulated the data with XploRe. The value of the spectrum frequency of the seven bands is to be extracted, where the total observations that we have $256 \times 256 \times 7 = 458752$ pixels. We then apply fuzzy clustering to all of the seven bands.

As a default, we set the value of $m=2$ and $c=5$. The choice of $m=2$ is due to its fast convergence compared to other numbers of m . Meanwhile the $c=5$ what we have chosen is based on the previous research [2].

We also conducted with the similar data using ISODATA algorithm. This process done by ERDAS software, in which this algorithm already implemented in ERDAS.

4 Result and Discussions

By analyzing the data, it is assumed that there is no more noise. Multi-resolution image is usually high correlated. Either when it is seen visually or numerically [6].

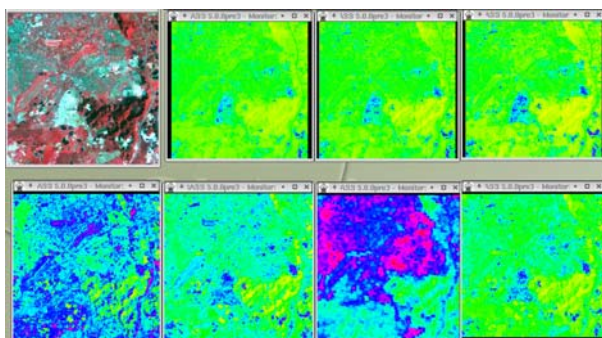


Figure 2: Scenes from test data and 7 band image of Ipoh City.

Figure 2 show the test data and all 7 bands of Ipoh City. [2] has carried out classification using an unsupervised ISODATA classification method. They classified into 15 classes. These classes were then clustered into 5 classes and named as: water, agriculture, grass, urban areas, barren and mining area.

However, it is known that classifying an urban area using Landsat TM data is not easy. Especially when it is done in a settlement and rural surrounded with meadows, brush land, and agricultural fields. When the area is fallow, then the brightness value obtained will be the same with the urban area itself. This result to be false by process of classification.

Next, we will see the performance algorithm of the suggested unsupervised classification (fuzzy clustering) and then compare this with the ISODATA classification methods.

We got the cluster results of the fuzzy clustering method from XploRe and transferred to GRASS for display purposes. We only choose the observation of pixels where the membership value is greater than 0.5. Pixels which have membership value smaller than 0.5, we leave it for further investigation. There

are around 800 pixels which are not belong to certain cluster but to two or three cluster together.

We also got the cluster results of ISODATA from ERDAS and transferred to GRASS for display purposes. All of the two clustering results together with the test images are shown in Figure 3.

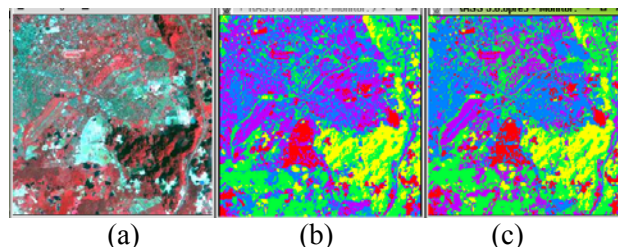


Figure 3. (a) Scenes of test image, (b) cluster with ISODATA method and (c) Fuzzy method.

Table 1 and Table 2 show the accuracy level of the two methods both ISODATA clustering and fuzzy clustering.

Table 1 : Accuracy assessment for ISODATA clustering

Class name	Referenced data						user 's accuracies
	W/S	F/A	U/B	G	B/M	Total	
Water/shadow	33	2	2	1	1	39	0.85
Forest/agriculture	1	108	8	5	2	124	0.87
Urban/built-up	2	6	67	7	3	85	0.79
Grass land	2	2	4	77	3	88	0.88
Barren/mining area	1	0	2	1	26	30	0.87
Total	39	118	83	91	35	366	
producer 's accuracies	0.85	0.92	0.81	0.85	0.74		
Overall accuracy = 0.85 %							

W/S=Water/Shadow;

F/A=Forest/Agriculture;

U/B=Urban/Built-up;

G=Grass Land;

B/M=Barren/Mining Area

Table 2: Accuracy assessment for fuzzy clustering

Class name	Referenced data						user 's accuracies
	W/S	F/A	U/B	G	B/M	Total	
Water/shadow	55	2	0	1	1	59	0.93
Forest/agriculture	1	112	2	1	0	116	0.97
Urban/built-up	0	6	59	7	3	75	0.79
Grass land	1	2	0	65	3	71	0.92
Barren/mining area	0	0	2	2	41	45	0.91
Total	57	122	63	76	48	366	
producer 's accuracies	0.96	0.92	0.94	0.86	0.85		
Overall accuracy = 0.91 %							

W/S=Water/Shadow;
 F/A=Forest/Agriculture;
 U/B=Urban/Built-up;
 G=Grass Land;
 B/M=Barren/Mining Area

We also presented the accuracy assessment of both methods. These accuracy assessment were done by ERDAS. We choose the pixel to be compared randomly. From the Figure 2(a), it shown clearly that an airport is lying on Grass Land. There is a relationship with the table of accuracy assessment, where for the Grass Land class, the accuracy is increased from 88% to 92%.

5 Conclusions

The application of fuzzy clustering to landsat TM data, specifically the landsat data from The Ipoh City area in Malaysia has been done. Using the previous result, it has detected 5 clusters for the area of investigation which are listed as follows: water or shadow, forest or agriculture, urban or built-up, grass land, and barren or mining area.

By choosing this 5 clusters, we conducted unsupervised classification with fuzzy clustering. Here we have come to the conclusion that the accuracy of the fuzzy clustering method is better than the ISODATA method of cluster analysis. For example, if we look at the Grass Land class, the accuracy is increased from 88% to 92%. The overall accuracy is also sinificantly increased from 85% to 91%.

References:

- [1] Ball, G. and Hall, D., 1967 A clustering technique for summarizing multivariate data, *Behavioral Science*, **12**:152:155.
- [2] Bawahidi, K., Affan, M., Md. Said, M.A. and Razali, R, 2003, *The application of remote sensing and GIS in monitoring Land-Use/Land Cover changes for a study area Ipoh*, Kuala Lumpur Conference.
- [3] Bezdek, J.C., Ehrlich, R., and Full, W., 1984, FCM: the Fuzzy c-Means Clustering Algorithm, *Computers and Geosciences*, **10**, 191-203.
- [4] Guenther, O., Radermacher, F. J., and Riekert, W.-F., 1995. *Environmental Monitoring: Models, Methods and Systems*, : Avouris, N. M. Page, B. (eds.), Environmental Informatics, Kluwer Academic Publishers.
- [5] Hellendorn, H. and Driankov, D., 1998, *Fuzzy: Model Identification*, Springer-Verlag, Heidelberg.
- [6] Schowengerdt, R.A., 1997, "Remote Sensing : Models and Methods for Image Processing", Second Edition, Academic Press.
- [7] Zadeh, L.A. 1965. „Fuzzy sets. Information and control”, **8**, 338-353.