ANALYSIS OF THAICHOTE BAND CHARACTERISTICS USING UNSUPERVISED PIXEL-BASED CLASSIFICATION

Tanee KAMKHET

Geo-Informatics Scientist, Institute of Space Knowledge Development,
Geo-Informatics and Space Technology Development Agency (Public organization),

120 (Building B) Chaeng Wattana Road, Laksi District, Bangkok 10210, Thailand;
Tel: +66(0) -2141-4410; Fax. +66(0) -2143-9594 to 5
E-mail: Tanee@gistda.or.th

KEY WORDS: Thaichote, land-cover, Iso-data, K-means

Abstract: Although object-based classification has been widely applied to classify remotely sensed data, unsupervised pixel-based classification was not yet thoroughly applied to each band of Thaichote (THEOS). Objectives of this research are to study characteristics of each band of Thaichote using unsupervised pixel-based classifications and to present an option of applying classification methods to Thaichote band by band. To identify spectral response of each band, the northeast region of Surin province was selected as a study area due to its land cover diversity.

Thaichote images acquired in April 2012 were geometrically corrected. Then, each band was individually classified using Iso-data and K-means methods. The classified images were validated against data from Land Development Department, the national agency in land use planning; visual interpretation and data from field survey, and then kappa statistics were calculated. Kappa index of each land cover type based on each band classification show how good a single band can classify each land cover type leading to a guideline to select a single band, an appropriate classification method that gives most accurate result.

In conclusion, all the results can be distinguishable on a difference method and the effect of any factor. Furthermore, the advantage or disadvantage of all results could be demonstrated and could support to determine its adaptation to the requirement remote sensing data researcher.

INTRODUCTION

Classification from remote sensing includes two principles: supervised and unsupervised classifications. Each was selected as a difference method, (Lewis and Disney, 2012) depending on the objective, accuracy, or requirement of the researcher. Mostly, at present the data from satellite, with multi-spectral characteristics, was calculated from color composite arrangement.

In this study, the primary purpose is to test unsupervised classification from multi-spectral of Thaichote (THEOS) imagery. However, the consideration to split the wavelength as independent. The result showed the advantage and disadvantage, limitation and accuracy of analysis comparing with land-cover from local administration offices.

Another objective of the research is to test the result of unsupervised classification in Iso-data and K-mean methods. However, in order to reduce a self-conflict of any data, it was necessary to use the same condition and statistics. The results of using which type are to be decided by the researcher.

STUDY AREA AND DATA USED

The study area chosen for this research is Surin province (15^{o} 14’ to 15^{o} 20’ N latitude and 103^{o} 46’ to 103^{o} 52’ E longitude), with 93 sq.km covering five sub-district: Nong Bua of Thatum District; Phon Ko of Sanom District and Rattanaburi; Nam Khiao; Nong Bua Ban Sub-district, Ratanaburi District.

The reasons for selection of this area are based on the Thaichote image being clear, the area having various types of land use with support for the interpretation. Thaichote multispectral data was acquired on April 20, 2012. with 15 meter pixel size, WGS 1984 datum and UTM zone 48 north.

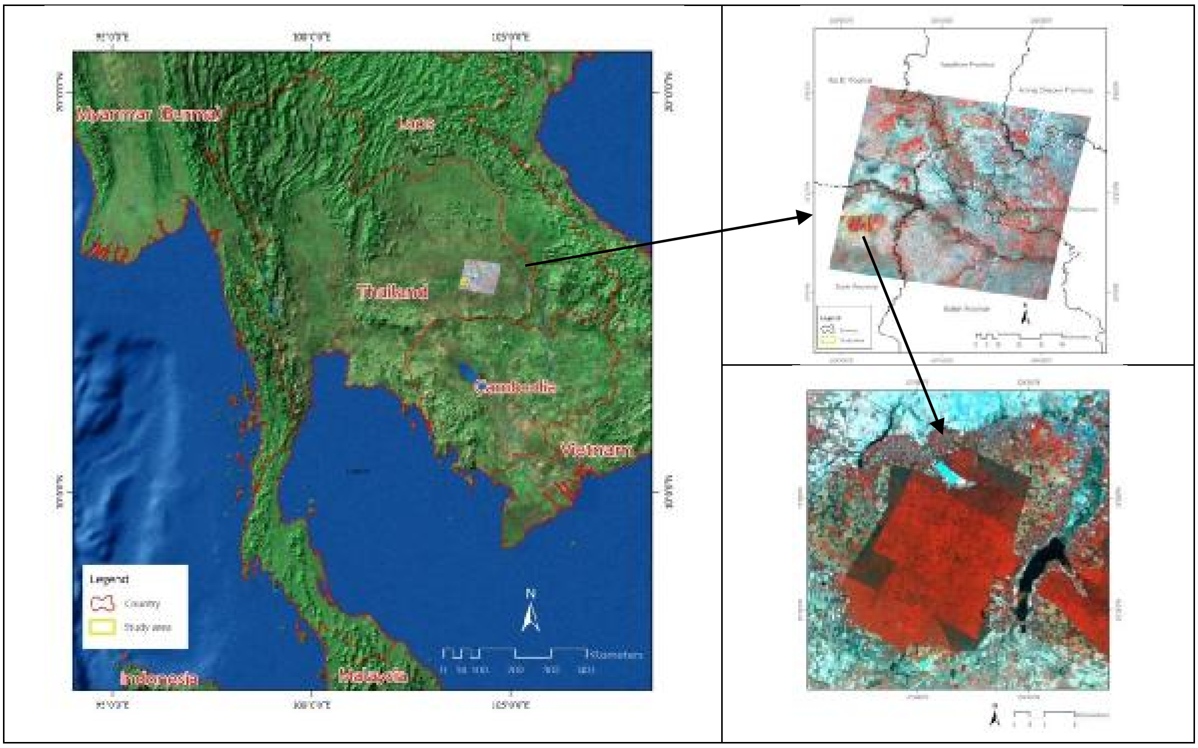


Figure 1: The study area

METHODOLOGY

The analysis started from pre-processing, then unsupervised classification by Iso-data and K-mean method was carried out with final processing in post-classification and accuracy assessment.

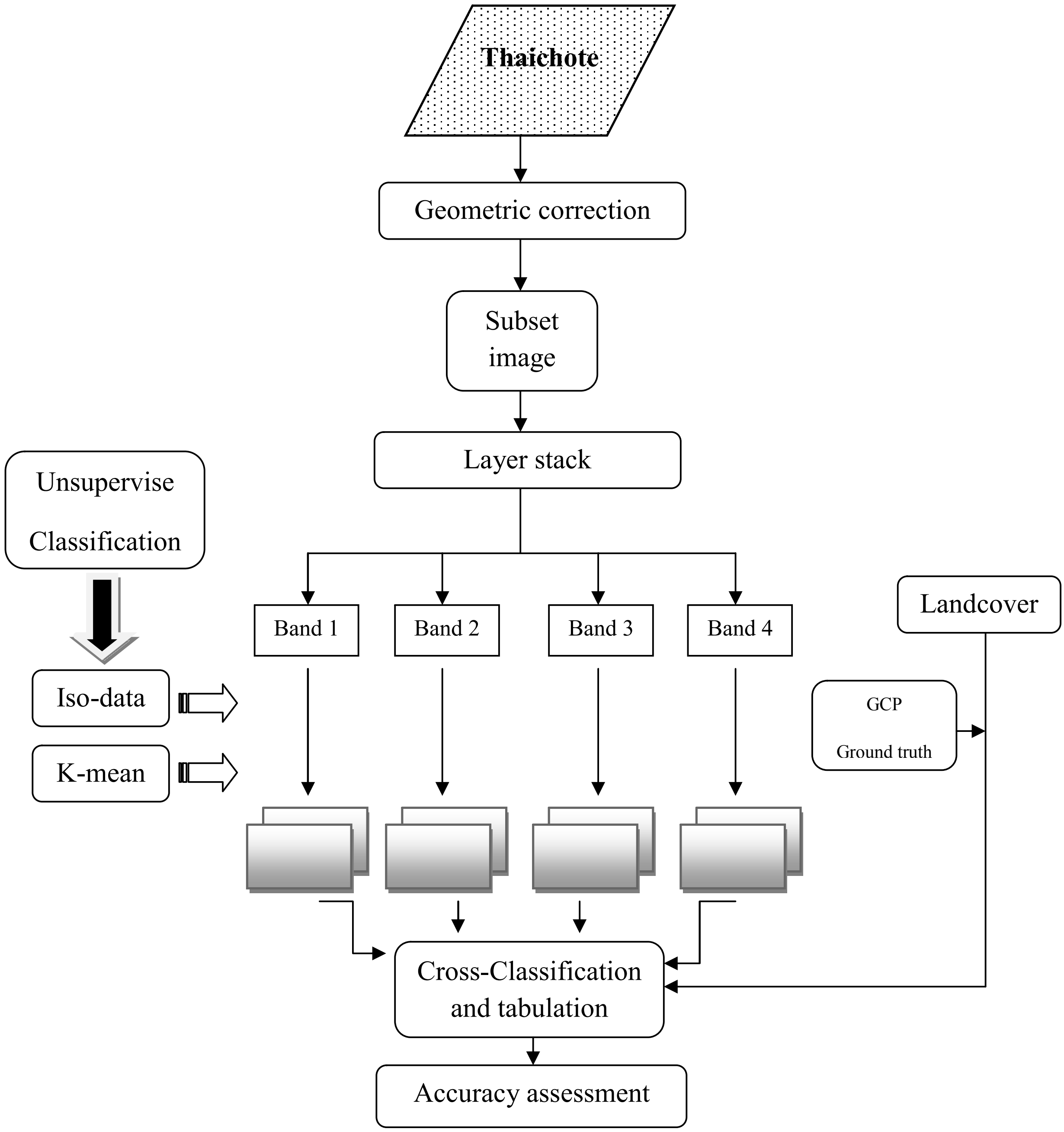


Figure 2: Flow diagram of the study

1. Pre-processing

Pre-processing is for the geometric correction of the images to map registration. The reference point based on the series L7018 of the Royal Thai Survey Department, with the scale of 1:50,000 was used. After that, image correction was performed, that is image to image registration which will provide the satellite and land cover data, with the same position reference. In addition, the process on control point covers the whole image, coupling with the nearest neighbor calculation, and the equation of polynomial order 1, including RMS error, as given, within 1 pixel, or 15*15 meters.

The corrected image, then, being cropped with the study boundary, were layer stacked to find each profile. Band 1 to 4 in raster format, with data type using 8-bit, with 0-255 range and a degree of difference in each band were used with the software Imagine 9.2 processing.

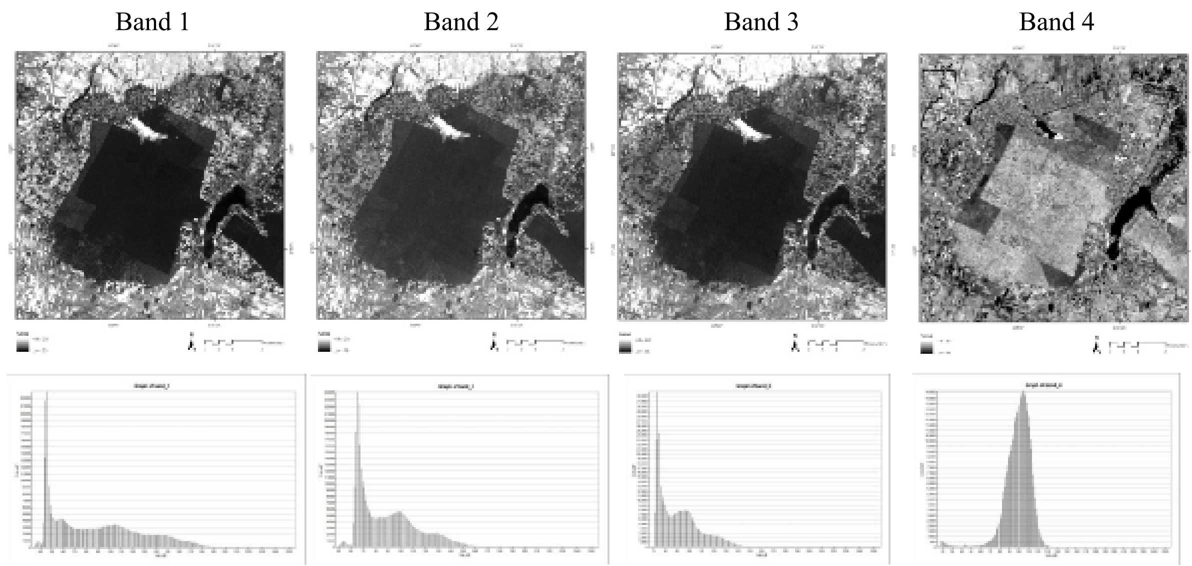


Figure 3: Band 1-4 and profile

	Wavelength (µm)	Pixel data		Mean	Median	Mode	Std. Dev.
		Min.	Max.				
Band 1 <Red>	0.45 – 0.52	33	255	87.96	81	46	38.73
Band 2 <Green>	0.53 – 0.60	48	255	89.17	84	65	24.06
Band 3 <Blue>	0.62 – 0.69	66	249	90.97	88	73	17.49
Band 4 <Near infrared>	0.77 – 0.90	14	167	98.75	101	105	14.84

Table 1: Statistic of Band 1- 4

2. Unsupervised classification

The classification of the data layer is based on the administrative which land cover, has 5 classes: Agricultural land, Forest land, Miscellaneous land, Urban and Built-up land, and Water body.

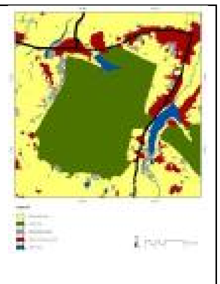
Administrative Landcover	Definition	Number of Pixel (sum 406,836)	Area (sq.km)	
Agricultural land	Paddy field, Cassava, Mixed field crop	199,428	44.87	
Forest land	Dense forest plantation	135,430	30.47	
Miscellaneous land	Grass and Scrub	15,128	3.40	
Urban and Built-up land	Village and Abandoned area	43,956	9.89	
Water body	Reservoir and Farm pond	12,894	2.90	

Table 2: Administrative Landcover

Iso-data unsupervised classification calculates class means evenly distributed in the data space then iteratively clusters the remaining pixels using minimum distance techniques.

Each iteration recalculates means and reclassifies pixels with respect to the new means. Iterative class splitting, merging, and deleting is done based on input threshold parameters.

K-mean unsupervised classification calculates initial class means evenly distributed in the data space then iteratively clusters the pixels into the nearest class using a minimum distance technique. Each iteration recalculates class means and reclassifies pixels with respect to the new means. (Tou and Gonzalez, 1974)

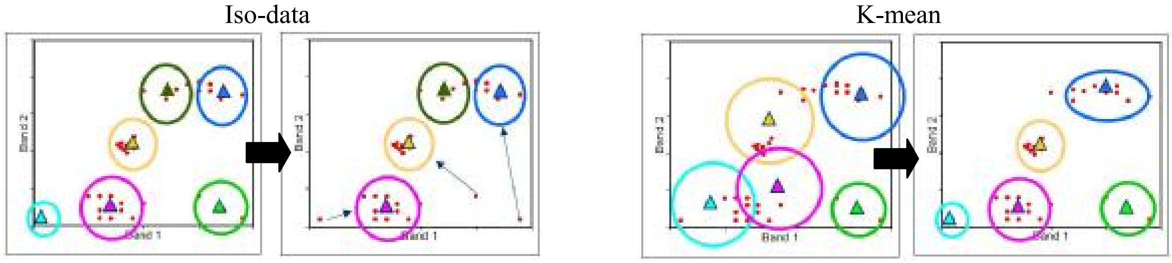


Figure 4: Iso-data and K-mean (Hutson, 2005)

The author uses software Envi 4.7 in the data classification, unsupervised, with Iso-data, under these parameters : number of class equals 5, maximum iterations equal 100, change threshold equals 5.00, minimum pixel in class equals 1, maximum class Stdv. equals 1, minimum class distance equals 5, and maximum merge pairs equal 2. Furthermore, the k-mean given number of class equals 5, maximum iterations equals 100, and change threshold equals 5.00.

After the classification processing, the result in each band was re-checked with the accuracy, by the existing land cover data, by cross classification and tabulation. In this, the yields should be classified into two parts: overall accuracy of the whole image and accuracy of each land cover.

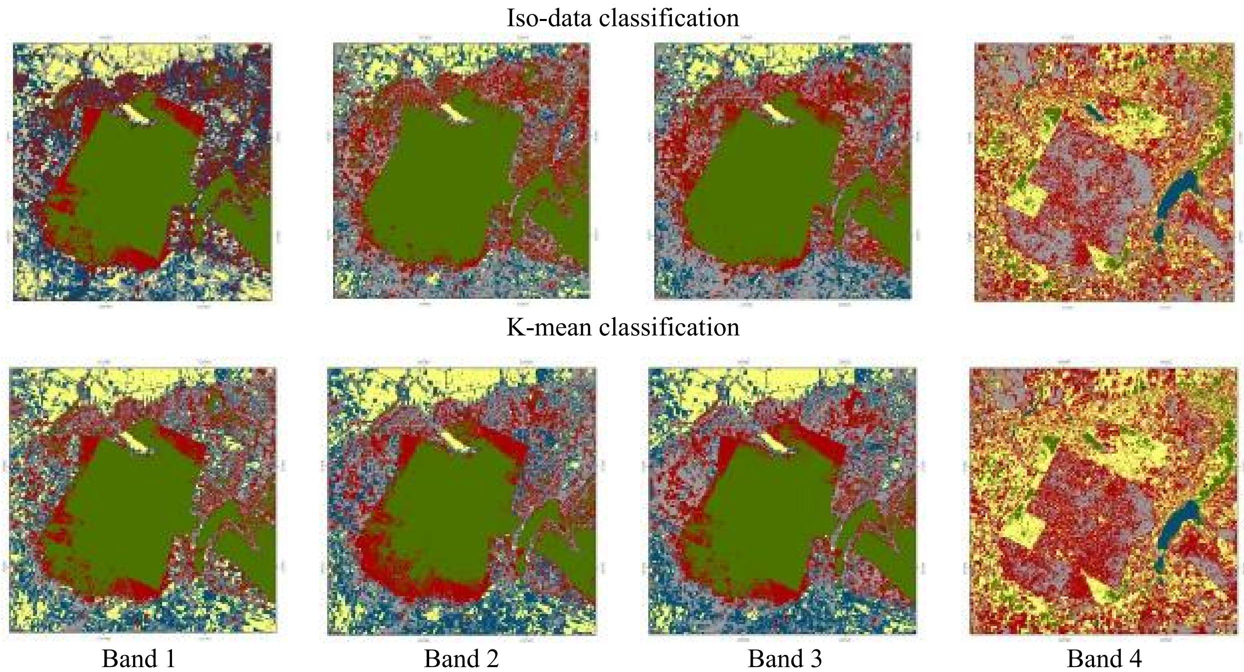


Figure 5: Results of Iso-data and K-mean classification

3. Result and Accuracy Assessment

This study used the Kappa Index, which usually attributed to Cohen' s Kappa [1960]. This is an index of interrater reliability that is commonly used to measure the level of agreement between two appraisement when both classify the same object. (Landis and Koch, 1977)

The classification of land cover, all 5-band, used Iso-data and K-mean methodologies. Although, the results of classification and degree of accuracy may differ, but the Kappa index from the analysis shows that the band 1 classification, with K-mean, gives high accuracy, with Kappa index .3050; on the other hand, the band 4 classification by K-mean gives the lowest accuracy, with Kappa index .0289. In the conclusion, the overall of the accuracy is not the indicator for each land cover.

Iso-data	Number of Pixel			
	Band 1	Band 2	Band 3	Band 4
Agricultural land	44,240	31,427	27,374	105,343
Forest land	78,199	161,350	148,704	36,269
Miscellaneous land	67,799	88,417	99,840	109,161
Urban and Built-up land	88,918	74,435	78,330	148,367
Water body	127,680	51,207	52,588	7,696
Overall Kappa	0.0723	0.2884	0.2789	0.0301
K-mean	Number of Pixel			
	Band 1	Band 2	Band 3	Band 4
Agricultural land	64,473	52,553	46,449	135,152
Forest land	131,863	113,562	119,134	33,192
Miscellaneous land	65,284	76,323	88,697	71,253
Urban and Built-up land	66,122	76,265	72,669	161,655
Water body	79,094	88,133	79,887	5,584
Overall Kappa	0.3050	0.2667	0.2772	0.0289

Table 3: Results from Iso-data and K-mean classification

In term of Kappa Index of Agreement (KIA) which classified as each land cover, the results are as follows:

- Agriculture land of band 1, with K-mean, the KIA is .1623 or equals to 12.90 sq.km/ 28.80%
- Forest land of band 2, with Iso-data, the KIA is .8820 or 28.23 sq.km/ 92.64%
- Miscellaneous land of band 3, with K-mean, the KIA is .2891 or 1.48 sq.km/ 43.52%
- Urban and built-up land of band 2, with Iso-data, the KIA is .3101 or 4.25 sq.km/ 42.97%
- Water body of band 4, with Iso-data, the KIA is .4126 or 1.23 sq.km/ 42.41%

Conclusions

Unsupervised classification of land cover with Iso-data and K-mean yields a different degree of accuracy. In this study, the overall accuracy has insignificance, in order to justify that the outcomes may be used as reference in all land cover. Hence, if the data is to be cited as a reference, this must be chosen in a specific study. Furthermore, there are some preposition and assumption of the study, among others:

1. Each land cover classification consorts to the characteristic of the reflection of the object in a particular wave, which is enough for the primary data on land cover.

2. In considering agricultural land, the result suggested to the highest reaction in Band 4 and and on the contrary, the highest number is in Band 1, with the study showing that most of the land cover is the paddy field. The data was taken in dry season, therefore these were no plant covering the field.

3. For miscellaneous land, which, in fact, is unable to clearly classify the land cover; but in the real term, there are some scattered plants, or grass; then, the outcomes of the classification is not consorted with the hypothesis.

4. In water body, with the highest accuracy in Band 4, having lowest reflection; since the absorption of water, from .8 micron upwards, the yield is fine in water land cover classification.

5. In case of each landuse classification, the study might be specified, in order to obtain the good outcome, with only one wave range such as the waters.

6. Difficulties in grouping the data layer of the land cover for reference, since the spectral characteristic in satellite image is unable to be grouped in each boundary; then, there might be some error and inaccuracy of classification.

7. In this study, the specific area has been partially analyzed, the of each spectral wave length represents only that particular one; in addition, the completeness of the image, statistical parameters of the testing, satellite data acquisition period, or even the real landcover during the study, are all to be borne in mind in each study.

Reference:

2005, R. Hutson. Remote Sensing and Land Classification, Remote Sensing Group, Plymouth Marine Laboratory.

2012, P. lewis and M. Disney. Environmental Remote Sensing GEOG 2021. Department of Geography; University College; London.

1974, J. T. Tou and R. C. Gonzalez. Pattern Recognition Principles, Addison-Wesley Publishing Company, Reading, Massachusetts.

1977, J. R. Landis and G. C. Koch. The measurement of observer agreement for categorical data, Biometrics, 33:159-174.