2011 International Conference on Network and Electronics Engineering
IPCSIT vol.11 (2011) © (2011) IACSIT Press, Singapore

Finding and Detection of Outlier Regions in Satellite Image

Kitti Koonsanit and Chuleerat Jaruskulchai

Department of Computer Science, Faculty of Science, Kasetsart University,
Bangkok, Thailand
sc431137@hotmail.com, fscichj@ku.ac.th

Abstract. Outlier cluster detection has always been a hot research field of data mining. Outlier cluster detection is important in many fields. Automatic determination of the outlier cluster is often needed to eliminate that outlier cluster. In this paper, a method has been developed to determine the outlier regions in satellite image using a data mining algorithm based on the co-occurrence matrix technique in order to determinate that outlier. Our method consists of four stages, the first stage estimate a number of cluster by co-occurrence matrix, the second stage cluster dataset by automatic clustering algorithm, the third stage detect outlier regions by automatic threshold value and the final stage defines outlier regions, which are lower than threshold value, to be outlier regions. The results from the tests confirm the effectiveness of the proposed method in finding the outlier regions.

Keywords: Outlier regions, Anomaly Detection, determination outliers cluster, co-occurrence statistics, Outlier detection

1. Introduction

In 1980, Hawkins made the definition of it: an outlier is an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism [1]. Usually, this kind of data has special behaviour or model. In effective data set, outlier is a small part and recognized as the by product of clustering [2]. So, outlier is always cancelled or neglected simply. However, certain outlier probably is the real reflection of normal data. These data are worthy to be study more.

In this paper, we propose a new easy method for automatically estimating the outlier regions in unlabeled data set. Pixel clustering technique in a colour image is a process of unsupervised classification of hundreds thousands or millions pixels on the basis of their colours. In this paper, a method has been developed to determine the outlier regions in satellite image clustering application using a data mining algorithm based on the co-occurrence matrix technique.

2. Related Work

2.1. Background

Mutispectral imaging is characterized by its ability to record detailed information about the spectral distribution of the received light. Mutispectral imaging sensors typically measure the energy of the received light in tens or hundreds of narrow spectral bands in each spatial position in the image, so that each pixel in a mutispectral image can be represented as a high-dimensional vector containing the sampled spectrum. Since different substances exhibit different spectral signatures, mutispectral imaging is a well-suited technology for numerous remote sensing applications including target detection. When no information about the spectral signature of the desired targets is available, a popular approach for target detection is to look for objects that deviate from the typical spectral characteristics in the image. This approach is commonly referred to as anomaly detection [3], and is related to what is often called outlier detection in statistics. If targets are small compared to the image size, the spectral characteristics in the image are dominated by the background. An important step in outlier regions detection is often to compute a metric for correspondence with the background, which then can be threshold to detect objects that are unlikely to be background objects.

Two approaches are of particular interest. One was developed by Reed and Yu [4][5][6] and is referred to as the RX detector (RXD), which has shown success in outlier detection for multispectral and hyperspectral images [7][8]. Another was proposed in [9][10] and is referred to as low probability detection (LPD), which was designed to detect targets with low probabilities in an image. The benchmark of anomaly detection is RX algorithm which is derived from the Generalized Likelihood Ratio Test (GLRT) with the assumption of Gaussian background [11]. However, background may be consisted of different ground cover types in real remote sensing images, such as water body, grass land, trees. This will lead to miss detection in complex background. Many researches attempt to use the Gaussian mixture model [12][13]. In reference [12], Ashton employs K-means cluster clustering the image into a number of statistical clusters and models each cluster with the Gaussian distribution. Subsequently, Carlotto proposes a similar approach using vector quantization to reduce the computational time [13]. However, the K-means based algorithm only considers the spectrum information of background. This will lead to miss clustering background pixels during the cluster process. Besides, all the methods, these methods are unsuitable for our application that needs to implement software fast and to ease the difficulty of implement software for beginners. In this paper, propose an outlier regions detection algorithm.

3. The Proposed Algorithm

In this paper, we propose a new method for determination of the outlier regions, which is based on co-occurrence matrix scheme. While a traditional co-occurrence matrix specifies only the transition within an image on horizontal and vertical directions. The proposed method can be used to automatically select a k range in multispectral satellite image. The proposed technique consists of four main steps: estimation, clustering, threshold and region detection.

3.1. Estimate number of cluster

The proposed technique first, the co-occurrence matrix scheme is employed to automatically segment out the object region in an image. Then, the local maximum technique is used to count a number of regions, which a number of cluster.

Our definition of a co-occurrence matrix [15][16][14][17] is based on the idea that the neighboring pixels should affect region of clusters. Hence, we define a definition for a co-occurrence matrix by including the transition of the gray-scale value between the current pixel and adjacent pixel into our co-occurrence matrix illustrated in figure 1 and figure 2.

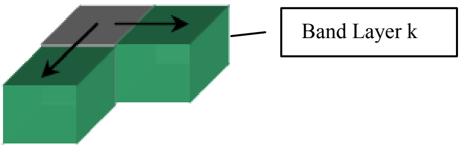
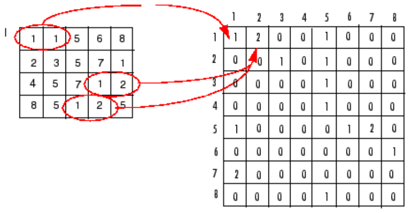


Fig. 1: right and bottom of pixel in a co-occurrence matrix



1	1	5	6	8
2	3	5	7	1
4	5	7	1	2
8	5	1	2	5

	1	2	3	4	5	6	7	8
1	1	2	0	0	1	0	0	0
2	0	0	1	0	1	0	0	0
3	0	0	0	0	1	0	0	0
4	0	0	0	0	1	0	0	0
5	1	0	0	0	0	1	2	0
6	0	0	0	0	0	0	0	1
7	2	0	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0

Fig. 2: Creating a Co-Occurrence Matrix

Let F be the set of images. Each image has dimension of P×Q. Let t_{ij} be an element in a co-occurrence matrix depending upon the ways in which the gray level i follows gray level j

$$t_{ij} = \sum_{x=1}^{P} \sum_{y=1}^{Q} \delta \begin{cases} (F_k(x,y) = i) \ and \ (F_k(x,y+1) = j) & or \\ (F_k(x,y) = i) \ and \ (F_k(x+1,y) = j) \end{cases}$$

Where $\delta = 1$

$, \delta = 0 \quad otherwise .$ (1)

where F_k denotes the kth band in the image set, F

If $s, 0 \le s \le L-1$ is a threshold. Then s can partition the co-occurrence matrix into 4 quadrants, namely A, B, C, and D shown in Figure.3.

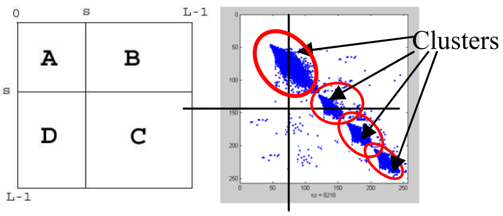


Fig. 3: An example of blocking of co-occurrence matrix

Since two of the quadrants shown in Figure.3, B and D, contain information about edges and noise alone, they are ignored in the calculation. Because the quadrants, which contain the object and the background, A and C, are considered to be independent distributions.

The idea of proposed method is to select the results of co-occurrence matrix into a diagonal matrix. After threshold processing, the result of diagonal matrix was shown in figure 3 (right hand side). Diagonal matrix is used to show some clustered pixels. The gray level corresponding to local maximum which give the optimal number for object- classification in image as shown in figure 6.

3.2. Clustering

After K estimate processing, we got a number of cluster, we cluster data by clustering algorithm such as k-means which is the most popular clustering techniques.

3.3. Automatic Thresholding

After cluster processing, thresholding techniques was assigned an outlier score to each top instance as shown in figure 6 (A,B,C,D and X) in the test data depending on the degree to which that instance is considered an outlier.

In this paper, we apply a method for automatic thresholding, which is based on standard deviation. Standard deviation is a statistical evaluate of spread or variability. First, we have to find the mean for standard deviation. Mean is represented by the division of sum of all values and the total number of values. The standard deviation is the root mean squares deviation of the values from their arithmetic mean. It is calculated by take the square root of the variances and is symbolized by s.d, or s. as shown in (2).

$$\sigma = \sqrt{\frac{\sum (x - Mean)^2}{n-1}} \tag{2}$$

Where Σ = sum of , x = individual point
$Mean$ = Mean of all point and n = sample size (Number of point)

The proposed method can be used to select $Mean - \sigma$ is an appropriate automatically threshold values in order to estimate a cut-off threshold value to select the outlier cluster as shown in Figure.6. Thus the output of our techniques use a cut-off threshold to select the outlier clusters which less than $Mean - \sigma$ value.

3.4. Automatic Outlier Regions

Finally, after threshold processing, we get outlier region which is an outlier cluster as shown in fig 4-7.

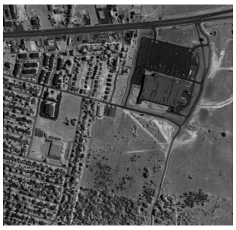


Fig. 4: an example of original image

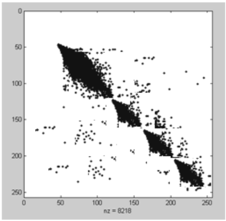


Fig. 5: an example of co-occurrence matrix of image from Figure 4.

Figure 5 illustrates anomalies in a simple 2-dimensional data set. The data has 4 normal regions B, C, D and X, since most observations lie in this one region. Points that are different from the regions, e.g., points in region X, are anomalies or outlier region.

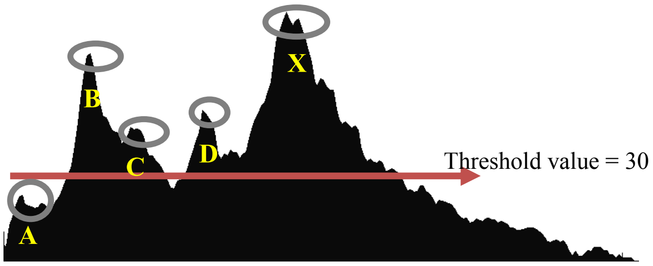


Fig. :6 Example histogram from diagonal of co-occurrence

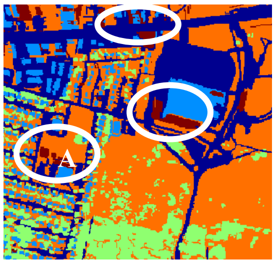


Fig. 7: "A Cluster" = a result of outlier region

4. Experiment and Results

4.1. Dataset

We used the sets of raw data from different CCD Multispectrum images. [18] The dataset are obtained from small multi mission satellite project (SMMS), a department of Electrical Engineering, Kasetsart University. We would like to analyze data, which was registered in Thailand.

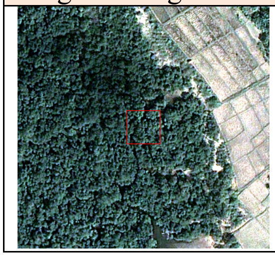
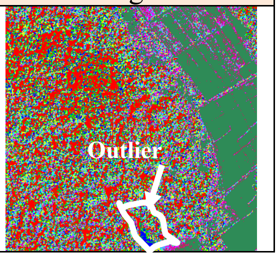
4.2. Unsupervised Classification Method

The experiments performed in this paper use the simple K-mean from the Weka software package. [19]

4.3. Experimental Result

Our experiment was tested with CCD Multispectrum images and shown in Table.2. The experiments demonstrate the robustness and effectiveness of the proposed algorithm.

Table 2. Result of finding outlier regions

Original image	Outlier Regions
	

From the experimental result, it was found that outlier regions solved by co-occurrence statistics techniques gives the nearest outlier regions with ground truth. It can be noticed that the outlier regions in clustering between original images and solving by co-occurrence statistics techniques are very closed.

The outcome of this research will be used in further steps for analysis tools in satellite image mining that finds anomalous clusters to visualize the satellite image such as natural resources and agricultural. A result of this research was developed to provide users have been processed to view and analyses the satellite image. We hope that it can be used as a tool and help develop research in satellite image software.

5. Conclusions

In this paper, a method has been developed to determine the outlier regions in satellite image using a data mining algorithm based on the co-occurrence matrix technique in order to determinate that outlier regions. Our method consists of four stages, the first stage estimate a number of cluster by co-occurrence matrix, the second stage cluster dataset by automatic clustering algorithm, the third stage detect outlier regions by automatic thresholding and the final stage defines regions, which are lower than threshold value, to be outlier regions. The proposed method was tested using data from unknown number of clusters with multispectral satellite image in Thailand. The results from the tests confirm the effectiveness of the proposed method in finding the outlier regions.

6. Acknowledge

This work was supported by budget for overseas academic conference from the faculty of science, Kasetsart University and the graduate school, Kasetsart University. The authors would like to thank TGIST.

7. References

[1] Hawkins. D. Identification of Outliers. London: Chapman and Hall. 1980.

[2] Liu Junling. Study and implementation of clustering and outlier detection algorithm[D]:[Master thesis].LIAO NING: Shenyang Institute of Computing Technology Chinese Academy of Sciences . 2006.

[3] D. W. 1. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," IEEE Signal Process. Mag., vol. 19, pp. 58-69,2002.

[4] I. S. Reed and X. Yu, “Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution,” IEEE Trans. Acoust., Speech, Signal Processing, vol. 38, pp. 1760–1770, Oct. 1990.

[5] X.Yu, I. S. Reed, and A. D. Stocker, “Comparative performance analysis of adaptive multispectral detectors,” IEEE Trans. Signal Processing, vol.41, pp. 2639–2656, Aug. 1993.

[6] X. Yu, L. E. Hoff, I. S. Reed, A. M. Chen, and L. B. Stotts, “Automatic target detection and recognition in multispectral imagery: A unified ML detection and estimation approach,” IEEE Trans. Image Processing, vol.6, pp. 143–156, Jan. 1997.

[7] E. A. Ashton and A. Schaum, “Algorithms for the detection of sub-pixel targets in multispectral imagery,” Photogram. Eng. Remote Sens., pp. 723–731, July 1998.

[8] C. M. Stellman, G. G. Hazel, F. Bucholtz, J. V. Michalowicz, A. Stocker, andW. Scaaf, “Real-time hyperspectral detection and cuing,” Opt. Eng., vol. 39, pp. 1928–1935, 2000.

[9] J. C. Harsanyi, “Detection and classification of subpixel spectral signatures in hyperspectral image sequences,” Ph.D. dissertation, Dept. Elect. Eng., Univ. Maryland-Baltimore County, Baltimore, MD, 1993.

[10] J. C. Harsanyi,W. Farrand, and C.-I. Chang, “Detection of subpixel spectral signatures in hyperspectral image sequences,” in Proc. Amer. Soc. Photogram. Remote Sens., Reno, NV, 1994, pp. 236–247.

[11] R. Reed and X. Yu, "Adaptive multi-band CFAR detection of an optical pattern with unknown spectral distribution," IEEE Trans. Acoust.,Speech, Signal Process., vol. 38, pp. 293-305, 1990.

[12] E. A. Ashton, "Detection of Subpixel Anomalies in Multispectral Infrared Imagery Using an Adaptive Bayesian Classifier," IEEE Trans. Geosci. Remote Sensing, vol. 36, pp. 506-517, 1998.

[13] M.1. Carlotto, "A cluster-based approach for detecting man-made objects and changes in imagery," IEEE Trans. Geosci. Remote Sensing, vol. 43, pp. 374-387,2005.

[14] Kitti Koonsanit and Chuleerat Jaruskulchai, Automatic Determination of the Initialization Number of Clusters in K-means Clustering Application by Using Co-occurrence Statistics Techniques for Multispectral Satellite Image, The 2010 International Conference on Information Security and Artificial Intelligence (ISAI 2010), December 17-20, 2010, Chengdu, China

[15] N. R. Pal and S. K. Pal, "Entropic thresholding," Signal processing, vol. 16, pp. 97-108, 1989.

[16] T. Chanwimaluang and Guoliang Fan, “An efficient algorithm for extraction of anatomical structures in retinal images,” ICIP 2003 Proceedings, Sept 4-17, 2003

[17] K. Koonsanit, T. Chanwimaluang, D. Gansawat, S. Sotthivirat, W. Narkbuakaew, W. Areeprayolkij, P. Yampri and W. Sinthupinyo, “Metal Artifact Removal on dental CT Scanned Image by Using Multi-layer Entropic Thresholding and Label Filtering Technique for 3-D Visualization of CT images,“ Proc. of International Conference on Biomedical Engineering : ICBME 2008.IFMBE Proceedings, 13th International Conference on Biomedical Engineering , December, Singapore

[18] Small Multi-Mission Satellite (SMMS) Data Retrieved: May 26, 2010 from the World Wide Web:http://smms.ee.ku.ac.th/index.php

[19] Remco R. Bouckaert , “WEKA Manual,” WAIKATO University, pp.1-303, January 2010. Retrieved: May 26, 2010 from the World Wide Web: www.cs.uu.nl/docs/vakken/dm/WekaManual.pdf