

A Simple Anomaly Detection for Spectral Imagery Using Co-occurrence Statistics Techniques

Kitti Koonsanit, Chuleerat Jaruskulchai, and Apisit Eiumnoh

Abstract—Anomaly detection has always been a hot research field of data mining. Anomaly detection is important in many fields. Automatic determination of the anomaly cluster is often needed to eliminate that anomaly cluster. In this paper, a method has been developed to determine the anomaly regions in satellite image using a data mining algorithm based on the co-occurrence matrix technique in order to determinate that anomaly. 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 anomalous clusters by threshold value and the final stage defines clusters, which are lower than threshold value, to be anomalous clusters. 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 anomaly regions.

Index Terms—Anomaly detection, determination outlier cluster, co-occurrence statistics, outlier detection

I. INTRODUCTION

Clustering is a popular tool for data mining and exploratory data analysis. One of the major problems in cluster analysis is the determination of the outlier clusters or anomaly clusters in unlabeled data, which should be eliminated. Outlier is the data which has obviously difference with clustering. 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 behavior or model. In effective data set, outlier is a small part and recognized as the byproduct of clustering [2]. So, outlier is always canceled 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 anomaly clusters in unlabeled data set. Pixel clustering technique in a color image is a process of unsupervised classification of hundreds thousands or millions pixels on the basis of their colors. In this paper, a method has been developed to determine the anomaly clusters in satellite image clustering application using a data mining

algorithm based on the co-occurrence matrix technique. Therefore, automatic determination of the anomaly clusters can greatly help with the unsupervised classification of satellite Image.

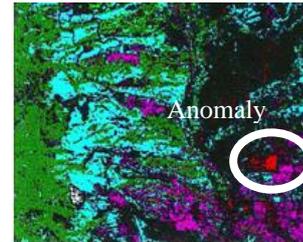


Fig.1. Example of anomaly detection on satellite image

II. RELATED WORK

Hyperspectral imaging is characterized by its ability to record detailed information about the spectral distribution of the received light. Hyperspectral 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 hyperspectral image can be represented as a high-dimensional vector containing the sampled spectrum. Since different substances exhibit different spectral signatures, hyperspectral 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 anomaly detection is often to compute a metric for correspondence with the background, which then can be thresholded to detect objects that are unlikely to be background objects.

Two approaches are of particular interest. One was developed by Reed and Yu [4]-[6] and is referred to as the RX detector (RXD), which has shown success in anomaly 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 mix-

Manuscript received June 9, 2012; revised July 27, 2012. This work was supported by Thailand Graduate Institute of Science and Technology (TGIST) is gratefully acknowledged. The scholar student ID is TG-22-11-53-005D and the grant number is TGIST 01-53-005.

Kitti Koonsanit is with Kasetsart University, Bangkok, Thailand (e-mail: sc431137@hotmail.com).

Chuleerat Jaruskulchai is with Kasetsart University, Bangkok, Thailand.

Apisit Eiumnoh is with National Center for Genetic Engineering and Biotechnology, Patumthani, Thailand.

ture 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 anomaly detection algorithm. The new method is compatible with the k-means algorithm and it overcomes the limitation of having to indicate the anomaly clusters by co-occurrence matrix and threshold defiend which is a apply technique in this proposed paper.

III. THE PROPOSED ALGORITHM

In this paper, we propose a new method for determination of the outlier clusters, 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 as shown in Table. I.

TABLE I: ALGORITHM FINDING OUTLIER CLUSTER

Algorithm Simple Anomaly detection	
Input:	A spectral image
Output:	Anomalous clusters
Processing	
1. Estimation:	Estimate a number of cluster by co-occurrence matrix [14][15][16][17]
2. Clustering:	Cluster image by a clustering algorithm and input a number of cluster from step 1.
3. Threshold:	Detect anomalous clusters by automatic threshold or input from user
4. Detection :	Define clusters which are lower than thresholding value as anomalous clusters

A. Estimation

The proposed technique consists of four main steps: estimation, clustering, threshold and detection. 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], [17], [14] 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 Fig. 2 and Fig. 3.

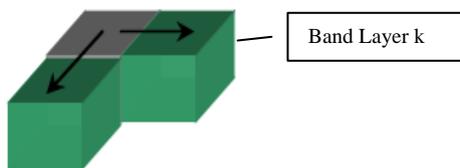


Fig. 2. Right and bottom of pixel in a co-occurrence matrix

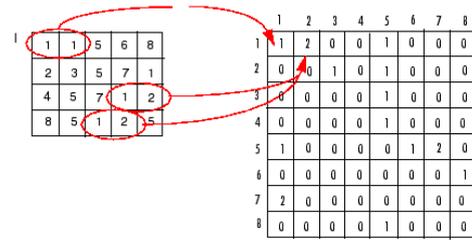


Fig. 3. Creating a co-occurrence matrix

Let F be the set of images. Each image has dimension of $P \times Q$. 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

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

where $\delta = 1, \delta = 0$ otherwise

where F_k denotes the k^{th} band in the image set, F

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

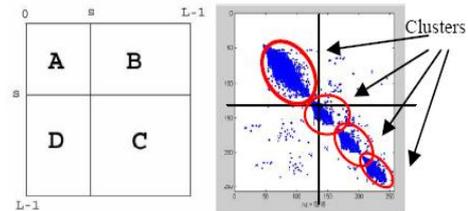


Fig. 4. An example of blocking of co-occurrence matrix

Since two of the quadrants shown in Figure.8, 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 6. 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 7.

B. 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.

C. Thresholding

After cluster processing, thresholding techniques assign an anomaly score to each instance in the test data depending on the degree to which that instance is considered an anomaly. Thus the output of such techniques is a ranked list of anomalies. An analyst may choose to either analyze top few anomalies or use a cut-off threshold to select the anomalies.

D. Detection

Finally, after threshold processing, we get anomaly region which is an outlier cluster as shown in fig 5- fig 8.

IV. EXPERIMENT AND RESULTS

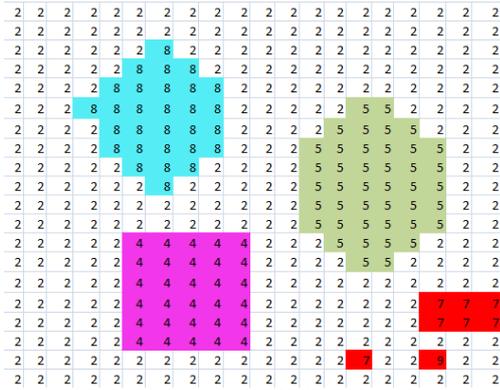


Fig. 5. An example of original image

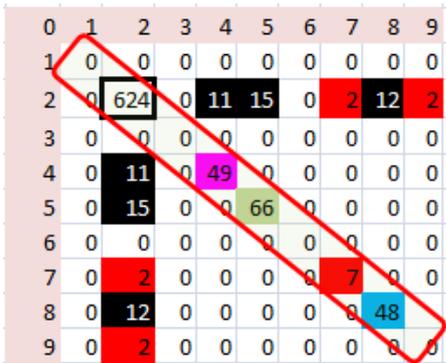


Fig. 6. An example of co-occurrence matrix of image from Figure 5.

Fig. 6 illustrates anomalies in a simple 2-dimensional data set. The data has 4 normal regions, A, B, C and D, since most observations lie in this one region. Points that are different from the regions, e.g., points in region X, are anomalies

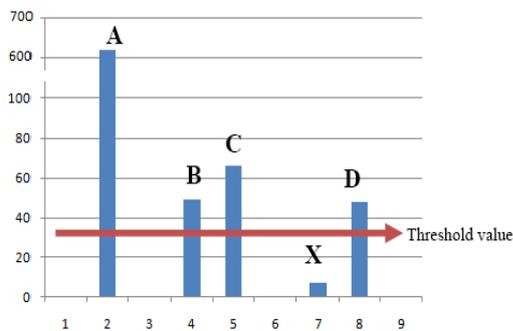


Fig. 7. Example histogram from diagonal of co-occurrence matrix (with threshold value = 30)

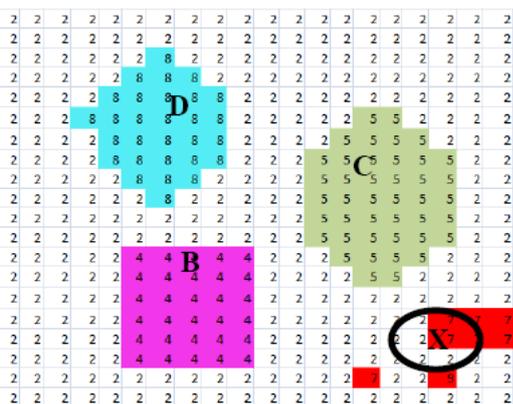


Fig. 8. X Cluster = a result of anomaly detection

A. Dataset

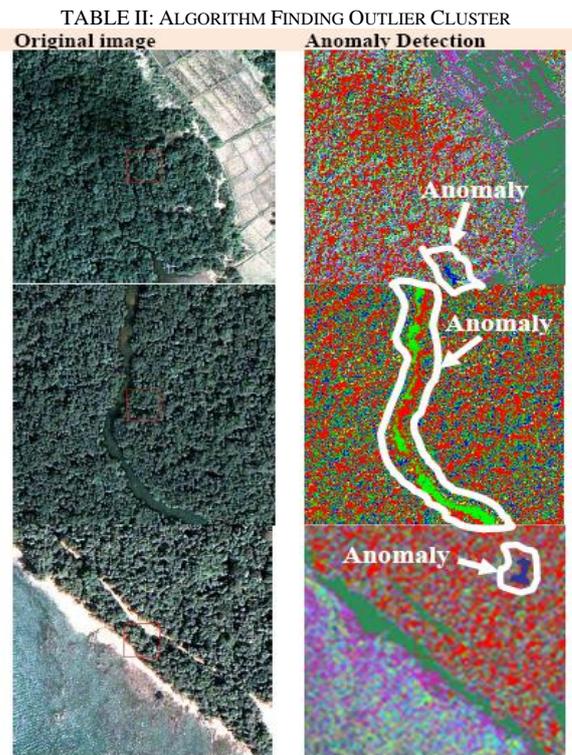
We used the eight sets of raw data from different CCD Multi-spectrum images [18]. The dataset are obtained from small multi mission satellite project (SMMS), a department of Electrical Engineering, Kasetsart University. We would like to analyze data, which was registered in Thailand and thus try to determinate of the initialization number of clusters in interesting areas of Thailand.

B. Unsupervised Classification Method

The experiments performed in this paper use the simple K-mean from the Weka software package [19]. The simple K-Mean is the common unsupervised classification method used with remote sensing data. The effectiveness of the K-Mean depends on reasonably accurate estimation of the k cluster for each spectral class. We solve a number of clusters by co-occurrence statistics.

C. Experimental Result

Our experiment was tested with CCD Multi-spectrum images and shown in Table.II. The experiments demonstrate the robustness and effectiveness of the proposed algorithm.



We created using java with weka develop on netbean IDE version 6.8 and tested the proposed process on satellite image data such as small multi-mission satellite (SMMS), and the experimental results show that our proposed process can determine the initialization of the anomaly cluster in satellite image clustering effectively. The algorithm provides promising performance in determining of the initialization of the anomaly cluster in clustering application by using co-occurrence statistics techniques for multispectral satellite image.

From the experimental result, it was found that anomaly clustering solved by co-occurrence statistics techniques

gives the nearest anomalous clusters with ground truth. It can be noticed that the anomaly cluster 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 data mining software in the future.

V. CONCLUSION

Anomaly detection has always been a hot research field of data mining. Anomaly detection is important in many fields. Automatic determination of the anomaly cluster is often needed to eliminate that anomaly cluster. In this paper, a method has been developed to determine the anomaly regions in satellite image using a data mining algorithm based on the co-occurrence matrix technique in order to determinate that anomaly. 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 anomalous clusters by threshold value and the final stage defines clusters, which are lower than threshold value, to be anomalous clusters. 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 anomaly regions.

ACKNOWLEDGMENT

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. The financial support from Thailand Graduate Institute of Science and Technology (TGIST) is gratefully acknowledged. The scholar student ID is TG-22-11-53-005D and the grant number is TGIST 01-53-005.

REFERENCES

- [1] D. Hawkins, "Identification of Outliers," London: Chapman and Hall, 1980.
- [2] L. 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. I. 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*, 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, and W. 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. I. 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] K. Koonsanit and C. Jaruskulchai, "Automatic Determination of the Initialization Number of Clusters in K-means Clustering Application by Using Co-occurrence Statistics Techniques for Multispectral Satellite Image," in *Proc. of 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 G. 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. Areepayolkij, 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," in *Proc. of International Conference on Biomedical Engineering : ICBME 2008. IFMBE Proceedings, 13th International Conference on Biomedical Engineering*, December, Singapore, 2008.
- [18] Small Multi-Mission Satellite (SMMS) Data Retrieved. [Online]. Available: <http://smms.ee.ku.ac.th/index.php>
- [19] R. R. Bouckaert, "WEKA Manual," WAIKATO University, [Online]. Available: www.cs.uu.nl/docs/vakken/dm/WekaManual.pdf



Kittit Koonsanit received his M.S. degree in computer science from Kasetsart University in 2008. He is currently a Ph. D. candidate in computer science, Kasetsart University, Thailand. His fields of interest include clustering, image processing, image segmentation, band selection, multispectral image, and medical imaging.



Chuleerat Jaruskulchai received her D.Sc. degree in computer science from George Washington University, School of Engineering and Applied Science, USA in 1998. She is currently an Associate Professor and lecturer in the Department of Computer Science, Kasetsart University, Thailand. Her fields of interest and research areas include information retrieval, clustering, text classification, and statistic modeling.



Apisit Eiumnoh received his Ph.D. degree in Soil Genesis & Classification from North Carolina State, UK. He is currently an Associate Professor in National Center for Genetic Engineering and Biotechnology, Patumthani, Thailand.