PAPER Special Section on Recent Advances in Multimedia Signal Processing Techniques and Applications

# Automatic Determination of the Appropriate Number of Clusters for Multispectral Image Data

Kitti KOONSANIT<sup>†a)</sup>, Nonmember and Chuleerat JARUSKULCHAI<sup>†</sup>, Member

SUMMARY Nowadays, clustering is a popular tool for exploratory data analysis, with one technique being K-means clustering. Determining the appropriate number of clusters is a significant problem in K-means clustering because the results of the k-means technique depend on different numbers of clusters. Automatic determination of the appropriate number of clusters in a K-means clustering application is often needed in advance as an input parameter to the K-means algorithm. We propose a new method for automatic determination of the appropriate number of clusters using an extended co-occurrence matrix technique called a tri-co-occurrence matrix technique for multispectral imagery in the pre-clustering steps. The proposed method was tested using a dataset from a known number of clusters. The experimental results were compared with ground truth images and evaluated in terms of accuracy, with the numerical result of the trico-occurrence providing an accuracy of 84.86%. The results from the tests confirmed the effectiveness of the proposed method in finding the appropriate number of clusters and were compared with the original co-occurrence matrix technique and other algorithms.

key words: determination a number of clusters, number of classes, tri-cooccurrence, clustering, co-occurrence, multispectral image

## 1. Introduction

Unsupervised classification is a popular tool for unlabeled datasets in data mining and exploratory data analysis. One of the major problems in cluster analysis is the determination of the appropriate number of clusters in unlabeled data, which is a basic input for most clustering algorithms. In this paper, we propose a new, easy method for automatically estimating the appropriate number of clusters in an unlabeled dataset. One of the most popular clustering techniques is the k-means technique. The results of the k-means technique depend on different factors such as a method of determination of the number of clusters. Such sensitivity is an important disadvantage of the k-means technique. Determination of the number of clusters is consistent with the aim of our research in order to find a number of clusters. In this paper, a method has been developed to determine the appropriate number of clusters in a satellite imagery clustering application using a data mining algorithm based on a novel co-occurrence matrix technique as shown in Fig. 1. Therefore, automatic determination of the appropriate number of clusters can greatly help with the unsupervised classification of satellite imagery.

a) E-mail: kitti.koonsanit@nectec.or.th





Fig. 1 Example of K-mean algorithm in a satellite imagery application.

# 2. Related Work

Many approaches to image classification have been proposed over the years [1]. Of these various methods, clustering is one of the simplest, and has been widely used in the clustering of grey level images [2]–[4]. Techniques such as k-means [5], isodata [5], and fuzzy c-means [6], [7] have been around for quite a while; however, their application to color images has been limited. Although color images have increased dimensionality by requiring three bands such as red, green and blue, clustering techniques can be easily extended to cope with this. The k-means and fuzzy c-means algorithms require the number of clusters to be known beforehand [8]-[11]. In order to supply the information required by the aforementioned algorithms, the user must have some knowledge about the image using own judgment. Our new method is compatible with the k-means algorithm and it overcomes the limitation of having to indicate the number of clusters by a novel co-occurrence matrix which is a technique proposed in this paper.

#### 3. Multispectral Imagery

Multispectral imaging has been gaining popularity and has been gradually applied to many fields besides remote sensing. Multispectral data provides unique information about material classification and reflectance analysis in general. However, due to the high dimensionality of the data, both

Manuscript received April 29, 2011.

Manuscript revised September 9, 2011.

<sup>&</sup>lt;sup>†</sup>The authors are with Kasetsart University, Thailand.



Fig.2 Example of a multispectral image containing BGR and NIR bands.



**Fig. 3** Example of a multispectral wavelength containing BGR and NIR bands.

human observers as well as computers have difficulty in interpreting this wealth of information. A multispectral image is one that captures image data at specific frequencies across the electromagnetic spectrum. The wavelengths may be separated by filters or by the use of instruments that are sensitive to particular wavelengths, including light from frequencies beyond the visible light range such as infrared. Multispectral imaging was originally developed for space-based imaging and can allow the extraction of additional information that the human eye fails to capture with its receptors for red, green and blue frequencies. Multispectral images are the main type of images acquired by remote sensing (RS) radiometers. Dividing the spectrum into many bands, multispectral is the opposite of panchromatic which records only the total intensity of radiation falling on each pixel. Usually satellites have 4 to 7 or more radiometers as shown in Fig. 2. Each one acquires one digital image (in remote sensing, this is called a scene) in a small band of visible spectra, ranging from 0.7  $\mu$ m to 0.4  $\mu$ m, called the red-green-blue (RGB) region, going to infra-red wavelengths of  $0.7 \,\mu\text{m}$  to 10 or more  $\mu$ m, classified as NIR-Near InfraRed, MIR-Middle InfraRed and FIR-Far InfraRed or Thermal as shown in Fig. 3. An example of the four scenes comprising a four-band multispectral image is:

- Blue, 450–515...520 nm, used for atmospheric and deep water imaging, reaching to a depth of 150 feet (46 m) in clear water.
- Green, 515...520–590...600 nm, used for imaging of vegetation and deep water structures, reaching to 90 feet (27 m) in clear water.
- Red, 600...630–680...690 nm, used for imaging of man-made objects, reaching in water to 30 feet (9.1 m) deep, and for soil and vegetation.
- Near infrared, 750–900 nm, primarily for imaging of vegetation.

Multispectral data provides unique information about material classification and reflectance analysis in general.

Га	ıbl	le	1	]	K	-n	nea	ns	a	lg	or	it	h	n	1
----	-----	----	---	---	---	----	-----	----	---	----	----	----	---	---	---

k- means Algorithm				
Input:	k : the number of desired clusters			
Output:	A set of k clusters			
Processing				
1. Select k objects from D as initial cluster centers				
2. Form k clusters by assigning each object to its closest center				
3. Recomputed the center of each cluster				
4. Until co	4. Until centers do not change			

Although multispectral images provide abundant information about bands, their high dimensionality also substantially increases the computational burden. However, due to the high dimensionality of the data, both human observers as well as computers have difficulty interpreting this wealth of information. An important task in multispectral data processing is to segment the multispectral image effectively. In general [12], a good number of clusters can be obtained by running an algorithm many times with a different number of clusters and comparing the results with a specified criterion. In this paper, we propose a new determination of the appropriate number of clusters in the application of satellite image segmentation for multispectral imagery.

## 4. K-means Method

The K-means algorithm [13], [14] is the simplest clustering algorithm and is widely used. K-means requires an input which is a predefined number of clusters. This input is named k which is consistent with our aim of research in order to find an appropriate number of clusters. The k-means method aims to minimize the sum of squared distances between all points and the cluster center. This procedure, shown in Table 1, consists of the following steps, as described by Tou and Gonzalez [5].

Initially, the K number of clusters should be chosen as a start. The initial step involves choosing a set of K instances as centers of the clusters. The set is often chosen such that the points are mutually "farthest apart" in some way.

Next, the algorithm considers each instance and assigns it to the cluster which is closest. The cluster centroids are recalculated either after each instance assignment or after the whole cycle of re-assignments. This process is iterated.

## 5. Problem Statement

From Table 1, it is obvious that the final result of clustering depends on number of clusters. An optimal number of clusters can be determined numerically by running an algorithm many times each with different number of clusters and comparing the results using a specific criterion [12].

The first step of K-means algorithm is to choose a number of clusters. Once the K-means algorithm has been processed, the input parameters of K-means algorithm need to be estimated and defined. Several parameter estimation methods are available for K-means clustering algorithm for example Xmeans. The aim of our research is to find the proper number of clusters for a multispectral imagery. In this section, we present typical parameter estimation methods for the K-means clustering algorithm. This section presents an overview of the previous estimation input parameter estimation approaches of the K-means algorithm, without paying much attention to the algorithm in detail. Basically, there are number of papers and articles that have been published regarding the approaches for choosing the optimal number of clusters.

## 5.1 Original Approach

Basically, the idea of this approach is to find the balance between two variables: the number of clusters (K) and the average variance of the clusters [15]. Of course, as the number of clusters increases, the average variance decreases (up to the trivial case of K = n and variance = 0). We assume that experimenter has already run the algorithm for several values of K. Then the experimenter can use the number of clusters at the intersection point of the curves. In basic data analysis, there is no universal approach that works for all cases. The experimenter still has to use his own judgment [15].

## 5.2 Model Selection Approach

The idea of this approach is to find a reasonably number of cluster. The experimenter only specifies a range in which the true K reasonably lies, and the output is not only the set of centriod, but also a value for K in this range, which scores best by model selection criterion such as Bayesian information criterion (BIC), AIC (Akaike Information Criteria), Davis Bouldin Index, and Confusion Matrix. An example of the popular ideas is Xmean [15] which based on BIC method to maximize the Bayesian information criterion (BIC).

## 5.3 Minimum Description Length Approach

The idea of this approach is to start with a large value for k and keep removing centroids (reducing k) until it no longer reduces the description length [16].

## 5.4 Gaussian Distribution Approach

The idea of this approach is to start with one cluster, then keep splitting clusters until the points assigned to each cluster have a Gaussian distribution [17]. Greg Hamerly et al, show some evidence that his works better than BIC, and that BIC does not penalize the model's complexity strongly enough.

Besides, all the methods, these methods are unsuitable for satellite imagery application because one needs to run these methods for several times. Therefore, we decide to define the appropriate number of clusters using the extended co-occurrence matrix technique to be input parameters of Kmeans algorithm in our research.

#### 6. The Proposed Algorithm

In this paper, we propose a new method for the determination of the appropriate number of clusters, which is based on a novel co-occurrence matrix. While a traditional cooccurrence matrix specifies only the transition within an image in the horizontal and vertical directions, in this work, we embrace the transition of the gray-scale value between the current band and its prior band as well as the current band and its next band into our novel co-occurrence matrix which was called tri-co-occurrence. The proposed method can be used to automatically select a k range in a multispectral satellite image as shown in Table 2.

The proposed technique consists of three main steps. First, band selection [18] is used in the multispectral image to select the best band. Second, the tri-co-occurrence matrix scheme is employed to automatically segment out the object class in the multispectral image. Then, the local maximum technique is used to count the number of clusters, which is represented as the number of clusters, as shown in Fig. 4.

 Table 2
 Algorithm finding the number of clusters.

Algorithm finding the number of clusters					
Input:	D1 D2 D3 : a dataset containing n objects				
Output:	K: the number of desired clusters				
Processing	Processing				
1. Read D1	1. Read D1, D2, D3 images				
2. Transform	D1, D2, D3 to G1, G2, G3 gray scale images				
3. Read G1, G2, G3 gray scale images					
4. Transform	G1, G2, G3 to T co-occurrence matrix				
5. Histogram = T.diagonal()					
6. K = Finding Local Maximum (Histogram)					



Input Multispectral Image

**Fig.4** Block diagram of the proposed automatic determination of the appropriate number of clusters algorithm.

#### 6.1 Original Co-occurrence Matrix

A definition of a co-occurrence matrix [19]–[26] is based on the idea that the neighboring pixels should affect the number of clusters. Hence, the original method defines a cooccurrence matrix by including the transition of the grayscale value between the current pixel and the adjacent pixel into the co-occurrence matrix as illustrated in Fig. 5 and Fig. 6.

Let F be the set of images. Each image has dimensions 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* 

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

where  $\delta = 1$ 

,  $\delta = 0$  otherwise.

where  $F_k$  denotes the *k*th band in the image set, *F* 

Gray level co-occurrence matrix, which is representative of statistical texture analysis, was introduced by Haralick et al. [19]. They defined a gray level co-occurrence matrix, which is the one of fourteen texture measures of gray levels of a texture image and proposed. They expected to represent the characteristics of texture based on it. Gray level co-occurrence matrix texture measurements have been use for segmentation using image texture. Their measure is the uniformity of texture obtained using the second-order moment of the matrix. The contrast of texture obtained from the average of the square of the difference in gray level, and the direction of texture is obtained through the correlation matrix. Then, Umeda [20] and Takiyama and Yano [21] reported that texture images were discriminated with high recognition rates using a gray level co-occurrence matrix directly as a texture feature. Their results from the tests confirmed the effectiveness of the proposed method. Shiranita et al. [22] describes a method for extracting a texture feature from a meat image using the gray level co-occurrence



Fig. 5 Right and bottom of pixel in a co-occurrence matrix.



Fig. 6 Creating a co-occurrence matrix.

matrix and confirmed to be in good result.

Besides, there are many publications that have used the gray level co-occurrence matrix directly as the texture feature for gray image. The results were confirmed the effectiveness of the gray level co-occurrence matrix for segmentation in their images. Although, a number of papers and articles have been published regarding the subject; however, none of the approaches can be employed to incorporate in a multispectral image. Besides, all the methods, to the best of our knowledge, are performed on one band as gray scale image; nevertheless, multispectral images which have more than one band as multiband. We assume that the neighboring bands of multispectral images can enhance the number of clusters of the current band. Therefore, in this work, we estimate the appropriate number of clusters using the tri-cooccurrence matrix calculated from the current band and the neighboring bands to help define the real structure of a multispectral image and the appropriate number of clusters.

#### 6.2 Tri-Co-occurrence Matrix

A new definition of a tri-co-occurrence matrix is based on the idea that the neighboring bands should affect the number of clusters. In this work, we embrace the transition of the gray-scale value between the current band and its prior band as well as the current band and its next band into our novel tri-co-occurrence matrix as shown in Fig. 7 and Fig. 8.

Because image pixel intensities are not independent of each other, a tri-co-occurrence technique is employed. Specifically, we implement a multi-band tri-co-occurrence method which can preserve the structural details of an image. Our definition of a tri-co-occurrence matrix is based on



Fig. 7 Example layers from band number.



**Fig. 8** Comparison between original co-occurrence matrix (a) and tri-co-occurrence matrix (b).



Fig. 9 Examples of satellite images (a) and ground truth data (b).



Fig. 10 An example of blocking of tri-co-occurrence matrix.

the idea that the neighboring bands should affect the current bands. Hence, we define a new definition for a cooccurrence matrix by including the transition of the grayscale value between the current band and its prior band as well as the current band and its next band into our tri-cooccurrence matrix, illustrated in Fig. 8 (b).

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

$$t_{ij} = \sum_{x=1}^{P} \sum_{y=1}^{Q} \delta \begin{cases} (F_k(x,y) = i) \text{ and } (F_k(x,y+1) = j) \text{ or} \\ (F_k(x,y) = i) \text{ and } (F_k(x+1,y) = j) \text{ or} \\ (F_k(x,y) = i) \text{ and } (F_{k-1}(x,y) = j) \text{ or} \\ (F_k(x,y) = i) \text{ and } (F_{k+1}(x,y) = j) \end{cases}$$
(2)

where  $\delta = 1$ 

 $\delta = 0$  otherwise.

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

Our test data are satellite images which consist of different gray levels as shown in Fig. 9 (a) and the ground truth data is as shown in Fig. 9 (b).

From the tri-co-occurrence matrix in Fig. 10, if  $s, 0 \le s \le L - 1$  is a threshold (L = gray level = 256), then s can partition the co-occurrence matrix into four quadrants—namely, A, B, C, and D, as shown in Fig. 10.

The tri-co-occurrence matrix considers the relation be-



**Fig. 11** Example of diagonal of tri-co-occurrence matrix (k = 5): (a) Diagonal matrix plot; (b) Diagonal matrix plot with smooth filter;.

tween four adjacent pixels at a time, called the one reference and the four neighbor pixels, shown in Fig. 8 (b). In the Fig. 10, a relation between the one reference and the neighbor pixels were plotted and shown using Eq. (2). Since two of the quadrants shown in Fig. 10 (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. However, for in the proposed trico-occurrence matrix, we are ignored all quadrants in the calculation, but we considered only diagonal matrix, which is represented and shown by dash line. In the next session, we show how to estimate the number of clusters using the diagonal of the tri-co-occurrence matrix.

6.3 Diagonal Matrix

The proposed method selects the results of the tri-cooccurrence matrix into a diagonal matrix. After diagonal matrix processing, the result of the diagonal matrix is shown in Fig. 11 (a). The diagonal matrix is used to show some clustered pixels and adapted by smoothing and a threshold filter, as shown in Fig. 11 (b). The gray level corresponds to a local maximum which gives the appropriate number of clusters, which is represented as the number of classes in the multispectral image, as shown in Fig. 11 (b).

## 7. Experiment and Results

#### 7.1 Dataset

We will use the seven sets of raw data from different multispectral images [27]–[31] and associated ground truth data. We would like to analyze the data and thus try to determine the appropriate number of clusters. Table 3 shows some multispectral images that contain freely available standard data which was used for our research.

## 7.2 Experimental Results

Our experimental results with the multispectral images are shown in Table 3. The experiments demonstrate the robustness and effectiveness of the proposed algorithm.

The testing was carried out using MATLAB version R2006b, on a Pentium 4 processing chip, with a CPU of 2.0 GHz and used the proposed process on satellite image data labeled Urban I [27], Urban II [28], Landsat I [29], Landsat II [29], IndianPines [30] and SMMS [31].

Dataset	Tri-co-	Co-occurrence	Xmeans	Isodata	Ground
	occurrence	matrix[19]-[26]	[15]	[32]	truth data
	matrix				
Urban I [27]	6	2	8	12	5
Urban II [28]	6	3	9	13	5
Landsat I [29]	4	4	13	15	6
Landsat II [29]	4	4	13	14	6
IndianPines[30]	7	6	16	18	7
SMMS I [31]	6	6	5	9	6
SMMS II [31]	6	5	5	11	6

Table 3Results from experiment.

 Table 4
 Comparative accuracy results for different algorithms compared with ground truth data.

A 1	Results of the accuracy				
Algorithm	Accuracy (%)				
Co-occurrence matrix	62.86				
Tri-co-occurrence matrix	84.86				
Isodata	7.14				
Xmeans	32.29				
Ground truth data	100.00				

This approach was successfully compared with the original co-occurrence matrix, ground truth data and other algorithms—namely, the Xmeans algorithm [15] and iso-data algorithm [32], which is in the top five most frequently used unsupervised classification algorithms applied in remote sensing.

From the experimental results, it was found that clustering using K solved by the tri-co-occurrence statistics techniques provided the nearest number of clusters when compared with the ground truth data. These selected K were used as inputs for K-mean clustering algorithms. Table 3 shows the number of clusters obtained from various datasets for an example experiment. It can be noticed that there are only slight differences in the number of clusters between the ground truth and tri- co-occurrence statistics techniques.

The experimental results show that our proposed process can effectively determine the appropriate number of clusters in satellite image clustering. The performance of the determination a number of clusters algorithm is conventionally measured using accuracy. The percent accuracy rate between the calculated values and the actual values is defined by Eq. (3).

$$% Accuracy rate = 100 - \left(\frac{|K_{calculated} - K_{actual}|}{K_{actual}} \times 100\right)$$
(3)

where  $K_{calculated}$  is the number of classes from our proposed process and  $K_{actual}$  is the number of classes from the ground truth data, respectively.

The algorithm was tested on seven sets of multispectral images. The experimental results were compared with the ground truth images and are evaluated in terms of accuracy in Table 4.

The experiment demonstrates the robustness and effectiveness of the proposed algorithm. The algorithm provides a promising performance in the determination of the number of clusters with an 84.86% accuracy rate.

For the misclassification results on the number of clusters, we considered on the failed results of our proposed method. It is obvious that accuracy rate is not perfect because for an actual multispectral imagery, there is often no sharp boundary between clusters. Thus, there are closed grey levels that strongly affect our method. Our method for automatic determination of the number of clusters depends on the difference of grey level in multispectral imagery. For example, there are two clusters on multispectral imagery, which possess closed grey levels; therefore, it is hard to separate those areas. Those two clusters may be merged into one cluster. So, it affects resulting number of clusters. This is the main cause for those misclassification results generated by our method.

## 8. Discussion

Our definition of a tri-co-occurrence matrix is based on the idea that the neighboring bands should affect the current bands. Because the neighboring bands can use preserved features that separate different object classes on the current band, and both the representation information and class specific information are included.

The strongest point of our proposed is that more information can be added through neighboring bands. The number of clusters depends on the difference of pixel intensities. Therefore, the pixels on the adjacent bands that classify and separate the different objects should be preserved. The pixels of the neighboring bands are accordingly defined as the current bands that not only maintain the major representation of the current band, but also maximally preserve information and more features of the current band that separate different object classes. Because the co-occurrence method does not necessarily guarantee that the resulting transformation will preserve the classification information. The tri-cooccurrence can preserve and help the current band to separate different object classes and the appropriate number of clusters.

Finally, the tri-co-occurrence method, which is proposed in this study, helps achieve a better performance to segment the current band and provide the appropriate number of cluster.

## 9. Conclusion

In this paper, we define a tri-co-occurrence matrix which can preserve the structure within an image and provide the appropriate number of cluster. The approach can be applied to automatically indicate an appropriate number of cluster ranges in satellite images in the pre-clustering steps. The proposed technique consists of three main steps. First, band selection is used in the multispectral imagery. Second, the co-occurrence matrix scheme is employed to automatically segment out the object class in an image. Finally, the local maximum technique is used to count the number of clusters. The experimental results were compared with ground truth images and evaluated in terms of accuracy, with the numerical result of tri-co-occurrence providing an accuracy of 84.86%. The algorithm shows promising performance in determining the appropriate number of clusters in a K-means clustering application by using a tri-co-occurrence statistical technique for multispectral satellite imagery. The outcome of this research will be used in further steps to develop

## Acknowledgments

The authors would like to thank Graduate School, Kasetsart University, Faculty of Science, Kasetsart University and TGIST for financial support.

analysis tools for satellite image mining where the K-mean

method will help to process the satellite imagery.

## References

- N.R. Pal and S.K. Pal, "A review on image segmentation techniques," Pattern Recognit., vol.26, no.9, pp.1277–1294, 1993.
- [2] G.B. Coleman and H.C. Andrews, "Image segmentation by clustering," Proc. IEEE, vol.67, pp.773–785, 1979.
- [3] A.K. Jain and R.C. Dubes, Algorithms for Clustering Data, Prentice Hall, New Jersey, 1988.
- [4] R. Nevatia, In T.Y. Young, and K.S. Fu (Eds.), Handbook of Pattern Recognition and Image Processing, Orlando: Academic Press, 1986.
- [5] J.T. Tou and R.C. Gonzalez, Pattern Recognition Principles, Addison-Wesley, Massachusetts, 1974.
- [6] M.M. Trivedi and J.C. Bezdek, "Low-level segmentation of aerial images with fuzzy clustering," IEEE Trans. Syst. Man. Cybern., vol.SMC-16, no.4, pp.589–598, 1986.
- [7] Y.W. Lim and S.U. Lee, "On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques," Pattern Recognit., vol.23, pp.935–952, 1990.
- [8] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," J. Royal Statistical Society, Series B, 2001.
- [9] M.M.-T. Chiang and B. Mirkin, "Intelligent choice of the number of clusters in K-means clustering: An experimental study with different cluster spreads," J. Classification, vol.27, no.1, pp.3–40, 2010.
- [10] M.M.-T. Chiang and B. Mirkin, "Experiments for the number of clusters in K-means," Proc. Progress in Artificial Intelligence, 2007.
- [11] S. Ray and R.H. Turi, "Determination of number of clusters in Kmeans clustering and application in colour image segmentation," Proc. 4th International Conference on Advances in Pattern Recognition and Digital Techniques (ICAPRDT'99), pp.137–143, 1999.
- [12] T.N. Tran, R. Wehrens, and L.M.C. Buydens, "Clustering multispectral images: A tutorial," Chemom. Intell. Lab. Syst., vol.77, pp.3–17, 2005.
- [13] P. Bradley and U. Fayyad, "Refining initial points for K-means clustering," Proc. 15th International Conf. on Machine Learning, pp.91– 99, 1998.
- [14] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "Knowledge discovery and data mining: Towards a unifying framework," Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining (KDD'96), pp.82–88, Portland, Oregon, Aug. 1996.
- [15] D. Pelleg and A. Moore, "X-means: Extending K-means with efficient estimation of the number of clusters," Proc. 17th International Conf. on Machine Learning, pp.727–734, Morgan Kaufmann, San Francisco, CA, 2000.
- [16] H. Bischof, A. Leonardis, and A. Selb, "MDL principle for robust vector quantisation," Pattern Anal. Appl., vol.2, pp.59–72, 1999.

- [17] G. Hamerly and C. Elkan, "Learning the k in k-means," Proc. Seventeenth Annual Conference on Neural Information Processing Systems (NIPS), pp.281–288, 2003.
- [18] K. Koonsanit and C. Jaruskulchai, "Band selection for hyper spectral image using principal components analysis and maxima-minima functional," Lecture Notes in Artificial Intelligence, LNAI, April 2011.
- [19] R.M. Haralick, M. Shanmugan, and I. Dinstein, "Texture features for image classification," IEEE Trans. Syst., Man. Cybern., vol.3, no.6, pp.610–621, 1973.
- [20] M. Umeda, "Textural image discrimination using a PDP model," IEICE Technical Report, PRU90-123, 1991.
- [21] R. Takiyama and K. Yano, "Textural image discrimination based on the co-occurrence matrix," IEICE Technical Report, PRU91-148, 1992.
- [22] Shiranita, K., Miyajima, T., and Takiyama, R., "Determination of meat quality by texture analysis," Pattern Recognit. Lett., vol.19, pp.1319–1324, 1998.
- [23] N.R. Pal and S.K. Pal, "Entropic thresholding," Signal Process., vol.16, pp.97–108, 1989.
- [24] T. Chanwimaluang and G. Fan, "An efficient algorithm for extraction of anatomical structures in retinal images," Proc. IEEE ICIP, pp.1093–1096, Barcelona, Spain, Sept. 2003.
- [25] Seo, S.T., et al, "Co-occurrence matrix-Based image segmentation," IEICE Trans. Inf. & Syst., vol.E93-D, no.11, pp.3128–3131, Nov. 2010.
- [26] M. Hauta-Kasari, J. Parkkinen, T. Jaaskelainen, and R. Lenz, "Multispectral texture segmentation based on the spectral co-occurrence matrix," Pattern Anal. Appl., pp.275–284, 1999.
- [27] "Urban and terrain hyperspectral data set, "HYDICE sensor imagery," Army Geospatial Center, http://www.agc.army.mil/ Hypercube, accessed April 3, 2011.
- [28] M. Fong and Z. Hu, "HYPERACTIVE: Hyperspectral image analysis toolkit," UCLA Dept of Math, http://www.math.ucla.edu/ ~wittman/hyper/hypermanual.pdf, accessed April 3, 2011.
- [29] LANDSAT Thematic Mapper image, "LANDSAT Thematic mapper image," Matlab, http://www.mathworks.com, accessed April 3, 2011.
- [30] "Indian Pine hyperspectral data set, "HYDICE sensor imagery," http://cobweb.ecn.purdue.edu/~biehl/MultiSpec, accessed April 3, 2011.
- [31] SMMS image data, "Small Multi-Mission Satellite (SMMS) Data" Kasetsart U., http://smms.ee.ku.ac.th, accessed April 3, 2011.
- [32] J.R. Jensen, "Introductory digital image processing," in A Remote Sensing Perspective, pp.197–256, Prentice Hall, New Jersey, 1996.



Kitti 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.