Journal of Advanced Computer Science and Technology Research, Vol.5 No.4, December 2015, 113-122

An Unsupevised Package for Multi-Spectral Image Processing for Remote Data

Muhsin. A. Zaid, Akram. M. Zeki
Department of Information Systems
International Islamic University Malaysia
Kuala Lumpur, Malaysia
muhsin_shamry@yahoo.com, akramzeki@iium.edu.my

ISSN: 2231-8852

ABSTRACT

The ability to match digital images and technique combination in the computer world had revolutionalised the trend. This paper researched on the unsupervised classification of the Multi-Spectral Image. All the two classes under the unsupervised classification were presented and explained. That is the K-Means (KM) and Kohonen Neural Network (KNN). A package for Multi-Spectral Images is designed with the ability to read data, apply Principal Component Analysis (PCA) as a feature extraction, then apply False Colour Composite (FCC) as one of the classification techniques in multi-spectral images. The unsupervised classification method is considered throughout in this research.

Keywords: **Index - K-Means (KM), Kohenen Neural Network (KNN), Principal Component Analysis (PCA), False Colour Composite (FCC).**

1. Introduction

Remote sensing data is received from airplanes or satellites; remote sensing data is a multi-spectral image that means it is an image has several images for the same area (Unsalan & Boyer, 2011). It is an array of three dimensions, as shown in figure (1.1) the (X,Y) dimensions represent the images, where Z dimension represents the band of sensors. Number of these images depends on the number of sensors that is hold by airplane or satellite (Jansse & Gorte, 2004). The most important specification of Remote Sensing satellite is the resolution. This specification has four different levels of resolution (Junling & Tao, 2007). Spatial resolution: refer to the distance between two objects that makes them different from each other. A satellite image with 30 meter spatial resolution means that each pixel in the image represent a ground area with 30 x 30 meter.Spectral resolution: refer to how many bands that sensor has and their size.Radiometric resolution: refer to the width of the radiation. i. e. How many pixel is it 6, or 8 pixels.Temporal resolution: refer to the time that the Remote Sensing satellite takes to scan the same area again.As example of Remote Sensing sensor's Images is the Thematic Mapper sensor. This sensor gives an image with seven bands which they are: first three bands are the visible bands (Blue, Green, and Red), the fourth band is the Near Inferred (NIR) band, Mid Inferred (MIR) bands (5 and 7) with different wavelength intervals, and last band is the Thermal Inferred band (TIR) with different spatial resolution.

Remote Sensing Images is a very rich source for information, each band gives information different from other bands, due to this fact one image of remote sensing with seven bands gives a lot of information about one area. As example, one band gives information about the type of soil, another band gives information about area's vegetation and so on. By adding time factor to the previous fact, because Remote Sensing Satellites spin around earth in orbits and take many images for the same area in different intervals times, the importance and value of Remote Sensing Images rise to a very high level (Weng, 2010).

Unsupervised classification defer from supervised classification is the pixels are assigned to the classes without any intrude or requires any previous knowledge of user of classes. Besides, unsupervised segmentation holds its proximity to feature extraction and clustering (Dey et al., 2010). The figure (1.4) shows Image Classification steps for Unsupervised Classification.

The paper is organised as follows: after abtsract, introduction is the first. Literature review is the next and after it is the proposed enhacement. Then results ad discussion were presented. Lastly the paper ends with the conclusion and references.

2. Related Work

This is entirely based on the statistics of the image data distribution, and is often called clustering. The process is automatically optimized according to cluster statistics without the use of any knowledge- based control (i.e. ground truth). The method is therefore objective and entirely data driven (Liu & Mason, 2013). In more details unsupervised classification can be defined as partitioning of multi-spectral feature space into different regions (called classes or clusters) according to inherent similarity between multi-spectral pixels (pattern vectors). In other words, the distribution of pixels belonging to any homogeneous surface forms cluster of point in the feature space.

For Classical Approach for Unsupervised Classification, several algorithms of multidimensional data clustering exist. These are the simple cluster seeking, the maximum and the K-Means (KM) clustering algorithms. The most important among them is the KM algorithm as it is more accurate and less affected by the order upon which the pattern vectors are exposed to the algorithm (Al-Kuhla, 2014). However, it requires more computation time until reaching the convergence state. The authors of (Maheshwary & Srivastav, 2008) used KM clustering as on of their approach to extract information from Remote Sensing data. KM was the main axis when (Patil & Jondhale, 2010) had palpated there paper "Edge based technique to estimate Number of Clusters in k-means Color Image Segmentation".

The Artificial Neural Network Approach for Unsupervised Classification of Remote Sensing data comprises of many examples of the most common Unsupervised network are Kohonen network which was introduced by (Kohonen, 1984) and the Adaptive Resonance Theory (ARTl and ART2) which were introduced by (Carpenter & Grossberg, 1987) and (Carpenter & Grossberg, 1988). However, Kohonen network is most commonly used for classifying Remote Sensing data. (Junling & Tao, 2007) proposed an energy function as the convergence condition of network, and as such improved Kohonen network unsupervised-learning algorithm. Kohonen network used as one of method of classification under title of “Remote Sensing of Impervious Surfaces in the Urban Areas: Requirements, methods, and trends” by (Weng, 2012).

Table 1 summarizes what had mentioned above about using unsupervised classification.

Table 1: Summary Table

Unsupervised Classification	K-Means	Luke et al, 2005
		Momm et al., 2006
		Aires & Prigent, 2007
		Maheshwary & Srivastav, 2008
		Panchal et al., 2009
		Patil & Jondhale, 2010
	Kohonen	(Junling & Tao, 2007)
		Yuan et al., 2009

3. Proposed Enhancement

Unsupervised classification can be defined as partitioning of multi-spectral feature space into different regions (called classes or clusters) according to inherent similarity between multi-spectral pixels (pattern vectors). In other words, the distribution of pixels belonging to any homogeneous surface forms cluster of point in the feature space.

$$x \in Si(n), \text{ if and only if, } \| x - zi(n) \| \leq \| x - Zj(n) \| \text{ for } i = 1,2,\ldots\, m \text{ and } i \neq j$$

Classical approach

MacQueen introduced K-Means clustering algorithm in 1967, and it is based on Euclidean distance measurement between the tested pattern vectors and the updated mean vectors. The most important among them is the KM algorithm as it is more accurate and less affected by the order upon which the pattern vectors are exposed to the algorithm (Al-Kuhla, 2004). However, it requires more computation time until reaching the convergence state. (Maheshwary & Srivastav, 2008) used KM clustering as on of their approach to extract information from Remote Sensing data. KM was the main axis when (R. Patil & Jondhale, 2010) had palpated there paper “Edge based

technique to estimate Number of Clusters in k-means Color Image Segmentation”. The procedures of its implementation are as follows:
Step One: Assume m is the number of clusters in the available data set and assume any pattern vectors being selected arbitrarily as the centers of these clusters. There should be m pattern vectors, each representing the center of one cluster. These centers are z1,z2, ...zm. Also cluster sets S1, S2,.... Sm, initially (at iteration zero) each set contains only one member who is the cluster center z.
Step Two: Input new pattern (x) to be classified and calculate Euclidean distance between this pattern and the centers of the assumed clusters.

Where n is the iteration number that equals zero at the beginning.
Step Three: Compare the distances and assign the pattern vector x to the nearest cluster using the following decision rule:
Step Four: Calculate the new centers of each set via the following equation:
Where Ni is the number of patterns in the cluster set Si at iteration n.
Step Five: Compare the new centers zi(n+ 1) with the old ones zi(n) and perform the following test:

if (Zi (n + 1) equals Zi (n)) for all clusters then go to step six
else
n = n + 1
go to step Two
Step Six: Stop

Kohonen Network Architecture

Kohonen network consists of one input layer and one output layer. The number of features used in the classification usually determines the number of neurons in the input layer. In the case of Remote Sensing data, this represents the number of wavebands in the satellite sensor. The number of neuron in the output layer is determined arbitrarily, which is the number of output class for the case of Remote Sensing data classification.

Kohonen Network Training

The training of Kohonen network is done as follows:
Step One: Initialize the weights over the connectionists randomly and they must be normalized to unit length of one. Also pre-define the initial value of the dynamic rate (a) and the neighborhood size. The neighborhood size is the number of neurons that are close to the winner and are influenced by weight update. In addition, set the iteration number to one t = 1.
Step Two: Input the pattern vector after being normalized and calculate the activation function using the Equation (2):

$dj = \Sigma_{i=1} w_{ji} xi$ ….. (2)

$Zi(n) = 1/Ni (\Sigma w_{x€Si(n)} X_i$ …. (3)

Where, dj is the output value at neuron j , Wji is the weight vector that connects the input neurons i to the output neuron j and Xi is the input pattern vector. According to (Fu, 2003) neuron with maximum activation value is the winner since
Both W and X are normalised.
Step Three: Update the weight using Equation (4) :
$W_{ji}(t+1)=W_{ji}(t) + \sigma(t).[X_i - W_{ji}(t)]$ .... (4)
Step Four: Compare the weights of the current iteration (t+l) with those of the previous iteration (t) and follow the condition below:

if (Wji (t+1) = Wji (t)) for all output neurons go to step Five
else
t = t +1
$\sigma(t+1) = \sigma(t) - E$
go to step Two

The symbol (E) is the decrement that is set by dividing the initial value of σ (O) over the number of patterns used in the training.
Step Five: Stop

4. Results and Discussion

IRS is stand for Indian Remote Sensing satellite, it is a sensor with four bands matching the first for bands of TM-Sensor, but with spatial resolution of 42 meter and with size (256X512). The Images represent an area in the Red sea; area mainly includes two main mangrove stands, mud flats, algal mats and sargassum. Table (2) shows the Variance/Covariance matrix of the four original bands and the Eigen-Vector and Eigen-Value of the PCA.

Table 2: Variance/Covariance Matrix for IRS-Red Sea Area

	1	2	3	4
1	925.91528 4707568	730.16917 5856370	1232.3851 3772669	1063.80083 405035
2	730.16917 5856370	629.37898 3914701	1137.4892 0748725	992.192052 224911
3	1232.3851 3772669	1137.4892 0748725	2227.6179 4761235	1986.53954 313275
4	1063.8008 3405035	992.19205 2224911	1986.5395 4313275	1858.79691 225797
Eigen Value Matrix				
	1	2	3	4
1	5330.6232 9266156	-4.55E-13	-4.55E-13	-5.68E-14

2	-5.68E-13	264.67398 8042654	-8.53E-14	1.07E-14
3	-3.59E-13	4.97E-14	41.748344 2189105	5.85E-13
4	5.93E-14	1.15E-13	5.78E-13	4.66350356 946366
Eigen Vector Matrix				

What can be observed from table (2) the band 2 is the least variance, and that make FCC made from bands (4, 3, 1). Figure (1) shows the original images of IRS-Red Sea and Figure (2) shows the corresponding PCA images for them. Figure (3) the bar graph show the advantage of using PCA (Data redundancy). Figure (4) shows the three colour composite original images and the PCA images composite.

Visual inspection of Figure (4) shows that the FCC were made from PCA images gave better results than the one were made from the original images. Different features have different colours that make it easy to notice them (Deep water, shallow water, and off-shower area).

To take advantage of above feature extraction, classification was applied to this data, which are MD and KM. Each approach was applied two times one on the original images and the second on the three PCA image. Figure (2) and Figure (3) show the MD and KM for both original and PCA images. Figure (4) shows the six clustering images resulted from using KM.

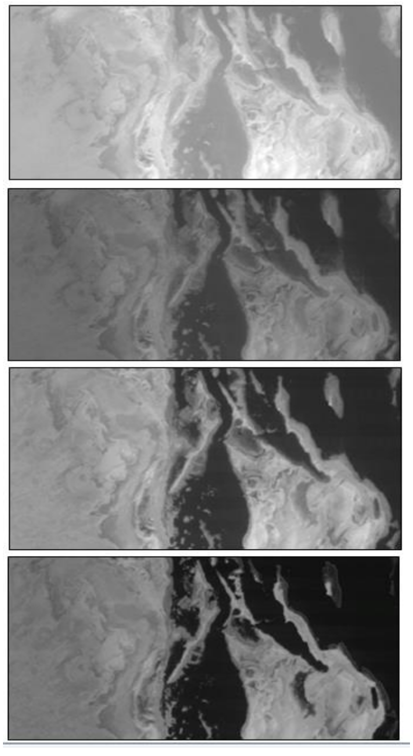

Fig. 1: Four original images of IRS-Red Sea

Fig. 2: Two PCA images for Red Sea area

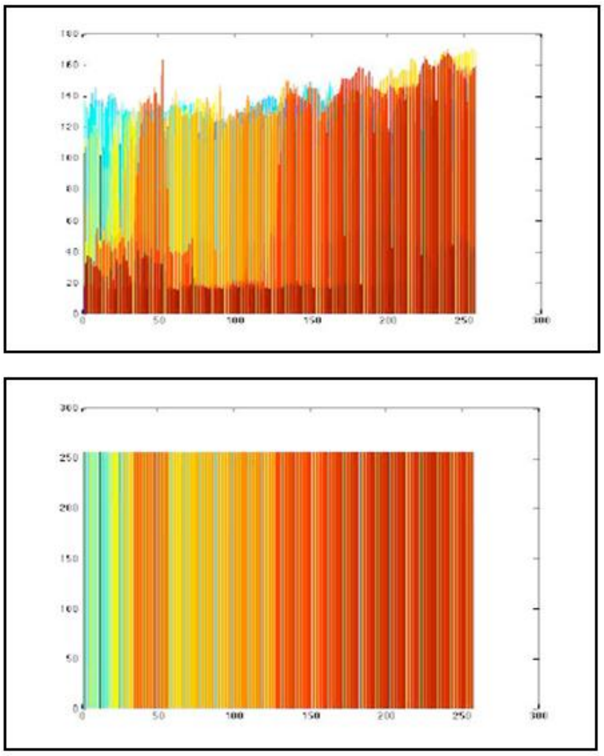

Fig. 3: Two Bar graph images for original and PCA of Band 1 for Red Sear area

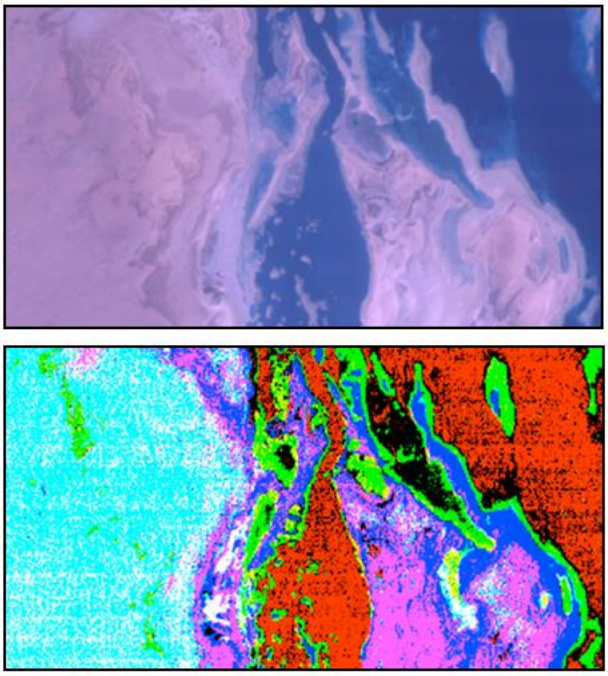

Fig. 4: Two FCC for both Original and PCA images for Red Sea area

5. Conclusion

Before carrying out classification, it is very important to apply Feature Extraction process in order to make classification process faster and more accurate. One of Feature Extraction is used in research, PCA. Unsupervised method (KM) clustering algorithm was applied

In conclusion, results showed that after applying Feature Extraction processes and classification of data, PCA and classification processes illustrates very useful processes in terms of reducing the redundancy of data. Results showed that using FCC technique images from PCA has more colour-interpreted features than the one from original images. It has more colour variants and each feature has a specified colour that makes it easily to be differentiated from other features. The classification process showed that the classification result from manipulating PCA image is better in terms of both accuracy and the speed of classification process than the classification made on the original images.

References

Aires, Filipe, & Prigent, Catherine. (2007). Sampling techniques in high-dimensional spaces for the development of satellite remote sensing database. *Journal of Geophysical Research, 112*(D20), D20301.

Al-Kuhla, L. T. (2004). Integration Between Conventional and Neural Network Approaches for Unsupervised Classification of Remote Sensing Data. (MSC), University of Mosul, Iraq.

Carpenter, Gail A, & Grossberg, Stephen. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer vision, graphics, and image processing, 37*(1), 54-115.

Carpenter, Gail A, & Grossberg, Stephen. (1988). *ART 2: Self-organization of stable category recognition* codes *for analog input patterns*: US Army Research Office.

Jansse, L. L. F., & Gorte, B. G. H. (2004). *Principles of Remote Sensing*: International Institute for Geo-Information Science and Earth Observation.

Junling, WU Ke1 FANG Qiang2 ZHANG, & Tao, WENG. (2007). Application of Kohonen Network in RS Image Classification. *Journal of Geomatics, 2*, 019.

Junling, WU Ke1 FANG Qiang2 ZHANG, & Tao, WENG. (2007). Application of Kohonen Network in RS Image Classification. *Journal of Geomatics, 2*, 019.

Kohonen, Teuvo. (1984). Self-Organization and Associative Memory. *Self-Organization and Associative Memory, 100 figs. XV, 312 pages.. Springer-Verlag Berlin Heidelberg New York., 1*.

Liu, Jian Guo, & Mason, Philippa J. (2009). *Essential Image Processing and GIS for Remote Sensing* (First ed.): Wiley-Blackwell.

Luke, Sean, Sullivan, Keith, Panait, Liviu, & Balan, Gabriel. (2005). *Tunably decentralized algorithms for cooperative target observation.* Paper presented at the Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems.

Maheshwary, Priti, & Srivastav, Namita. (2008). *Retrieving similar image using color moment feature detector and k-means clustering of remote sensing images.* Paper presented at the Computer and Electrical Engineering, 2008. ICCEE 2008. International Conference on.

Maheshwary, Priti, & Srivastav, Namita. (2008). *Retrieving similar image using color moment feature detector and k-means clustering of remote sensing images.* Paper presented at the Computer and Electrical Engineering, 2008. ICCEE 2008. International Conference on.

Momm, HG, Easson, Greg, & Wilkins, Dawn. (2006). *Genetic programming as a preprocessing tool to aid multi-temporal imagery classification.* Paper presented at the Proceedings of the ASPRS 2006 Annual Conference.

Panchal, VK, Kundra, Mr Harish, & Kaur, Jagdeep. (2009). Comparative study of Particle Swarm Optimization based unsupervised clustering techniques. *International Journal of Computer Science and Network Security, 9*(10), 132-140.

Patil, RV, & Jondhale, KC. (2010). *Edge based technique to estimate Number of Clusters in k-means Color Image Segmentation.* Paper presented at the Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on.

Unsalan, Cem, & Boyer, Kim L. (2011). Multispectral *Satellite Image Understanding: From Land Classification to Building and Road Detection*: Spring.

Weng, Qihao. (2010). Remote Sensing and GIS Integration Theories, Methods, and Application: Mcrow-Hill.

Weng, Qihao. (2012). Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sensing of Environment, 117*, 34-49.

Yuan, Hui, Van Der Wiele, Cynthia F, & Khorram, Siamak. (2009). An automated artificial neural network system for land use/land cover classification from Landsat TM imagery. *Remote Sensing, 1*(3), 243-265. 001).