



Oil Spill Hyperspectral Data Analysis: Using Minimum Distance and Binary Encoding Algorithms

Sahar A. El_Rahman^{1,2} & Ali Hussein Saleh Zolait³

¹ Department of Computer Science, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia

² Electrical Department, Faculty of Engineering-Shoubra, Benha University, Cairo, Egypt

³ Department of Information Systems, University of Bahrain, Sakhir, Kingdom of Bahrain

Received: 11 Aug. 2016, Revised: 10 Dec. 2016, Accepted: 21 Sept. 2016, Published: (1 January 2017)

Abstract: Oil spill calamities have increased, threatening maritime ecosystems. This reinforces the need for accurate mapping of oil-spill calamities. The use of hyperspectral classifiers to extract areas of oil spill in a test site was achieved in this work. The paper describes the effects of utilizing a set of hyperspectral image analysis algorithms such as Minimum Distance (MD) and Binary Encoding (BE) algorithms to classify hyperspectral images of oil-spill areas in the Gulf of Mexico using Environment for Visualizing Images software. Hyperspectral image subsetting, region of interest and principal component analysis were performed in the preprocessing stage, which is used to reduce the vast amount of data and eliminate redundant data. The paper provides empirical insights on the classification accuracy of hyperspectral images. A confusion matrix is used to determine the accuracy of a classification by comparing a classification result with ground truth information. The overall accuracies were 94.6399% and 88.4422% for the MD and BE algorithms, respectively. Therefore, the two algorithms are accurate for classifying hyperspectral images of the Gulf of Mexico. However, the MD algorithm is more accurate than the BE algorithm.

Keywords: Hyperspectral Image, Supervised Classification, PCA, Minimum Distance Algorithm, Binary Encoding Algorithm.

1. INTRODUCTION

The concerned authorities and governments have found a way to obtain information concerning oil spill using hyperspectral images, and it is obtained by capturing, analyzing, and studying images from satellites. Multispectral and hyperspectral technologies have been developed for science and research applications. New applications appear by considering multispectral and hyperspectral imagery (Sykas et al., 2011; Vagni, 2007).

Hyperspectral images are used in several applications including food safety and quality, medical sciences, forensics, and agriculture (Headwall photonics company, 2006). Hyperspectral imaging involves gathering and processing data from across the electromagnetic spectrum. The human eye sees visible light in three bands – red, green, and blue (RGB) – however; spectral imaging divides the spectrum into more bands. Hyperspectral images contain a wealth of data, but interpreting them requires an understanding of exactly what properties of the ground materials are being measured, and how they relate to the measurements recorded by the hyperspectral sensor (Smith, 2012). Sensors in hyperspectral imaging

systems provide images with a large number of contiguous spectral channels per pixel. Hence, information about different materials within a pixel can be obtained (Bayliss et al., 1997).

Hyperspectral imaging systems acquire images with an abundance of contiguous wavelengths (usually less than 10 nm). Dozens or hundreds of images are usually obtained; hence, every pixel in a hyperspectral image has its own spectrum over a contiguous wavelength range (Wu and Sun, 2013). Hyperspectral imagery has the potential to extract more accurate and detailed information than that obtained in other cases involving remotely sensed data (Lugo. et al., 2004).

In this method, no prior knowledge of the sample is needed because an entire spectrum is assumed at each iteration and post-processing allows all available information from the dataset to be mined, which is regarded as the main advantage of hyperspectral imagery. Another advantage is the pixel-wise incorporation of a continuous spectral signature of hundreds of wavelengths into a two-dimensional image of the object under inspection. The main disadvantages of this method are the

associated cost and complexity. Faster computers, sensitive detectors, and large data storage spaces are needed for analyzing hyperspectral data (Bauriegel and Herppich, 2014).

2. SUPERVISED CLASSIFIER

Image classification is viewed as an important aspect of remote sensing, image analysis, and pattern recognition (Abbas and Rydh, 2012). Image classification in remote sensing involves assigning pixels or the basic units of an image to classes. It has the potential to gather groups of identical pixels in remotely sensed data, into classes that match the informational categories of user interest by comparing pixels to one another and to those of known identity (Perumal and Bhaskaran, 2010). There are two primary approaches in hyperspectral classification: supervised and unsupervised.

Supervised classification is the process of clustering pixels into classes based on specified training data. Training data are groups of pixels that represent areas for which the information class (land cover, geologic type, etc.) is already known, as shown in Figure 1 (ENVI, 1999).

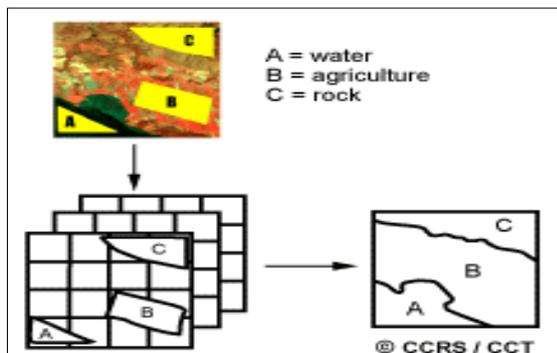


Figure 1: Supervised classification

In supervised classification, a band of “training areas” in the image is designated by an analyst, each of which is a known surface material that symbolizes a desired spectral class. For each training class, the average spectral pattern is computed using a classification algorithm, and then the remaining image cells are assigned to the most similar class. In unsupervised classification, the algorithm derives its own spectral class set from a spot sample of the image cells before performing class assignments (Smith, 2012).

The quality of the training sets is the main factor affecting the quality of supervised classification. All Training sets are created with digitized features.

supervised classifications usually involve a succession of operations that must be followed:

1. Setting the training sites.
2. Extracting signatures.
3. Classifying the image.

Usually, two or three training sites are selected. The greater number of training sites selected, the more effective the process. This process assures both accuracy of classification and correct interpretation of the results. After the training site areas are digitized, statistical characterizations of the information are performed. The specific patterns obtained are called signatures. Finally, the classification methods are applied (Perumal and Bhaskaran, 2010). Unlike the unsupervised methods that do not need to the training data. Unsupervised methods automatically cluster the data of image into various groups according to predefined criteria or a cost function (for example, clustering data based on minimum distance). These groups are then mapped to classes (Tsot and Olsenj, 2005).

There are various methods of supervised classification that have been developed to solve the hyperspectral data classification problem. In this work, two algorithms – the MD and BE algorithms – are applied and their results are compared.

A. MD Algorithm

The MD decision rule (also called spectral distance rule) computes the spectral distance between the mean vector for each signature and the measurement vector for the candidate pixel. It calculates the mean of the spectral values for the training set in each band and for each category, measures the distance from a pixel of unknown identity to each category, assigns the pixel to the category with the shortest distance, and denotes a pixel as “unknown” if the pixel is beyond the distances defined by the analyst. Figure 2 is a schematic of the classification (Al-Ahmadi and Hames, 2007).

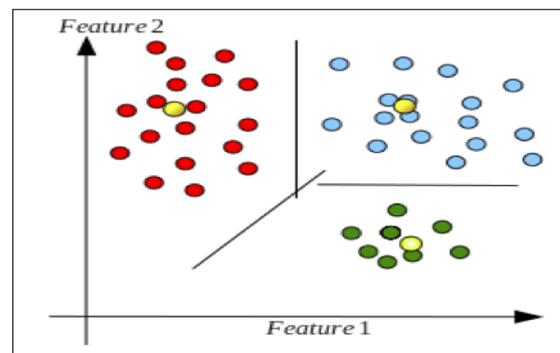


Figure 2: Minimum distance (MD) classification

In the MD algorithm:

- Each class is represented by its mean vector.
- Training is performed using objects (pixels) of a known class.
- The mean of the feature vectors for an object within a class is calculated.
- New objects are classified by determining the closest mean vector (Center for Image Analysis), as shown in Figure 3.

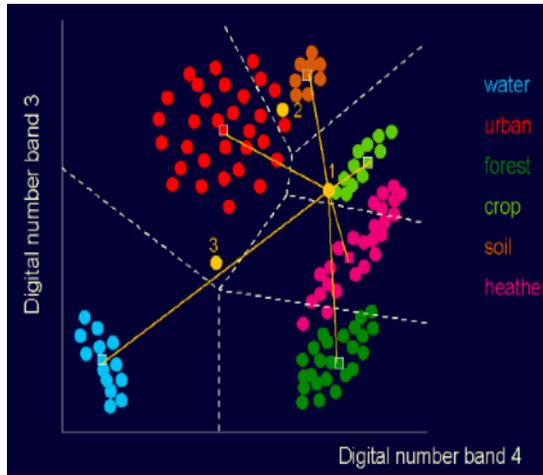


Figure 3: Minimum distance to means classification

As with all classification algorithms, every pixel in the image is analyzed to determine the class assignments. This can be time consuming depending on the file size. Hence, the standard MD classification procedure that's been modified to increase the computational efficiency. Under a normal MD classification, all the pixels are assigned to the nearest spectral class; no pixel is left unassigned.

Some algorithms allow the user to specify a threshold distance from the class mean, where beyond it a pixel will not be assigned and hence, remain unclassified (Al-Ahmadi and Hames, 2007). The Euclidean MD classifier is simple and computationally fast. It is a linear classifier, meaning that the decision surfaces are hyperplanes.

The decision function is:

$$G_i(x) = -r_i^2(x) = -(x - \mu_i)^T(x - \mu_i) \quad (1)$$

Where x is the n -dimensional pixel vector being classified, r is a tunable parameter, T distance-threshold parameter, and μ_i is the n -dimensional mean vector for class i . The function $G_i(x)$ is evaluated for each class, and the pixel is assigned to the class with a maximum value of $G_i(x)$.

The level set function is the signed MD from the pixel to the curve. This distance, by convention, is regarded as negative and positive for pixels outside and inside the contour C , respectively. The level set function ϕ of the closed front C is defined as (Airouche et al., 2009):

$$\phi(x,y) = \pm d((x,y),C) \quad (2)$$

Where $d((x,y),C)$ is the distance from point (x,y) to the contour C and the minus or plus sign is chosen depending on whether point (x,y) is outside or inside interface C , respectively. The interface is represented tacitly as the zero-th level set (or contour) of this scalar function (Airouche et al., 2009):

$$C = \{(x,y) / \phi(x,y) = 0\} \quad (3)$$

B. BE Algorithm

Binary encoding (BE) is a standard technique for classifying hyperspectral images. It reduces the amount of data while conserving as much information as possible (Xie et al., 2011). It reduces the information of a pixel (often represented as 8 bits per channel) to 1 or 2 bits per channel. The basic idea of BE is to compare the albedo of each pixel in every band with a threshold and then assign a code "0" or "1" to it (Nde and James, 2013).

$$S[i] = \begin{cases} 1 & \text{if } (X_i \geq T), \\ 0 & \text{else.} \end{cases} \quad (4)$$

where $S[i]$ is the code of the i -th band, X_i is the attribute of the original spectral vector, and T is the threshold (the mean of the spectral vectors is selected as the threshold).

The BE classification technique encodes the data and endmember spectra into 1s and 0s based on whether a band falls under or above the spectrum mean. It compares each encoded data spectrum with the encoded reference spectrum and produces a classified image. All the pixels are classified to the endmember with the greatest band number that matches unless the user specifies a minimum match threshold in which case some pixels may be unclassified if they do not meet the criteria (ENVI, 2004).

Proposed probability based improved binary encoding (PIBE) method includes two principal steps. The first step is to combine all useful information such as texture, shape, spectra, and height into binary codes, and the second step is to compute the matching probability between image objects and target classes according to the calculated distances between the corresponding binary codes (Xie et al., 2011 & Xie, Huan and Tong, X., 2013).

The encoding rule defines how to convert all useful data into binary codes. The two divisions of the code can be explained as follows:

- i. The spectra: The encoding rule for the spectra follows the traditional BE method.
- ii. Target classes: To be compatible with the input binary codes, the target classes need to be coded in the same way. While in principle, all necessary values can be obtained using training data and universal cognition.

With the BE rule, the target classes and image objects are converted to binary codes; each element is presented by $2L + 5N$, a bit-long binary code. Therefore, the binary codes of hyperspectral image objects are compared with those of the target class using a similar probability-based evaluation measure (Xie et al., 2011& Xie, Huan and Tong, X., 2013).

Binary spectral encoding is advantageous because it is simple, effective, and is a low-computational-load hyperspectral analysis method of classifying and identifying mineral components. Although this method provides a sound execution, it has some disadvantages, i.e., it has a low efficiency in some cases due to the high spatial resolution of modern hyperspectral sensors and because this method mainly operates on the pixels, the efficiency sometimes is low (Mazer et al., 1988).

3. METHODOLOGY

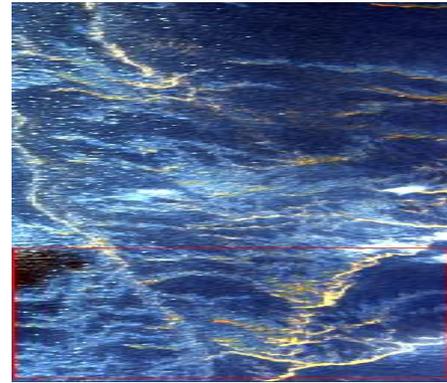
The hyperspectral dataset is first preprocessed. The preprocessing of the hyperspectral dataset includes determination of the region of interest (ROI) and applying PCA. The MD and BE supervised classification algorithms are used in the processing phase.

C. Preprocessing

The region of study is the Gulf of Mexico, USA. Hyperspectral data were acquired on 06 June, 2010, and have 360 bands. The test area was at a major oil spill site. The study region was downloaded from the SpecTIR site (SpecTIR, 2012). Figure 4a shows an image of the study area acquired from Google maps and Figure 4b indicates the dataset that will be analyzed.



(a)



(b)

Figure 4: Study Area

(i) Region Of Interest

An ROI is a part of an image selected either graphically or by methods such as thresholding. ROIs are used in supervised classification, but not unsupervised classification. The ROI selected in this work is shown in Figure 5.

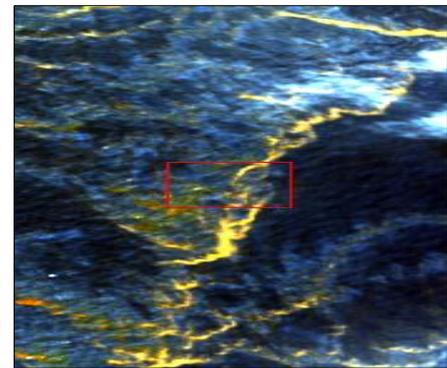


Figure 5 : Region of interest (ROI)

(ii) Principal Component Analysis

Principal component analysis (PCA) is a data-analysis technique used to reduce dimensionality. It reduces the high dimensional vectors to a set of lower dimensional vectors (Koonsanit et al., 2012). Number of bands of original dataset is reduced into a number of new bands that is called the principle components to increase the covariance and decrease redundancy in order to achieve lower dimensionality. Researchers use PCA to determine the best bands for classification, analyze their contents, and evaluate the correctness of classification using PCA images, as shown in Figure 6 and **Error! Reference source not found.** (Rodarmel and Shan, 2002).

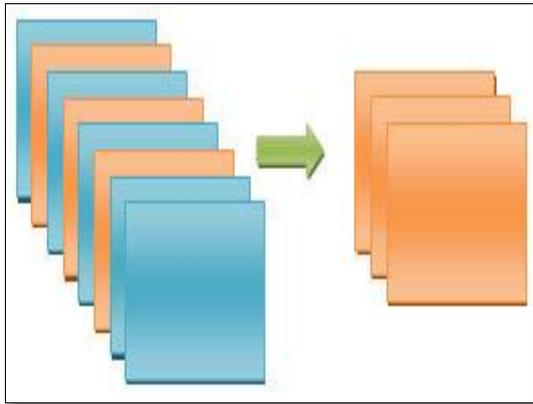


Figure 6: PCA dimension-reduction technique

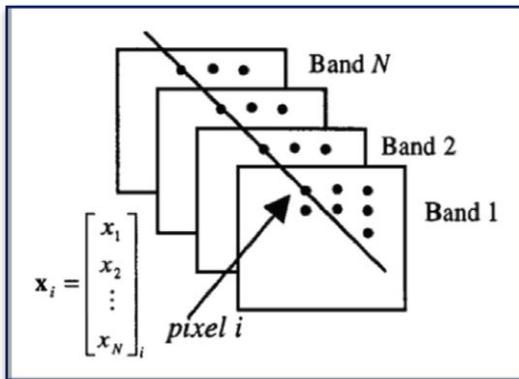


Figure 7 : Pixel Vector in PCA

Figure 8 shows the image obtained after applying PCA to the study area.

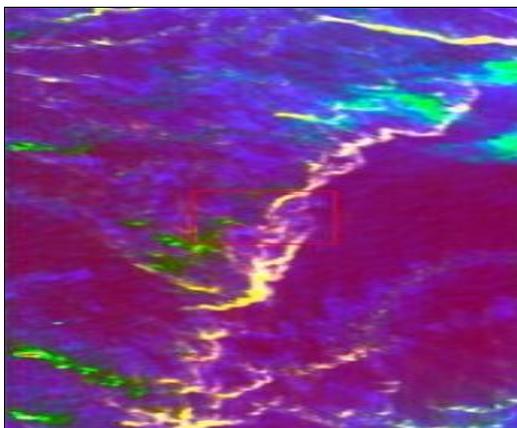


Figure 8: Image of area of oil spill after applying PCA

D. Processing

The overall aim of the work is to acquire an image of the study area. Preprocessing is required for the ROI of the image, and PCA is required to reduce the vast amount of data and eliminate the redundant data. After preprocessing, the main processing is performed by applying the supervised classification algorithms – MD algorithm or BE algorithm. The process is shown in Figure 9.

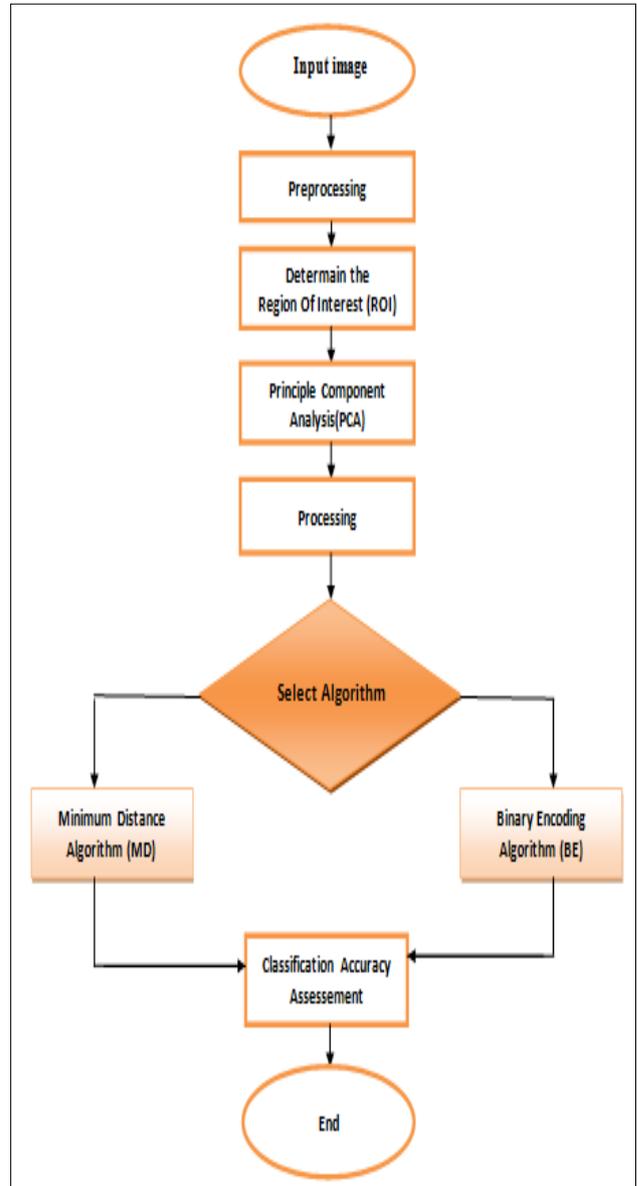
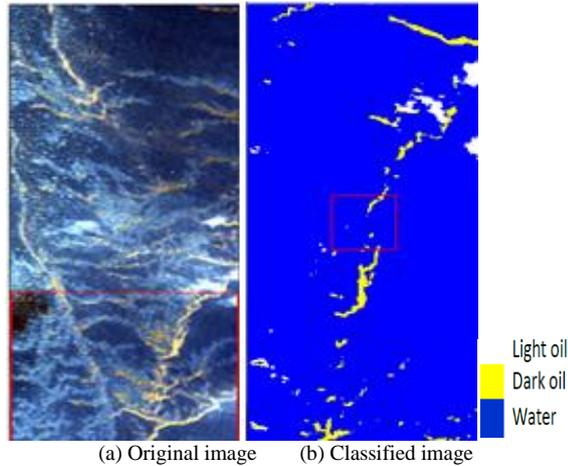


Figure 9: Proposed algorithm architecture

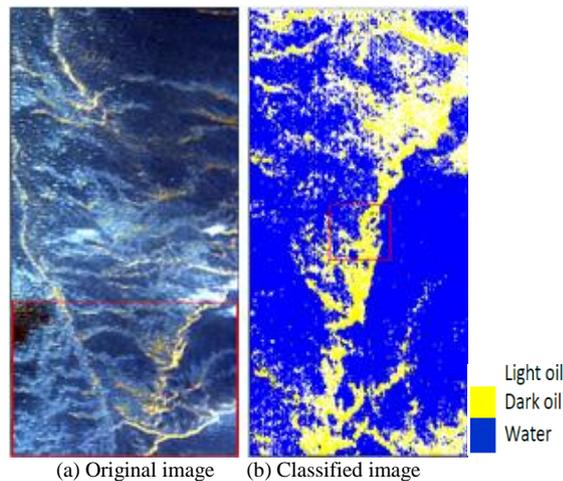
4. RESULT

The classification result of the study dataset obtained by using the MD and BE algorithms in the study area is shown in Figure 10 and Figure 11. A confusion matrix is used to show the accuracy of a classification by comparing a classification result with the ground truth information.



(a) Original image (b) Classified image

Figure 10: MD classification results



(a) Original image (b) Classified image

Figure 11: BE classification results

The confusion matrix is calculated using previously determined ground truth ROIs. Table 1 shows the confusion matrix for the MD algorithm and **Error! Reference source not found.** Table 2 shows the confusion matrix for the BE algorithm. The overall accuracies for the MD and BE algorithms are 94.6399% and 88.4422%, respectively. Hence, the MD algorithm performs better classifications for the study area than the BE algorithm.

Table 1 Confusion matrix classification for MD algorithm.

Overall Accuracy = (565/597) 94.6399%				
Kappa Coefficient = 0.9173				
Ground Truth (Pixels)				
Class	Water	Light oil	Dark oil	Total
Unclassified	0	0	0	0
Water [Blue]	128	5	0	133
Light oil [White]	0	214	0	214
Dark oil [Yellow]	0	27	223	250
Total	128	246	223	597
Ground Truth (Percent)				
Class	Water	Light oil	Dark oil	Total
Unclassified	0.00	0.00	0.00	0.00
Water [Blue]	100.00	2.03	0.00	22.28
Light oil [White]	0.00	86.99	0.00	35.85
Dark oil [Yellow]	0.00	10.98	100.00	41.88
Total	100.00	100.00	100.00	100.00

Table 2 Confusion matrix classification for BE algorithm

Overall Accuracy = (528/597) 88.4422%				
Kappa Coefficient = 0.8214				
Ground Truth (Pixels)				
Class	Water	Light oil	Dark oil	Total
Unclassified	0	0	0	0
Water [Blue]	128	1	0	129
Light oil [White]	0	195	18	213
Dark oil [Yellow]	0	50	205	255
Total	128	246	223	597
Ground Truth (Percent)				
Class	Water	Light oil	Dark oil	Total
Unclassified	0.00	0.00	0.00	0.00
Water [Blue]	100.00	0.41	0.00	21.61
Light oil [White]	0.00	79.27	8.07	35.68
Dark oil [Yellow]	0.00	20.33	91.93	42.71

5. CONCLUSION

This work has achieved the use of MD and binary coding classifiers to extract information about oil-spill areas in a test site, i.e., the Gulf of Mexico in USA. MD and BE algorithms were analyzed to evaluate their ability to classify pixels. PCA was used before the classification process to reduce the dimensionality of datasets. The overall accuracy of the MD and BE algorithms was 94.6399% and 88.4422%, respectively. This indicates that the MD algorithm produces more accurate results than the BE algorithm.



REFERENCES

- Abbas, K. and Rydh, M. (2012), "Satellite Image Classification and Segmentation by Using JSEG Segmentation Algorithm. I.J. Image", *Graphics and Signal Processing IJIGSP*, Vol. 4, No. 10, pp. 48-53.
- Airouche, M.; Bentabet, L.; and Zelmat, M. (2009), "Image Segmentation Using Active Contour Model and Level Set Method Applied to Detect Oil Spills", Proceedings of the World Congress on Engineering. Vol I, WCE 2009.
- Al-Ahmadi, F.S. and Hames, A.S. (2007), "Comparison of Four Classification Methods to Extract Land Use and Land Cover from Raw Satellite Images for Some Remote Arid Areas, Kingdom of Saudi Arabia", *Journal of King Abdulaziz University-Earth Sciences*, Vol. 20 No.1, pp. 167-191.
- Bauriegel, E. and Herppich, W.B. (2014), "Hyperspectral and Chlorophyll Fluorescence Imaging for Early Detection of Plant Diseases, with Special Reference to Fusarium spec. Infections on Wheat", *Journal of Agriculture*, Vol. 4 No. 1, pp. 32-57.
- Bayliss, J.; Gualtieri, J.; and Cromp, R. (1997), "Analyzing Hyperspectral Data With Independent Component Analysis", Proceedings of SPIE AIPR Workshop, volume 9.
- ENVI (2004). "ENVI User's Guide", Research Systems, Inc. All Rights Reserved, Australia, 2004.
- ENVI, (1999) "Laboratory Exercises in Image Processing: Image Classification", available at: <http://www.exelisvis.com/Portals/0/EasyDNNNewsDocuments/Repository/Classification.pdf>
- Headwall Photonics Company; (2006). "Hyperspectral Imaging – Methods, Benefits and Application", available at: <http://www.headwallphotonics.com/applications-old/>
- Koonsanit, K.; Jarusk, C.; and Eium, A. (2012), "Band Selection for Dimension Reduction in Hyper Spectral Image Using Integrated Information Gain and Principal Components Analysis Technique", *International Journal of Machine Learning and Computing*, Vol. 2 No. 3, pp. 251-248.
- Lugo, W.; Cruz, K.; Carvajal, C.L.; and Rivera, W. (2004), "Performance of hyperspectral imaging algorithms using Itanium architecture", *proceeding of the IASTED international conference circuits, signals and systems*.
- Mazer, A.S.; Martin, M.; Lee, M.; and Solomon, J.E. (1988), "Image processing software for imaging spectrometry data analysis", *Remote Sensing of Environment*, Vol. 24 No. 1, pp. 201-210.
- Nde, C.E. and James, E. (2013). *Assessment of Spectral Angle Mapper and Binary Encoding in the Quantification of the Built Environment from Multi-Spectral Landsat Imagery*. New York Science Journal; 6(9), pp. 107-111.
- Perumal, K. and Bhaskaran, R. (2010), "Supervised classification performance of multispectral images", *Journal of computing*, Vol. 2 No. 2, pp. 124-129.
- Rodarmel, C. and Shan, J. (2002), "Principal Component Analysis for Hyperspectral Image Classification", *Surveying and Land Information Systems*. Vol. 62 No. 2, pp.115-123.
- Smith, B. (2012). "Introduction to Hyperspectral Imaging. ©MicroImages, Inc", available at: <http://www.microimages.com> pp.1-24.
- SpecTIR. (2012). "Advanced Hyperspectral and Geospatial Solutions. SpecTIR", available at: <http://www.spectir.com/free-data-samples/>
- Sykas, D.; Karathanassi, V.; Charoula, A. and Polychronis, K. (2011), "Oil Spill Mapping Using Hyperspectral Methods and Techniques", *proceeding of the Tenth International Conference on the Mediterranean Coastal Environment, MEDCOAS*.
- Tsot, B. and Olsenj, R.C. (2005), "Combining spectral and spatial information into hidden Markov models for unsupervised image classification", *International Journal of Remote Sensing*, Vol. 26, No. 10, pp. 2114-2133.
- Vagni, F. (2007). "Survey of Hyperspectral and Multispectral Imaging Technologies. Technical Report", Neuilly-sur-Seine Cedex, France; North Atlantic Treaty Organization; The Research and Technology Organisation (RTO) of NATO.
- Wu, D. and Sun, D.-W. (2013), "Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review — Part I: Fundamentals", *Innovative Food Science and Emerging Technologies*, Vol. 19, pp. 1-14.
- Xie, Huan; Hei., Ch.; Loh., Pet.; Soe., Uwe; Shi, W. (2011). "A New Binary Encoding Algorithm for the Simultaneous Region-based Classification of Hyperspectral Data and Digital Surface Models". E. Photogrammetrie, Fernerkundung, Geoinformation, Schweizerbart'sche, Verlagsbuchhandlung, Stuttgart, Germany, Vol. (1), pp. 17-33.
- Xie, Huan; Tong, X. (2013). "A Probability-based improved binary encoding algorithm for classification of hyperspectral images", *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, Vol. 7, No. 6, pp.2108 -2118.



Sahar Abd El_Rahman was born in Cairo, Egypt, B.Sc. Electronics, Computer Systems & communication, Electrical Engineering Department, Benha University, Shoubra Faculty of Engineering, Cairo-Egypt. M.Sc. in an AI Technique Applied to Machine Aided Translation, Electronic Engineering, Electrical Engineering Department, Benha University, Shoubra Faculty of

Engineering, Cairo-Egypt, May2003. PHD. in Reconstruction of High-Resolution Image from a Set of Low-Resolution Images, Electronic Engineering, Electrical Engineering Department, Benha University, Shoubra Faculty of Engineering, Cairo-Egypt in Jan2008.

She is ASSISTANT PROFESSOR from 2011 till now at Princess Nourah Bint Abdulrahman University/Department of Computer Science, College of Computer and Information System. Also, She is ASSISTANT PROFESSOR from 2008 till now at Electronics & communication, and Computer Systems, Electrical Engineering Department, Faculty of Engineering, Shoubra., Benha University, Cairo, Egypt. She was a LECTURE in the same location from 2003 and INSTRUCTOR in the same location in 1998. Her research interests include computer vision, digital image processing, Signal processing, information security and cloud computing. Dr. Sahar A. El_Rahman is a member of IACSIT since 2013. A member of IAENG since 2011. She is a member of the Egyptian Engineers' Syndicate since 1997.



Ali Hussein Zolait is Assistant Professor of MIS in the department of Information System at college of Information Technology at the University of Bahrain. His research interests are management information systems (MIS), diffusion of innovation, security, and e-commerce application and performance. He was the Stoops Distinguished Assistant Professor of E-

commerce and Management Information Systems at Graduate School of Business, University of Malaya, Malaysia. Dr. Zolait also serves as the Visiting Research Fellow at the University of Malaya. He has developed hundreds of students at all levels- undergraduate, MBA, MM, Executive Development, and Doctoral. Dr. Zolait is a prominent scholar and leader in the field of Innovation Diffusion and Technology Acceptance. He has published many articles on aspects of information security, internet banking, mobile application, supply chain integration, information systems performance in organization, Web maturity evaluation, information systems, performance analysis and instructional technologies, and e-commerce application. He has research published in leading international journals such as Government Information Quarterly, Behaviour & Information Technology, Journal of Systems and Information Technology, and Journal of Financial Services Marketing. He is the Editor-in-Chief of the International Journal of Technology Diffusion (IJTD). He is the IEEE Senior member and currently is the IEEE Secretary for Bahrain section. Dr. Zolait is one of the Founders & Member of Board of Directors in the Society of Excellence & Academic Research, Kingdom of Bahrain. Member of Machine Intelligence Research Labs (MIR Labs), and program chair for the fourth and fifth e-learning conference, kingdom of Bahrain.