Satellite image classification using proposed singular value decomposition method

Bushra Q. Al-Abudi¹, Mohammed S. Mahdi², Noor ZubairKouder¹

¹Astronomy and Space Department, College of Science, Baghdad University

²Computer Science Department, College of Science, Al-Nahrain University

E-mail:noorzubair1984@yahoo.com

Abstract

Keywords

In this work, satellite images for Razaza Lake and the surrounding area district in Karbala province are classified for years 1990,1999 and 2014 using two software programming (MATLAB 7.12 and ERDAS imagine 2014). Proposed unsupervised and supervised method of classification using MATLAB software have been used; these are mean value and Singular Value Decomposition respectively. While unsupervised (K-Means) and supervised (Maximum likelihood Classifier) method are utilized using ERDAS imagine, in order to get most accurate results and then compare these results of each method and calculate the changes that taken place in years 1999 and 2014; comparing with 1990. The results from classification indicated that water and hills are decreased, while vegetation, wet land and barren land are increased for years 1999 and 2014; comparable with 1990. The classification accuracy was done by number of random points chosen on the study area in the field work and geographical data then compared with the classification results, the classification accuracy for the proposed SVD method are 92.5%, 84.5% and 90% for years 1990,1999,2014, respectively, while the classification accuracies for unsupervised classification method based mean value are 92%, 87% and 91% for years 1990,1999,2014 respectivety.

Satellite image, singular value decomposition, classification accuracy.

Article info.

Received: Aug. 2015 Accepted: Oct. 2015 Published: Dec. 2015

تصنيف صورة الأقمار الصناعية باستخدام الطريقة المقترحة تفكك القيمة المفردة بشرى قاسم العبودي¹، محمد صاحب الطائي²، نور زبير خضر¹ اقسم الفلك والفضاء، كلية العلوم، جامعة بغداد ²قسم الحاسبات، كلية العلوم، جامعة النهرين

الخلاصة

تم في هذا العمل تصنيف صور القمر الصناعي لاندسات لبحيرة الرزازة والمنطقة المحيطة بها في محافظة كربلاء للسنوات 1990و 1999و 2014 باستخدام برنامجين (MATLAB 7.12, ERDAS 2014) . تم تطبيق عدة طرق مختلفة من طرق التصنيف ومنها الطريقة المقترحة غير الموجهة (Maan value) والطريقة المقترحة الموجهة (MattLAB 7.12, ERDAS 2014) ، و والطريقة المقترحة الموجهة (K-Means) والطريقة الموجهة (Singular value decomposition) باستخدام والطريقة غير الموجهة (MattLAB 7.12, ERDAS 2014) ، و الطريقة نعير الموجهة (K-Means) والطريقة الموجهة (Singular value decomposition) باستخدام برنامج 2014 والله في دائلة الموجهة (K-Means) والطريقة الموجهة (Jacobie Classifier) باستخدام برنامج 2014) باستخدام برنامج 2014 والله في حالة الموجهة (Romanius) باستخدام برنامج 2014) باستخدام ومناطق التلال في حدثت في سنة 1999و و ذلك للحصول على أدق النتائج ومن ثم مقارنة نتائج هذه الطرق و حساب ومناطق التلال في حدثت في سنة 1999و وسنة 2014 بالمقارنة مع سنة 1990. أشارت النتائج أن المساحات المائية ومناطق التلال في حالة تناقص بينما المساحات الخضراء والأراضي الرطبة والقاحلة في حالة زيادة خلال المنوات 1999 و 2014 بالمقارنة مع سنة 1990. تم معان يا لوطبة والقاحلة في حالة زيادة خلال منطقة الدراسة في العمل الميداني والبيانات الجغرافية ومن ثم مقارنتها مع نتيجة التصنيف، دقة التصنيف ملطيقة الدراسة في العمل الميداني والبيانات الجغرافية ومن ثم مقارنتها مع نتيجة التصنيف، دقة التصنيف ملطيقة الدراسة في العمل الميداني والبيانات الجغرافية ومن ثم مقارنتها مع نتيجة التصنيف، دقة التصنيف ملطيقة الدراسة في الموجهة التي اقترحت (SVD) كانت 2.95% ،84.5% ،90% (Mean value) 2014 ، 1999، 1990 ، 19% ، 19%

Introduction

Remotely sensed imagery can be used in a number of applications, A principal application of remotely sensed data is to create a classification map of the identifiable or meaningful features or classes of land cover types in a scene[1]. Classification is one of the data mining methods which are used to classify the object into predefined group. It is the most frequently used decision making tasks of human activity. A classification problem occurs when an object needs to be assigned into a predefined group or class based on a number of observed attributes related to that object. The classification also plays very important role in the remote sensing and satellite image classification[2]. Image classification is a complex process that may be affected by many factors. Huge number of classification techniques can be found in the literature; mostly they have been categorized as either supervised or unsupervised methods. The supervised techniques are often required sort of prior knowledge in selecting correct region of interest "ROI", inadequate selection of "ROI" or the number of correct existed regions, often, yields an inadequate classification results, while the unsupervised methods need to identify the correct number of regions existed in the processed image. In this paper, we classified Landsat satellite

images for Razaza lake and the area surrounding district in Karbala province for years 1990,1999 and 2014 using two software programming (MATLAB 7.12 and ERDAS imagine 2014. Unsupervised and supervised classification method have been used to classify satellite image; these were unsupervised method (mean value), supervised methods (Singular Value Decomposition) using MATLAB software, and unsupervised method (K-Means), supervised method (Maximum Likelihood Classifier) using ERDAS imagine.

The study area

The study area chosen is Razaza lake and the surrounding area. The total area is 7101.4 km². It is bounded between the longitudes $42^{\circ} 83'$ to 43° 62' E and Latitude 32° 03' to 33° 42' N. it is lying 15 km west of Karbala. Razaza linked from the north by Habbanivah Lake by Nazim Al-Warawr Canal, and surrounded by the other three sides of the land of desert interspersed with some hills. The satellite image for the study area is capture from landsat-5, lansat-7 and landsat-8 for years 1990, 1999 and 2014 respectivety. The bands which are chosen (2, 3 and 4) with 30m spatial resolution. Fig.1 shows the original map of Iraq. While Fig. 2 shows the study area for period 1990,1999 and 2014.



Fig.1: Location of the study area on the map of Iraq.



Study area in 1990



Study area in 1999



Study area in 2014

Fig.2: The study area (Razaza lake and the area surrounding) for years 1990, 1999 and 2014.

The proposed classification methods

Two classification methods, were performed; unsupervised and supervised on the study area for the years 1990,1999 and 2014 using MATLAB programs as follows:

1. Unsupervised classificationusing mean value

Unsupervised classification is based on value of mean. The mechanisme of this classification

can be illustrated in the following steps:

1. Read the Satellite image.

2. Partitioned the Satellite image into fixed size blocks. In this work we used block size (3*3).

3. For each block compute mean value, and then determine the maximum and minimum value of mean for these blocks.

4. To get classes subtract the maximum value from the minimum then divided on number of class.

5. Add the results of step (4) to minimum value to get first class then add it to first class to get second class, and so on

6. Finally classification decision is made according to comparing the mean value for each block with mean values of classes.the block assigned to the class when the mean value of block lying in class.

2. Supervised classification using singular value decomposition

The need to minimize the amount of digital information stored and retrieved is an ever growing concern in the modern world, singular value decomposition (SVD) is an effective tool forminimizing data storage and data transfer. SVD is one technique of semantic indexing (SI). Singular value decomposition (SVD) is proposed to perform supervised classification method, it is consists of two phases: the training and classification; The training phase is responsible on storing the classes in the database, while the task of classification phase is to compute the similarity measure between the SVD of the target image and SVD of the classes found in the database.

The following steps illustrated the mechanism of this proposed classification:

1. Read the satellite image.

2. Partitioned the Satellite image into fixed size blocks. In this work we used block size (3*3). Each block of the satellite image represent the query vectors (q).

3. The satellite image that selecte consist of three band (green, red and near infrared) and five class (water, vegetation,Wet land, hills and Barren land). Select five blocks each block represents class, then putting it in document matrix A (m*n):

where m represents number of class element and n represents number of class.

4. Compute transpose of document matrix (A) then calculate value of (A^TA).

5. Compute (K) that represent largest eigenvalues and eigenvectors.

6. Compute the coordinates of documents in the k-dimensional orthogonalspace by the SVD-free approach [3]:

 $[V,S_2] = eigs (A^TA, k)$ (1)

4.Compute the coordinate of query vector q [3]:

 $qc=(((q^{T*}A)*V)*inv(S))*inv(S)$ (2) where S represents square root of (S₂) 5.Compute the similarity coefficients between the query vector and documents [3]:

 $z = qc*v(i,:)^{T/} (norm(qc)*norm(v(i,:)))$ (3) sim(i)=1-acos(z)(4)
where; sim represents the similarity.

6. Finally classification: the classification decision is made according to similarity measures of each block, the block assigned to the class that gives highest similarity

measure. Fig.3 shows the block diagram of mean value and singular value decomposition (SVD) classification methods for the satellite images using MATLAB software.



Fig.3: Block diagram of the proposed classification method using MATLAB software.

The results from applying mean value classification method can be shown in Figs.4-6, for years 1990,1999 and 2014, respectivety, Table 1 presents the change in the years 1999, 2014; comparing with 1990. The results from applying SVD classification method can be shown in Figs.7-9 for years 1990,1999, 2014, respectivety, Table 2presents the change in the years 1999, 2014; comparing with 1990. The results indicated that five classes found with comparison with the original image.These classes represent five major features in the study area (water, vegetation, wet land, hills and Barren land.



Fig.4: Unsupervised classification using mean value for Landsat-5 (1990).



Fig.5: Unsupervised classification using mean value for Landsat-7 (1999).



Fig.6: Unsupervised classification using mean value for Landsat-8 (2014).

Classes	Area (No. pixels) 1990	Area (No. pixels) 1999	Area (No. pixels) 2014	Changes between 1990 and 1999	Changes between 1990 and 2014
Water	110934	95841	44991	-15093	-65943
vi uter	20.5%	17.7%	8.3%	-13%	-59.4%
Vegetation	13743	9206	12600	-4537	1143
vegetation	2.5%	1.7%	2.3%	-33%	8.3%
Watland	25308	32922	17892	7614	-7416
Wet land	4.6%	6%	3.3%	30%	-29.3%
Hills	201708	181791	195327	-19917	-6381
пшя	37.3%	33.6%	36.1%	-9.8%	-3.1%
Downon land	188307	220239	269190	31932	80883
Barren land	34.8%	40.7%	49.8%	16.9%	42.9%

Table 1: The change in the years 1999, 2014 comparing with 1990 using mean valueclassification method.



Fig.7: Supervised classification using (SVD) for Landsat-8 (2014).



Fig.8: Supervised classification using (SVD) for Landsat-7 (1999).



Fig.9: Supervised classification using (SVD) for Landsat-8 (2014).

Table 2: The change in the years 1999, 2014 comparing with 1990 using SVD classification
method.

		met	hod.		
Classes	Area	Area	Area	Changes	Changes
	(No. pixels)	(No. pixels)	(No. pixels)	between 1990	between
	1990	1999	2014	and 1999	1990 and 2014
Water	114408	98370	44559	-16038	-69849
	21.1%	18.2%	8.2%	-14%	-61%
Vegetation	10269	11736	10404	1467	135
	1.9%	2.1%	1.9%	14.2%	1.3%
Wet land	25308	32607	25758	7299	450
	4.6%	6%	4.7%	28.8%	1.7%
Hills	185418	182538	166365	-2880	-19053
	34.3%	33.8%	30.8%	-1.5%	-10.2%
Barren land	204597	214749	292914	2439	8831
	37.8%	39.7%	54.2%	1.1%	4.3%

Satellite image classification using *ERDAS* imagine software

Unsupervised and supervised classification method are applied; these

are K-Means and Maximum likelihood Classifier. Fig.10 shows the block diagram of this software.



Fig.10: Block diagram of the classification methods using ERDAS imagine.

The results from applying unsupervised classification method (Kmean) are shown in Figs.11-13 for years 1990,1999, and 2014, respectivety and Table 3 presents the change in the years, 1999, 2014; comparing with 1990. The results from applying supervised classification method (Maximum likelihood) are shown in figures (14-16) for years

1990, 1999, 2014, respectivety and Table 4 presents the change in the years 1999, 2014; comparing with 1990. The results indicated that five classes found comparison with the original image. These classes represent five major features in the study area (water, vegetation, wet land, hills and Barren land).



Fig.11: Unsupervised classification using K-Means classification method for Landsat-5 (1990).



Fig.12: Unsupervised classification using K-Means classification method for Landsat-7 (1999).



Fig.13: Unsupervised classification using K-Means classification method for Landsat-8 (2014).

Table 3: T	he change in	•	99, 2014 comp Fication metho	aring with 1990 usi d.	ng K-Means
				Changes	Changes

Classes	Area(ha) 1990	Area(ha) 1999	Area(ha) 2014	Changes between 1990 and 1999	Changes between 1990 and 2014
Water	157711	130604	54343	-27107	-103368
water	22.2%	18.3%	7.6%	-17.1%	-65.5%
Vegetation	12538	34786	155130	22248	142592
vegetation	1.7%	4.8%	21.8%	177.4%	1137.2%
Wet land	133292	146667	162901	13375	29609
wet fanu	18.7%	20.6%	22.9%	10%	22.2%
Hills	222412	244148	214802	21736	-7610
пшя	31.3%	34.3%	30.2%	9.7%	-3.4%
Downon land	184187	153935	122964	-30252	-61223
Barren land	25.9%	21.6%	17.3	-16.4%	-33.2%



Fig.14: Supervised classification using Maximum Likelihood classifier method for Landsat-5(1990).



Fig.15: Supervised classification using Maximum Likelihood classifier method for Landsat-7 (1999).



Fig.16: Supervised classification using Maximum Likelihood classifier method for Landsat-8 (2014).

			u ciussijier me		
Classes	Area(ha) 1990	Area(ha) 1999	Area(ha) 2014	Changes between 1990 and 1999	Changes between 1990 and 2014
Water	156612	131552	50503	-25060	-106109
water	22%	18.5%	7.1%	-16%	-67.7%
Vegetation	13997	20437	16565	6440	2568
Vegetation	1.9%	2.8%	2.3%	46%	18.3%
Wet land	21881	39433	30668	17552	8787
wet fanu	3%	5.5%	4.3%	80.2%	40%
Hills	276879	272007	257353	-4872	-19526
111115	38.9%	38.3%	36.2%	-1.7%	-7%
Barren land	240771	246711	355321	5940	114550
Darren lanu	33.9%	34.7%	50%	2.4%	47.5%

 Table 4: The change in the years 1999, 2014 comparing with 1990 using Maximum

 Likelihood classifier method.

Accuracy assessment

Classification accuracy assessment is a general term for comparing the classification to geographical data that are assumed to be true to determine the accuracy of the classification process. Usually, the assumed true data are derived from ground truth. Ground truth or field survey is done in order to observe and collect information about the actual condition on the ground at a test site and determine the relationship between remotely sensed data and observed. object to be It is recommended to have a ground truth at the same time of data acquisition, or at time least within the that the environmental condition does not change, it is usually not practical to ground truth or otherwise test every pixel of a classified image, therefore a set of reference pixel is usually used, reference pixels are points on the classified image for which actual data will be known. The reference pixel is randomly selected[4]. The most common tool used for the classification accuracy assessment is in terms of a confusion (or error) matrix. A confusion matrix is a square array of dimension $n \times n$, where n is the number of classes. The matrix shows the relationship between two samples of measurements taken from the area that has been classified. The first set represents test data that have been collected via field observation. inspection of agricultural records, air photo interpretation, or other similar means. The second sample is composed of the labels of the pixels. allocated by the classifier. that correspond to the test data points.

The columns in a confusion matrix represent test data, while rows represent the labels assigned by the classifier.

The main diagonal entries of the confusion matrix represent the number of pixels that are given the same identification by the test data and the classifier, and these are the number of pixels that are considered to be correctly classified. Several indices of classification accuracy can be derived from the confusion matrix. The overall accuracy is obtained by dividing the sum of the main diagonal entries of the confusion matrix by the total number of samples. In order to assess the accuracy of each information class separately, the concepts of producer's accuracy and user's accuracy can be used. For each information class i in a confusion matrix. the producer's accuracy is calculated by dividing the entry (i, i) by the sum of column i, while the user's accuracy is obtained by dividing the entry (i, i) by the sum of row i. Thus, the producer's accuracy tells us the proportion of pixels in the test data set that are correctly recognized by the classifier. The user's accuracy measures the proportion of pixels identified by the classifier as belonging to class i that agree with the test data[5,6]. For example, Fig. 17 shows an error matrix with producers, and overall accuracy users calculations for a simple 6-class. In this work, the classification accuracy was done for years 1990, 1999 and 2014. Number of random points which chosen on the study area in field work compared were then with the classification results.



Fig.17: Example error matrix with producers, users, and overall accuracy calculations [6].

Tables 5-7 present error matrix with producers, users, and overall accuracy calculations for unsupervised classification using mean value for years 1990, 1999 and 2014, respectively. While Tables8-10 present error matrix with producers, users, and overall accuracy calculations for supervised classification using SVD for years 1990, 1999 and 2014, respectively.

 Table 5: Error matrix with producers, users, and overall accuracy calculations for unsupervised classification using mean value for Landsat-5 (1990).

Classes	water	Vegetation	Wet land	Hills	Barren land	Row total	user's accuracy
Water	47	2	0	0	0	49	95.9%
Vegetation	1	31	3	0	0	34	91.1%
Wet land	0	0	27	0	2	29	93.1%
Hills	0	3	0	39	4	46	84.7%
Barren land	0	0	0	2	40	42	95.2%
Column total	47	36	30	41	46	200	
Producer's Accuracy	97.9%	86.1%	90%	95.1%	86.9%		
Total accu	iracy		929	%			

 Table 6: Error matrix with producers, users, and overall accuracy calculations for unsupervised classification using mean value for Landsat-7 (1999).

	F	i etassijteatton					
Classes	water	vegetation	Wet land	Hills	Barren land	Row total	user's Accuracy
Water	39	2	0	0	0	41	95.1%
Vegetation	2	32	1	0	0	35	91.4%
Wet land	0	0	23	1	3	27	85.1%
Hills	0	0	4	43	9	56	76.7%
Barren land	0	0	0	4	41	45	91.1%
Column total	41	34	28	48	53	200	
Producer's Accuracy	95.1%	94.1%	82.1%	89.5%	77.3%		-
Total accu	Total accuracy 87%						

Classes	water	Vegetation	Wet land	Hills	Barren land	Row total	user's accuracy
Water	24	3	0	0	0	27	88.8%
Vegetation	0	32	2	0	0	34	94.1%
Wet land	0	0	27	1	2	30	90%
Hills	0	1	0	46	6	53	88.6%
Barren land	0	0	0	3	53	56	94.6%
Column total	24	36	29	50	60	200	
Producer's Accuracy	100%	88.8%	93.1%	92%	88.3%		
Total accu	racy		919	%			

Table 7: Error matrix with producers, users, and overall accuracy calculations for unsupervised classification using mean value for Landsat-8 (2014).

Table 8: Error matrix with producers, users, and overall accuracy calculations for supervisedclassification using SVD for Landsat-5 (1990).

Classes	Water	vegetation	Wet land	Hills	Barren land	Row total	user's accuracy
Water	39	2	0	0	0	41	95.1%
Vegetation	1	28	2	0	0	31	90.3%
Wet land	0	0	27	0	2	29	93.1%
Hills	0	0	4	48	2	54	87%
Barren land	0	0	0	2	43	45	95.5%
Column total	40	30	33	50	48	200	
Producer's Accuracy	97.5%	93.3%	81.8%	96%	89.5%		
Total accu	racy		92.59	%			

 Table 9: Error matrix with producers, users, and overall accuracy calculations for supervised classification using SVD for Landsat-7 (1999).

Classes	water	vegetation	Wet land	Hills	Barren land	Row total	user's accuracy
Water	31	3	0	0	0	34	91.1%
Vegetation	1	27	4	0	0	32	84.3%
Wet land	1	1	24	1	0	27	88.8%
Hills	0	0	5	49	7	61	80.3%
Barren land	0	2	2	4	38	46	82.6%
Column total	33	33	35	54	45	200	
Producer's Accuracy	93.9%	81.8%	68.5%	90.7%	84.4%		
Total accu	racy		84.5%	, D			

Classes	water	vegetation	Wet land	Hills	Barren land	Row total	user's accuracy
Water	30	2	1	0	0	33	90.9%
Vegetation	4	29	2	0	0	35	82.8%
Wet land	2	2	21	2	0	27	77.7%
Hills	0	0	2	48	1	51	94.1%
Barren land	0	0	0	2	52	54	96.2%
Column total	36	33	26	52	53	200	
Producer's Accuracy	83.3%	87.8%	80.7%	92.3%	98.1%		
Total accu	racy						

Table 10: Error matrix with producers, users, and overall accuracy calculations for supervised classification using SVD for Landsat-8 (2014).

Conclusions

The basic idea of our work is to classify Landsat satellite images at different times for the study area using several methods and then compare the results of each method. From the results obtained the following points are chosen for the present conclusions: The results from applying SVD 1. classification method show that the water and hills in decrease, water decreased about 14% in 1999, 61% in 2014 and hills decreased about 1.5% in 1999, 10.2% in 2014, while vegetation, wet land and barren land in increase, vegetation increase about 14.2 % in 1999, 1.3% in 2014, while wet land increased about 28.8 % in 1999, 1.7 % in 2014 and barren land increased about 1.1% in 1999,4.3% in 2014; comparable with 1990.

2. Maximum likelihood Classification method show that the water and hills in decrease, water decreased about 16% in 1999, 67.6% in 2014 and hills decreased about 1.7% in 1999, 7% in 2014. while vegetation, wet land and barren land in increase, vegetation increase about 46% in 1999, 18.3% in 2014, while wet land increased about 80.2% in 1999, 40% in 2014 and barren land increased about 2.4% in 1999, 47.5% in 2014; comparable with 1990.

3. The classification accuracy for the proposed SVD method is 92.5%, 84.5% and 90% for years 1990,1999,2014, respectivety, This proved superiority the method based on SVD for classifying satellite images.

4. Despite the use of many method of unsupervised classification, results of the proposed method based on mean value proved it's superiority because the classification accuracy are 92%, 87% 91% and for vears 1990,1999,2014 respectivety, it is so nearest from supervised classification results. While the fluctuations from using K-mean method make them not trusty, this is because of the many thresholds used in their performance maximum number iterations, (i.e. minimum number of points in the maximum class standard class. deviation, maximum class distance, maximum number of merge pairs, maximum std from mean, and maximum distance error).

5. The results from classification indicated that water and hills in decrease, while vegetation, wet land and barren land in increase for years 1999 and 2014; comparable with 1990. The cause of contraction water area due to scarcity of rainfall and desertification, conversely the vegetation and wet land increase. While the hills in decreasing conversely increase in the barren land this due to erosion factors.

References

[1] K.Perumal, and R.Bhaskaran, Journal of Computing, 2 (2010) 124-129.

[2] U.Anand, K. S.Santosh, G.S.Vipin, International Journal of Application or Innovation in Engineering and Management, 3, October (2014) 311-317.

[3]P. Pavel, M. Libor, S.Vaclav, " Iris Recognition Using the SVD-Free Latent Semantic Indexing", Fifth International Workshop on Multimedia Data Mining, P. 67-70, USA, 2004.

[4] S. A.Safaa," Studying Galal Badra Dam Basin Area using R.S and GIS Tools", M.Sc. Thesis Submitted to Physics Department, College of Science, Baghdad University, Iraq, 2014.

[5] M. Paul Mather, "Classification Methods For Remote Sensing Data", Second Edition, CRC Press, London, New York, 2009.

[6] R. G.Congalton, and K. Green, "Assessing the Accuracy of Remotely Sensed Data: Principles and Practices", Lewis Publishers, Boca Raton, FL,1999.