Effect of classification procedures on the accuracy of land-cover mapping

Hafez Abbas Afify
Transportation Eng. Dept., Faculty of Eng., Tanta University, Tanta, Egypt

A series of experiments has been undertaken to evaluate the influence of some of the most common procedures used at different stages of multispectral classification process on the accuracy of land-cover mapping. The original three bands of SPOT XS imagery, and two transformed principal component images were used. Single pixel and block sampling procedures of training data were evaluated. Three supervised classification techniques were applied; namely, Minimum Euclidean Distance, Minimum Mahalanobis Distance, and Maximum Likelihood. The Kappa coefficient of agreement and its variance were calculated for each classification results. Significance test was applied to compare the accuracy of two classifications carried out using different procedures. As a result, the factors and procedures involved were ranked according to their significant influence on the accuracy of the classified data.

في هذا البحث تم اجراء مجموعة من التجارب لتصنيف المعالم الارضية المختلفة من محاصيل و مناطق سكنية وتم اعتبار الاساليب المتعددة المستخدمة في كل خطوة من خطوات التصنيف وذلك بهدف الوصول الى مدى تأثير دقة التصنيف نتيجة تطبيق اسلوب معين دون غيره. وقد اجريت التجارب على صورة القمر الصناعي (سبوت) قبل وبعد معالجتها، وبتطبيق ثلاث من الطرق الشائعة الاستخدام في التصنيف. كما تمت مقارنة النتائج بأكثر من طريقة حيث امكن تحديد الاساليب المناسبة في كل من خطوات التصنيف والتي لها تأثير جوهري على دقة التصنيف لمنطقة الدراسة، وقد تم ايضا ترتيب هذة الاساليب تبعا لأهميتها في زيادة دقة تصنيف المعالم الارضية.

Keywords: Multispectral classification, Land-cover mapping, Principal component analysis, Kappa coefficient of agreement

1. Introduction

Remote sensing data acquired from high spatial sensors such as landsat Thematic Mapper (TM), and SPOT High Resolution Visible (HRV) has been utilized for mapping land-cover categories. Both supervised and unsupervised techniques can be applied to multispectral images to produce thematic maps. Among the most frequently used procedures of supervised classification are the Minimum Euclidean Distance, the Minimum Mahalanobis Distance, and the Maximum Likelihood classifiers [1]. They perform classification based only on the spectral signatures of specific pixels and do not consider the spatial information of those pixels nor the spectral information that might be obtained from surrounding pixels [2]. However, the classification process involves major stages, and each can be accomplished by various procedures and using different factors, which eventually influence the accuracy of the classified data. These major stages and corresponding procedures are:

1- Data preparation, this step includes several image-processing techniques that can be applied to the raw data such as radiometric and geometric corrections, creation of principal component images, and filtration using high and low pass filters [3].

2- Selection of training data, the purpose of this stage is to locate the training pixels, which represent the spectral characteristics of each land-cover class. Block and single pixel sampling strategies are common procedures used to collect the training samples of each class.

3. Classification algorithm, which refers to the decision rule used to assign each pixel to a certain class. The most-commonly-used procedures of supervised classification are the Minimum Euclidean Distance (MED), the

Alexandria Engineering Journal. Vol. 41 (2002), No. 3, 571-577

Minimum Mahalanobis Distance (MMD), and the Maximum Likelihood (ML) classifiers.

4- *Accuracy assessment*, including many different indices to evaluate the accuracy of the classification results such as the overall and average accuracy, the percentage of omission and commission errors, the Kappa coefficient of agreement, and the Z-value of the significance test [4].

The purpose of this paper is to evaluate various procedures used at different classification stages, and to determine the extent to which each procedure could significantly affect the accuracy of multispectral classification for land-cover mapping.

The experiments were carried out using the three original bands of SPOT XS image, and the first two transformed principal component images. The block and single pixel sampling procedures of training data were considered, and three commonly used classifiers of supervised classification were applied. The results of each classification carried out under various procedure and factor combinations were determined and compared to each other by different means of accuracy assessment.

2. The study area

The study site covers the town of Almahalla Al-Kubra, Al-Gharbiyyah, Egypt and the cultivated area to the east of it. This area is a part of the Nile Delta at the west side of the River Nile Dumyat branch. It provides a good site for the analysis as three cultivated land-cover types are involved in addition to the urban areas located in the north side of Bahr Shibin which flows through the study area as shown in fig. 1. The area is relatively flat with a mean ground elevation of 8 meters above the mean sea level.

The SPOT multispectral image used in this study was acquired on August 11, 1995. The study area constitutes a subscene of this image covering 400 x 400 pixels.

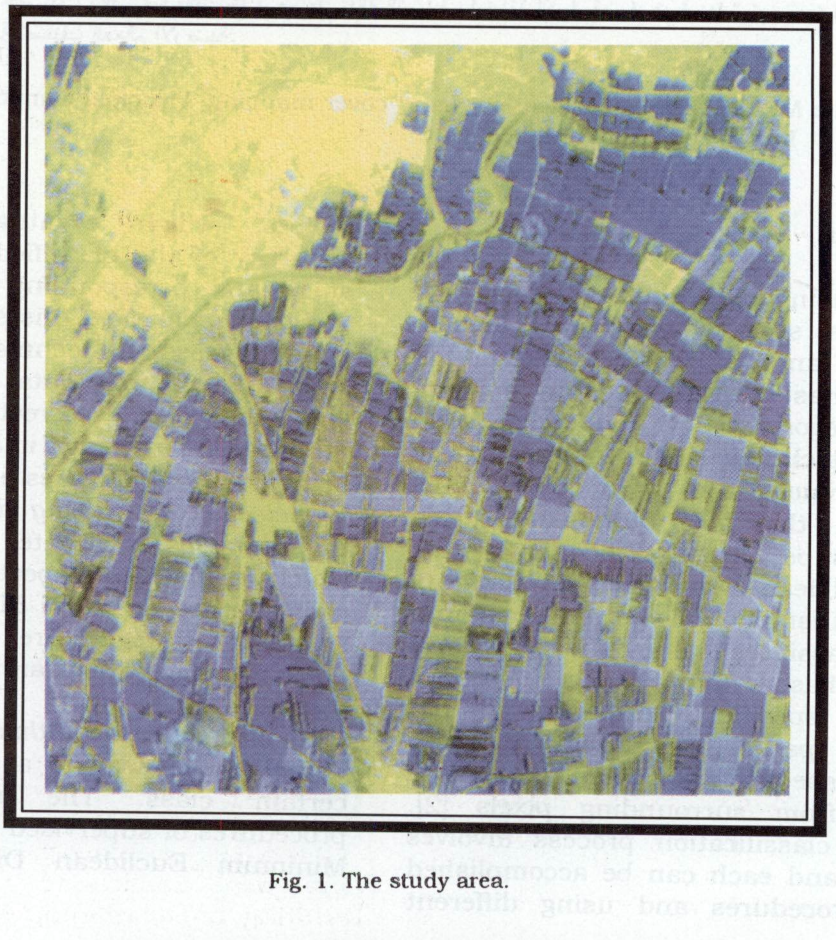

Fig. 1. The study area.

3. Data preparation

An examination of the histograms of the three original SPOT XS bands showed that the effective spectral ranges of the digital data are small as shown in table 1. A histogram equalization enhancement has been applied to the three bands aiming to occupy the full spectral range, and to facilitate the selection of the training samples.

Two transformed images were involved in the classification process. These two images were created by the principal component analysis (PCA) from the original bands of SPOT SX.

Table 1
Statistics of digital values of SPOT XS bands

	Band No.	Mean value	SD	Min. value	Max. value
Before enhancement	1	54.093	11.81	39	169
	2	44.21	17.71	25	184
	3	95.07	18.73	23	166
After enhancement	1	133.51	71.49	0	255
	2	131.64	71.47	0	255
	3	129.43	73.73	0	255

The correlation factor between each pair of bands is calculated and presented in table 2. The two visible bands of the SPOT data are highly correlated (0.99) while almost no correlation exists between the first two principal component images (-0.04). Moreover, the scatter diagram between the first two PCA images is presented in fig. 2. The three original bands of SPOT XS image, and the first two principal component images are used as two data sets for the classification analysis.

Table 2
The correlation factors between bands

Band	2	3	PC1	PC2
1	0.99	-0.33	0.88	0.42
2		-0.39	0.91	0.37
3			-0.73	0.70
PC1				-0.04

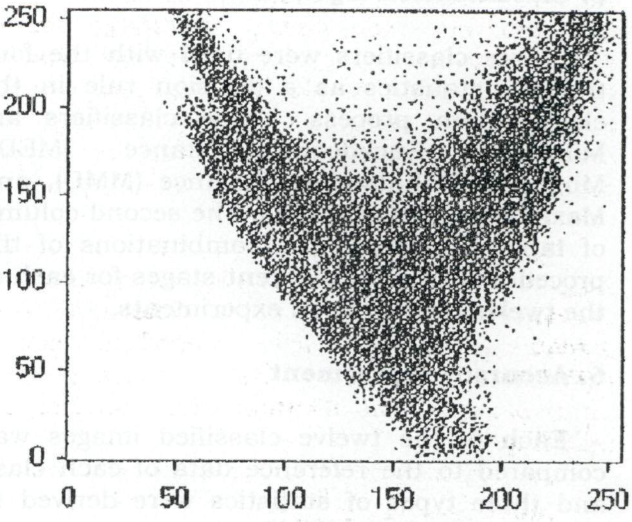

Fig. 2. The scatter plot of the two PCA images.

4. Determination of training samples

Bands 1, 2, and 3 of the SPOT XS image were displayed as a false-color composite image. The study area was segmented into relatively homogenous fields bounded by the irrigation and drainage systems. Unsupervised classification was carried out and the test of separability was applied. Only four clusters or themes were chosen because of their high spectral separability. The training and reference pixels of the four classes namely; crop1, crop2, crop3, and urban areas were located inside these fields and according to the unsupervised classification results. The selection of the training samples was also aided using a topographic map number NH36-I6C of scale 1 : 50,000 produced by the Egyptian General Survey Authority.

Two procedures of selecting training samples were considered. The first is block sampling (BS) procedure, which requires the selection of blocks of pixel from representative fields for each class. The second is single pixel sampling (SS) procedure, which requires the selection of a group of single pixels for each class [5]. The individual pixels were chosen to be at least ten pixels away from each other to avoid the auto-correlation between pixels, which are close together.

The two procedures of sampling training data were applied to the two data sets resulting in four training data statistics.

5. Classification algorithm

Three classifiers were used with the four training statistics as a decision rule in the classification process. These classifiers are Minimum Euclidean Distance (MED), Minimum Mahalanobis Distance (MMD), and Maximum Likelihood (ML). The second column of table 5 presents the combinations of the procedures used at different stages for each of the twelve classification experiments.

6. Accuracy assessment

Each of the twelve classified images was compared to the reference data of each class and three types of statistics were derived to evaluate their accuracies:

1- The confusion matrix (sometimes called the error matrix or the contingency table). Many indices can be derived from the confusion matrix. Some to describe the whole accuracy of the classified data such as the overall accuracy and the average accuracy. Others to represent the individual land-cover class accuracy such as the producer accuracy, and the user accuracy [6].

2- The Kappa (k) coefficient of agreement which is also derived from the confusion matrix. The estimate of Kappa is the proportion of agreement after chance agreement is removed from consideration. The result of performing Kappa analysis is a KHAT statistic, which is computed as;

$$\hat{K} = \frac{p_o - p_c}{1 - p_c},$$

where,

$$p_o = \sum_{i=1}^{m} p_{ii}, \qquad p_c = \sum_{i=1}^{m} p_{i+} \times p_{+i},$$

m is the number of rows in the confusion matrix,

p_{ii} is the proportion of pixels in row i and column i,

p_{i+} is the proportion of the marginal total of row i, and

p_{+i} is the proportion of the marginal total of column i.

The KHAT coefficient is more representative for the whole accuracy of the classification process as it takes into account all the elements of the matrix rather than the overall accuracy, which just considers the diagonal elements [7].

3- The Z statistic of significance test, also called the standard normal deviate is used to compare the performance of two independent classifications (two KHATs) carried out using different procedures. It can be derived by the formula:

$$z = \frac{\hat{K}_1 - \hat{K}_2}{[v(\hat{K}_1) - v(\hat{K}_2)]^{1/2}},$$

where,

$$v(\hat{K}) = \frac{1}{N(1-p_c)^4} \left\{ \sum_{i=1}^{m} p_{ii}[(1-p_c) - (p_{i+} + p_{+i})(1-p_o)]^2 + (1-p_o)^2 \sum_{\substack{i=1 \\ i \neq j}}^{m} \sum_{j=1}^{m} p_{ij}(p_{i+} + p_{+j})^2 - (p_o p_c - 2p_c + p_o)^2 \right\},$$

V is the variance of KHAT, and N is the total number of the reference pixels.

The Z statistic is used to examine if a certain procedure has a significant influence on the classification accuracy or not and to which confidence level. This was achieved by changing the examined procedure only in the two compared classifications and keeping all other procedures unchanged. The Z-values shown in table 3 are obtained from Gaussian normal distribution and are equal to values of the student's distribution where the degree of freedom is infinity. These values provide the confidence level to which the two compared classification results are significantly different.

7. Results and analysis

The confusion matrices of the twelve classification experiments have been generated. An example for experiment No.4 is shown in table 4.

Table 3
The relation between the Z value and the confidence level

Z value	1.645	1.960	2.326	2.576	3.090	3.291
Confidence level %	90	95	98	99	99.8	99.9

Table 4
The confusion matrix of experiment No. 4

		Reference data				Total	User accuracy %
		Crop1	Crop2	Crop3	Urban		
Classified data	Crop1	107	23	4	2	136	78.67
	Crop2	7	82	0	1	90	91.11
	Crop3	12	11	90	15	128	70.31
	Urban	0	0	0	177	177	100.00
	Total	126	116	94	195	531	
	Producer accuracy %	84.92	70.69	95.74	90.77		

Average accuracy = 85.53% Overall accuracy = 85.88%

The elements of the last row provide the percentage that a reference data of each class will be correctly classified. This percentage value is referred to as the producer accuracy and can be obtained by dividing the number of pixels correctly classified to a class by the total number of reference pixels of that class (107/126= 84.92% for crop1). In other words this value is a measure for the percentage of the omission errors (15.08% for crop1).

The elements of the last column provide the percentage that a sample of a class chosen from the classified image actually represents that class on the ground. This percentage value is referred to as the user accuracy and can be obtained by dividing the number of pixels correctly classified to a class by the total number of pixels classified as that class (107/136= 78.67% for crop1). In other words this value is a measure for the percentage of the commission errors (21.33% for crop1).

The average accuracy of a classified image is obtained by averaging the producer accuracies of the individual classes. The overall accuracy of a classified image is calculated by dividing the sum of the elements of the major diagonal by the total number of reference pixels [(107+82+90+177)/531= 85.88%].

The average accuracy, the overall accuracy, the KHAT coefficient, and the standard deviation of KHAT were calculated for each classification results and summarized in table 5.

It can be observed from table 5, that the use of single pixel sampling (SS) procedure has improved the classification accuracy whatever is the classifier or the data set used.

The accuracies of the classifications adopted using the original three SPOT XS bands are higher than those adopted using the two transformed principal component images except when the Maximum Likelihood classifier was employed.

Accuracy of classification results using MMD classifier proved to be higher than that using ML classifier. This is because the number of tested pixels located behind the threshold of ML classifier. These pixels reduce the producer, and consequently the average and the overall accuracies. However, the user accuracy of individual classes is higher using the ML classifier. The use of MED classifier produced the least accurate results of the three classifiers. This is referred to that MED classifier does not evaluate the standard deviation nor the probability density of the training data clusters.

In order to evaluate the confidence level to which each of the used procedure could significantly affect the classification accuracy, the significance test for comparing two classification results was adopted. Since there are more than two classification results to be

Table 5
Configuration and results of the classification experiments

Exp.	Configuration	Average accuracy %	Overall accuracy %	KHAT	SD
1	XS + BS + MED	71.83	73.90	0.64976	0.01446
2	XS + BS + MMD	79.19	81.82	0.73328	0.01316
3	XS + BS + ML	75.95	77.87	0.71157	0.01370
4	XS + SS + MED	85.53	85.88	0.80838	0.02018
5	XS + SS + MMD	94.47	94.73	0.90785	0.01321
6	XS + SS + ML	88.46	89.12	0.86560	0.01714
7	PC + BS + MED	68.95	70.81	0.60947	0.01496
8	PC + BS + MMD	77.45	80.09	0.73028	0.01352
9	PC + BS + ML	76.41	78.63	0.71254	0.01376
10	PC + SS + MED	84.70	84.56	0.79116	0.02090
11	PC + SS + MMD	93.87	94.16	0.90020	0.01382
12	PC + SS + ML	89.60	90.21	0.86916	0.01670

Table 6
The Z-values

BS versus SS			XS versus PC		
Group	Z-value	All pairs	Group	Z-value	All pairs
XS	7.397	7.487	BS	0.716	0.478
PC	7.576		SS	0.291	
MED	6.736	7.487	MED	1.138	0.478
MMD	9.065		MMD	0.280	
ML	7.134		ML	0.104	

MED versus MMD			MED versus ML			MMD versus ML		
Group	Z-value	All pairs	Group	Z-value	All pairs	Group	Z-value	All pairs
XS	4.167	4.620	XS	2.539	3.181	XS	1.569	1.380
PC	5.054		PC	3.812		PC	1.189	
BS	5.147	4.620	BS	4.102	3.181	BS	1.031	1.380
SS	4.238		SS	2.541		SS	1.692	

compared to explore the significant difference between two alternative procedures, an average KHAT and an average KHAT variance have been calculated for each procedure. The Z-value for each comparison has been computed and presented in table 6.

According to table 3, the higher the Z-value is an indication to the more confidence level there is that the two examined procedures are significantly different from each other. Thus, one can nominate the best procedure at each of the classification stages to be used in order to improve the classification results.

8. Conclusions

In this study, supervised classification experiments have been carried out using the most frequently used procedures at each step to evaluate the classification performance with different factor and procedure combinations. Several accuracy indices have been derived from the results and the following points were concluded.

- In all combinations presented, single pixel training procedure resulted in higher classification accuracies than block training procedure. The significant improvement of classification accuracy due to using single pixel training strategy rather than block training strategy has been demonstrated to have a confidence level of 99.9%.
- Minimum Mahalanobis Distance classifier led to more accurate classification results than either Minimum Euclidean Distance or Maximum Likelihood classifiers. However, the best user accuracy of individual classes was obtained when Maximum Likelihood classifier was used.
- Comparing the results of the two data sets, a small improvement of accuracy using the original SPOT bands was realized. Nevertheless, when Maximum Likelihood classifier was employed, the two transformed principal component images provided more accurate classification results than the three original SPOT XS bands.
- From the results of table 6, one can rank the procedures involved according to their significant improvement of the classification accuracy from the most to the least as follow:
    - ♦ Single pixel training procedure.
    - ♦ Minimum Mahalanobis Distance classifier.
    - ♦ The original SPOT multispectral bands.

It should be noted however, that if the application in hand is concerned with the user accuracy of individual categories, it would be recommended to use Maximum Likelihood classifier with the principal component images.

References

[1] T. M. Lillesand, and R. W. Kiefer "Remote Sensing and Image Interpretation ", Third Edition, John Wiley & Sons Inc. (1994).

[2] J. R. Jensen " Introductory Digital Image Processing", Englewood Cliffs, Prentice-Hall (1986).

[3] P. Gong and P. J. Howarth, "The Use of Structural Information for Improving Land Cover Classification Accuracies at Rural-Urban Fringe", Photogrammetric Engineering and Remote Sensing, Vol. 56 (1), pp. 67-73 (1990).

[4] P. Gong and P. J. Howarth, " Frequency-Based Contextual Classification and Grey-Level Vector Reduction for Land-Use Identification", Photogrammetric Engineering and Remote Sensing, Vol. 58 (4), pp. 423-437 (1992).

[5] R. G. Congalton "A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data", Remote Sensing of Environment, Vol. 37, pp. 36-46 (1991).

[6] M. Story and R. Congalton, " Accuracy Assessment: A User's Perspective", Photogrammetric Engineering and Remote sensing, Vol. 52 (3), pp. 397-399 (1986).

[7] G. H. Rosenfield and K. Fitzpatrick-Lins, "A Coefficient of Agreement as A Measure of Thematic Classification Accuracy", Photogrammetric Engineering and Remote Sensing, Vol. 52 (2), pp. 223-227 (1986).

Received April 7, 2002
Accepted May 11, 2002