Supervised Classification Of Multispectral Image & Accuracy Assessment

[1] B. Naga Jyothi, [2] K. S. R. Radhika , [3] Dr I. V. Murali Krishna

[1] Assoc. Prof., Dept. of ECE, DMS SVHCE, Machilipatnam, A.P., India,
[2] Assoc. Prof., Dept. of CSE, DMS SVHCE, Machilipatnam, A.P., India,
[3] Director (Retd.,), IST/R&D Center, JNTU, Hyderabad, India,

Abstract: In a multispectral image the brightness values in the different wavelength bands encode the spectral information for each image cell and can be regarded as a spectral pattern. Spectral classification methods seek to categorize the image cells on the basis of these spectral patterns. In supervised classification the analyst designates a set of "training areas" in the image, each of which is a known surface material that represents a desired spectral class. The classification algorithm computes the average spectral pattern for each training class, then assigns the remaining image cells to the most similar class. In this paper, accuracy assessment is done considering all pixels in the classified image using matlab programming. This is an advantage over some of the existing image processing softwares in which classification is performed considering only limited pixels. Finally Error matrix, Producer's accuracy (Omission Error), User's accuracy (Commission Error), Overall accuracy and Kappa coefficient are obtained for two different band combinations.

Keywords: *Multispectral Images, Supervised Classification, Training Samples, Accuracy Assessment*

---

1 INTRODUCTION

Multispectral image analysis in this paper is done using the 7 image files ch1.tiff, ch2. tiff, ch3.tiff, ch4.tiff, ch5. Tiff, ch6.tiff, ch7.tiff of Copenhagen City, Denmark obtained from LANDSAT 5 TM which provides 7 bands of data. [eduspace Home 1]. The images used here are 512*512 pixel size. The LANDSAT system gives images with a high resolution of 30 m * 30m. The 7 individual image files are combined to a single ab.mat file.

1.1 MULTISPECTRAL ANALYSIS

The LANDSAT 5 TM satellite 7 bands are given below in **Table: 1**

Band 1	0.45-0.52 µm	visible light, blue
Band 2	0.52-0.60 µm	visible light, green
Band 3	0.63-0.69 µm	visible light, red
Band 4	0.76-0.90 µm	near infrared
Band 5	1.55-1.75 µm	middle infrared
Band 6	10.4 -12.5 µm	thermal infrared
Band 7	2.08 -2.35 µm	middle infrared

Three-band composite images are created by using three LANDSAT spectral bands to control the amount of red, green and blue in a color image. **True-color composite images** with combinations of bands 1, 2 and 3 approximately match the spectral range of vision for the human eye, so these images appear to be close to what would be expected in a normal color photograph.

A false-color image is an artificial representation of a multispectral image.

The specific bands used in three-band composites are often identified by giving the band numbers used for red, green, and blue in a specific order. Thus, an image using band 7 for red, band 4 for green, and band 2 for blue would be designated (7,4,2). As can be seen in the simulation results [**Fig: 1**], the shortwave composite of the bands 7, 4, 2 shows the forest in red and the cultivated

areas in different shades of red and pink. Roads are visible as straight light-colored lines. The river is in dark blue color.

1.2 SUPERVISED CLASSIFICATION

In supervised classification ,the identity of land cover types are known priorly through some means such as aerial photography, map analysis and personal experience etc. The analyst locates specific sites that represent homogeneous examples in the remotely sensed data. These areas are referred as training samples and the spectral characteristics of these areas are used to train the classification algorithm of the remainder of the image. Multivariate statistical parameters such as mean, standard deviation, covariance matrices, correlation matrices are calculated for each training site. Every pixel both within and outside the training sites is then evaluated and assigned to the class of which it has the highest likelihood of being a member.

Thematic classification of an image involves the following steps:

· **Feature extraction**: Transformation of the multispectral image by a spatial or spectral transform to a feature image. Examples are selection of subset of bands, a PCT to reduce the data dimensionality, or a spatial smoothing filter. This step is optional i.e., the multispectral image can be used directly, if desired

· **Training:** Selection of pixels to train the classifier to recognize the different themes, or classes, and determination of decision boundaries which partition the feature space according to the training pixel properties. This step is either supervised by the analyst or unsupervised with the aid of a a computer algorithm. For supervised training, the analyst must select representative pixels for each of the categories . It is important that the training area be a homogenous sample of the respective class, but at the same time include the range of variability for the class.

· **Labeling**: Application of the feature space decision boundaries to the entire image to label all the pixels. If the training was supervised, the labels are already associated with the feature space regions; if it was unsupervised, the analyst must now assign labels to the regions. The output map consists of one label for each pixel.

1.3. THE ERROR MATRIX

Producer's Accuracy (Omission error) , User's Accuracy (commission error).

To perform a classification accuracy or error assessment , two sources of information are to be compared.

- Pixels in the remote sensing-derived classification map and
- Ground reference test information

The relationship between the two sets of information is summarized in an **error matrix.** The error matrix is used to assess the remote sensing classification accuracy of '*k*' classes. The central part of the error matrix is a square array of numbers $k \times k$ in size. The columns of the matrix represent the ground reference test information and the rows correspond to the classification generated from analysis of remotely sensed data. The intersection of the rows and columns summarize the number of sample units (pixels) assigned to a particular category(class) relative to the actual category as verified in the field. The total number of samples examined is N.

The diagonal of the error matrix summarizes those pixels that were assigned to the correct class. Every error in the remote sensing classification relative to the ground

reference information is summarized in the off-diagonal cells of the matrix.

Each error is both an omission from the correct category and a commission to a wrong category. The column and row totals around the margin of the matrix are used to compute errors of inclusion (commission errors) and errors of exclusion (omission errors). The outer row and column totals are used to compute **producer's and user's accuracy.**

1.4. ACCURACY ASSESSMENT

Kappa Analysis

Kappa analysis is a discrete multivariate technique of use in accuracy assessment.

K_{hat} Coefficient of Agreement: kappa analysis yields a statistic $\hat{K}$, which is an estimate of Kappa. It is a measure of agreement between the remote sensing – derived classification map and the reference data as indicated by a) the major diagonal and b) the chance agreement, which is indicated by the row and column totals (referred to as marginals).

$$\hat{K} = \frac{N \sum_{i=1}^{k} x_{ii} - \sum_{i=1}^{k} (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^{k} (x_{i+} \times x_{+i})}$$

where k is the number of rows in the matrix, x_{ii} is the number of observations in row i and column i, and x_{i+} and x_{+i} are the marginal totals for row i and column i respectively and N is the total number of observations. Values of $\hat{K} > 0.8$ (80%) represent strong agreement, values between 0.4 and 0.8 represent moderate agreement and values less than 0.4 represent poor agreement.

2. METHODOLOGY

The problem is to segment the image into meaningful classes like forest, water, urban area , roads and agricultural areas. Simulation is done using Matlab [2]. Simulation involves the sequence of following steps:

·The 7 individual image files ch1.tiff, ch2.tiff, ch3.tiff, ch4.tiff, ch5.tiff, ch6.tiff, ch7.tiff are combined to a single ab.mat file using the matlab command **savefile**

·The image data can be loaded into MATLAB using the command **load ab**

· It is also possible to view three bands at a time using color or pseudo color. **imtool** command is used showing bands (7,4,2) converted to R,G,B

· **imshow** command is used to view separate and also combinations of bands in order to visualize differences and to improve the classification results

·**label_im** command is used to create training data set for each class.

·Training areas are marked using coordinates for rectangles and training labels are shown using the command **imagesc(label_im)**

·Classification tools like classify, scatterplots are used for different bands. Individual bands are separated and classification result is observed using various band combinations using the command
ImC = ExtendedClassify(ab, label_im)

·Classification result is observed using 4,6 & 2,4,6,7 band combinations.

Finally, Accuracy Assessment is carried out taking all the pixels into consideration using a matlab program by comparing the classified output array with a reference image array generated on the basis of visual spectral inspection of the input image.

3. RESULTS

Fig: 1 Input Image with Band Combination [7,4,2]

Fig: 2 Training Labels

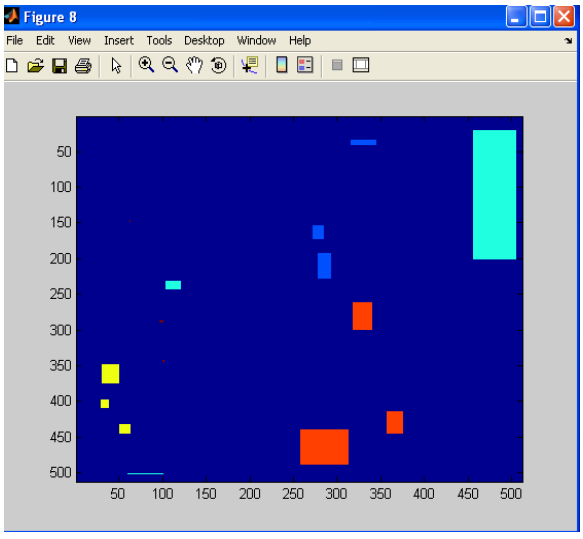


Fig: 3 Classification using bands 4, 6

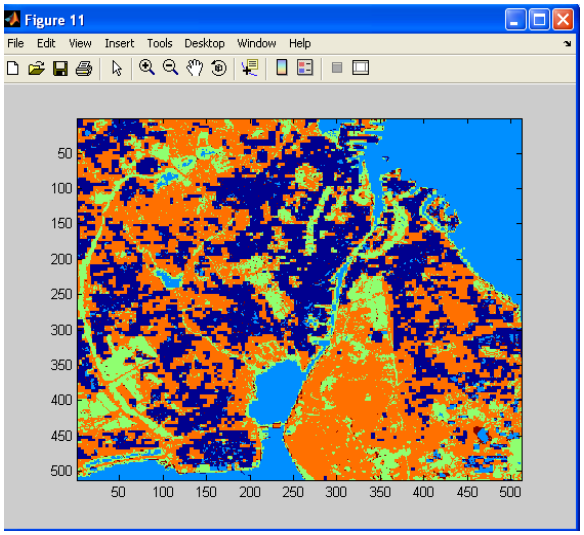
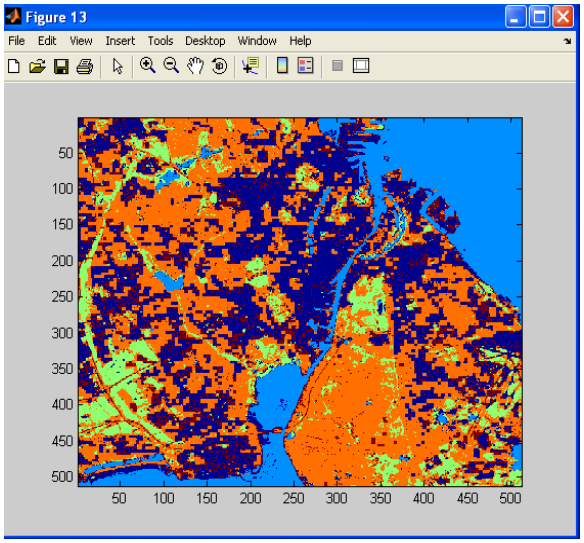


Fig: 4 Classification using bands 2,4,6,7



4. ACCURACY ANALYSIS:

Using Bands 4,6

ERROR MATRIX:

29782	0	135	45854
0	40151	358	4315
1372	3776	10935	23595
15615	0	93	86163

PRODUCER'SACCURACY-OMMISSION ERROR :

CLASS 1 :	63.68	36.32
CLASS 2 :	91.40	8.60
CLASS 3 :	94.91	5.09
CLASS 4 :	53.88	46.12

USER'S ACCURACY - COMMISSION ERROR :

CLASS 1 :	39.31	60.69
CLASS 2 :	89.57	10.43
CLASS 3 :	27.56	72.44
CLASS 4 :	84.58	15.42

OVERALL ACCURACY : 63.72%

KAPPA COEFFICIENT : 0.46

ACCURACY ANALYSIS:

Using Bands 2,4,6,7

ERROR MATRIX:

30293	2582	719	50024
0	41107	18	2593
280	238	9872	18732
16196	0	912	88578

PRODUCER'SACCURACY-OMMISSION ERROR :

CLASS 1 :	64.77	35.23
CLASS 2 :	93.58	6.42
CLASS 3 :	85.69	14.31
CLASS 4 :	55.39	44.61

USER'S ACCURACY - COMMISSION ERROR :

CLASS 1 :	36.23	63.77
CLASS 2 :	94.03	5.97
CLASS 3 :	33.90	66.10
CLASS 4 :	83.81	16.19

OVERALL ACCURACY : 64.79%

KAPPA COEFFICIENT : 0.47

5. DISCUSSION ON RESULTS

Different bands reveal information about different areas and so some combination of bands might give good classification result for one set of classes and poor for the remaining classes.

All the classes seem to be separated quite well, when **bands 2, 4, 6 and 7** are used for classification. Overall Accuracy is also found to be high for this band combination.

Softwares like Erdas can be used to compute Error matrix, Producer's accuracy (Omission Error), User's accuracy (Commission Error), Overall accuracy and Kappa coefficient . But the limitation with these softwares is that the above parameters are computed considering only few randomly selected pixels. If this few randomly selected pixels are classified

accurately then it results in 100% overall accuracy. If on the other hand , if all the randomly selected pixels are not classified correctly then it results in 0% overall accuracy. **Thus the overall accuracy is determined based only on the classification result of few randomly selected pixels.**

However in this paper a separate matlab program is developed as mentioned in the methodology and then the Error matrix, Producer's accuracy (Omission Error), User's accuracy (Commission Error), Overall accuracy and Kappa coefficient are computed by considering **all the pixels in the image. Thus the overall accuracy is computed based on the classification result of all the pixels in the output classified image. Obviously this method yields in much accurate output results.**

6. REFERENCES

[1] eduspace Home, http://Copenhagen Landsat Images and Band Combinations

[2] Centre for image Analysis, Swedish University of Agricultural Sciences,Uppsala University

[3] Schowengerdt, Robert A. (1997). *Remote Sensing: Models and Methods for Image Processing* (2nd ed.). New York: Academic Press. 522 p
.

[4] Remote Sensing of Environment (RSE) withTNTmips® TNTview®

[5]. D. LU and Q. WENG "A survey of image classification methods and techniques for improvingclassification performance" International Journal of Remote Sensing Vol. 28, No. 5, 823–870, 10 March 2007

[6]. Marçal, A.R.S., Borges, J.S., Gomes, J. A. and Costa, J.F.Pinto Da,." Land cover update by supervised classification of segmented ASTER images". *Int. J. Remote Sensing*, 26(7),pp.1347-1362., 2005

[7]. Foody,G.M.,2002 "Status of Land Cover Classification Accuracy Assessment" Remote Sensing of Environment,80:185-201

[8]. John R.Jensen "Introductory Digital Image Processing", A remote Sensing perspective, Third Edition, PPH, 338p, 499-500p

[9]. Erdas,2003, Erdas Field Guide, Atlanta:Leica GeoSystems

[10]. Congalton, R.G,1991, "A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data," Remote Sensing of Environment,37;35-46

[11]. Hudson,W and C.Ramm,1987,"Correct Formulation of the Kappa Coefficient of Agreement," Photogrammetric engineering & Remote Sensing,53(4):421-422

[12].B.Naga Jyothi, Dr. G. R. Babu, Dr.I.V.Murali Krishna, "Thematic classification of multispectral imagery, International Journal of Electronics and computer Science Engineering , V1N2-181-190,2012

[13].KSR Radhika, B.Naga Jyothi, Dr. Ch.Venkateswara Rao, Dr.V.Kamakshi Prasad, "Accuracy Assessment of per pixel Based Classification" Conference Proceedings, NC-Velasiem-2k12: Pg.73-77