International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-04, Issue-03, June 2018

Comparision of Pixel Based Classification Using Kappa Analysis

K.S.R.Radhika[1], Ch.Venkateswara Rao[2], V.Kamakshi Prasad[3]

[1] Associate Professor, DMS SVH College of Engineering,Machilipatnam, India

kammilisrr@gmail.com

[2] Group Head, National Remote Sensing Centre, Hyderabad, India

[3] Director of Evaluation, JNTU, Hyderabad, India.

Abstract: **The paper presents classification of LISS-III multispectral satellite image applying supervised classification technique. The output classified image is observed through maximum likelihood, minimum distance to means, parallelopiped and mahalanobis classification algorithms. In the classified image all pixels are considered while calculating the accuracy assessment. Comparision of pixel based classification algorithms is done using kappa analysis. In this analysis different accuracy parameters namely error matrix, commission and omission error, User's and producer's accuracy, overall accuracy and kappa coefficient are computed for different algorithms considering total pixels in the image.**

Keywords: Kappa Coefficient, Supervised Classifiers, Error Matrix, Accuracy Assessment.

I. INTRODUCTION

Data Collection in Remote sensing : Passive or active remote sensing systems are used in data collection. Electromagnetic radiation reflected or emitted from the targeted area of interest is recorded with passive sensors. Target is illuminated with electromagnetic energy and radiated energy, scattered back is recorded in active sensors. Analog and/or digital data is collected by remote sensing systems. Models (relationships) are developed between the amount of backscattered/emitted/reflected electromagnetic (EM) energy within specific bands. The process study of biological, chemical and physical properties forms the basis for investigating several remote sensing (RS) systems [5].

Energy is recorded in multiple bands of EM spectrum in multi spectral remote sensing systems. The size and number of bands in EM spectrum for which RS unit is sensitive is called spectral resolution. MSS(Multi Spectral Scanner) recorded data in four bands, Band 1, 2, 3 and 4 in 1970's and 1980's. Data is acquired in hundreds of spectral bands in hyperspectral remote sensing instrument. In RS system smallest distinguishable separation of two objects is called spatial resolution. Measure of smallest angular/linear separation amongst two objects that could be distinguished by remote sensing systems is spatial resolution. How periodically the sensor records data of a particular area is temporal resolution. Remote sensing detector sensitivity to signal strength differences while it records the energy backscattered/emitted/reflected from the terrain is radiometric revolution. Simply this defines the number of mere discriminable signal levels [6].

Fig 1: Input image

Multi based LISS-III satellite image relation captured in Narsapur region, outside of Hyderabad city is considered as an input image for study. Green, red, NIR, MIR serve as bands 1,2,3 and 4 of input LISS-III image. Input image uses standard False Color Composite (NIR, Red, Green). Hence the dominating classes namely vegetation is displayed in red, water in black etc. The input image is shown in Figure 1. The input image is downloaded from Bhuvan website of National Remote Sensing Center, Hyderabad [10].

 DOI : 10.18231/2454-9150.2018.0424

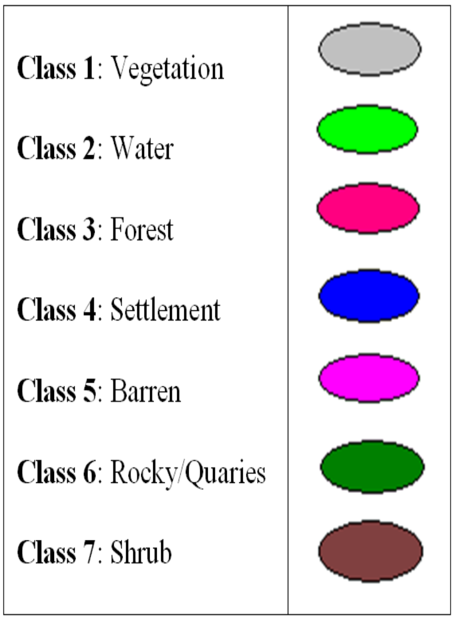

Fig 2: Different prominent classes

II. BACKGROUND

In [1] very high resolution (HR) imagery classification and a novel protocol for accuracy assessment is presented. While only the thematic accuracy is inadequate for characterizing the geometrical properties of HR classifying maps, a protocol in proposed basing on two indices family analysis namely 1.Conventional Thematic Accuracy Indices, 2.Set Involving Unique Geometric Indices. Modeling various geometric object characteristics identified in the map within this reference set of indices are presented characterizing five categories of geometric errors within the classification map. They are 1.Fragmentation, 2.Shape Distortion, 3.Under Segmentation, 4.Over Segmentation, 5.Edge Location. The paper proposes a novel method for varying free parameters related to supervised classifiers depending on multi objective criterion function aiming at choosing the parameter values resulting the classification map. This results in optimizing geometric and thematic error indices.

For determining the information quality obtained through classification accuracy assessment is used. Error matrix used with remote sensing accuracy assessment independent on evaluating the obtained classification across some ground truth information. Classification methods of serial photographs and parameters of accuracy assessment computation. In [2] fuzzy classifier is presented without ground data reference. Additionally for statistical analysis usage standard errors of accuracy are computed with stop resampling techniques.

Thematic map production such as land cover depiction with the usage of image classification serves as common remote sensing application. Significant research has been targeted towards various components of mapping, such as accuracy assessment. Various methods of classification accuracy assessment and the background are presented in [3]. Problem areas currently limiting the capability to assess, document, usage of thematic maps accuracy obtained from remote sensing are discussed.

A method in economics theory namely 'Pareto Boundary' is proposed to analyze quantitative trade off between commission and omission errors. This method is used for dichotomic classification [9].

III. SUPERVISED CLASSIFICATION

In supervised classification, land cover categories are identified using few means like aerial photography analysis through maps, personal experience [7]. Some sites of interest are to identified by the analyst representing homogeneous samples within the remote sensed data. These are called “training samples”. For training the classifying algorithm for the rest of the image, these training samples spectral characteristics are used. For each of these training samples, statistical parameter such as correlation matrices, mean, covariance matrices, standard deviation are found. Evaluation is carried for all pixels, within and outside the training areas and each pixel is then allotted to a particular class to which it is the maximum likelihood of being to be a member [4].

IV. KAPPA ANALYSIS

Kappa analysis is a discrete multivariate technique used with accuracy assessment. A static $\widehat{K}$, an approximate of Kappa is the outcome to Kappa analysis. It symbolises agreement between reference data of classification map and the classified data [8]. This is as given by (i) chance agreement expressed by row totals and column totals (marginal’s) and (ii) major diagonal. If K<0.4, it is poor agreement and if 0.4<K<0.8, it indicates moderate agreement, and if K>0.8 (80%) , it denotes stronger agreement. The kappa coefficient is given as below.

$$\widehat{K} = \frac{\sum_{i=1}^{k} X_{ii} - \sum_{i=1}^{k}(X_{i+} * X_{+i})}{N^2 - \sum_{i=1}^{k}(X_{i+} * X_{+i})}$$

where k : row number in matrix
N : total observations number
X_{+i} : marginal totals of column ‘i’
X_{i+} : marginal totals of row ‘i’
X_{ii} : observations number in row ‘i’ and column ‘i’.

V. ERROR MATRIX

For error assessment or classification accuracy computations, there is a need to compare two sources of information.

1.Classification map pixels obtained through remote sensing

2.Test information regarding ground reference.

 DOI : 10.18231/2454-9150.2018.0424

Error Matrix is the relationship summarized between the information of the above said two sets. Error matrix is useful in assessing the remotely sensed classification accuracy related to k different classes . k x k square array of numbers form the central part of the error matrix. While computing the error matrix, classified data is given as rows and ground reference data is given as columns. The diagonal elements of error matrix correspond to the summary of particular class pixels number that are correctly classified against actual class as found in reference data. 'N' represents total number of examined classes.

The pixels arranged to correct class are summarized by the diagonal in the error matrix. In remote sensing classification process with respect to the ground reference data, the errors are represented as off diagonal cells in the error matrix. Every error in simultaneously commission to wrong class and omission from correct class. the row totals and column totals are used in computing omission or exclusion errors and commission or inclusion errors. Producer's and user's accuracy are computed using error matrix: through the outer column totals and outer row totals.

VI. METHODOLOGY

Work presents classification of input image Figure 1, through different classification techniques. Classifiers used for analysis are maximum likelihood, minimum distance to means, parallelopiped and mahalanobis. For each method, accuracy assessment is done covering all pixels by comparing classified output obtained through each technique with the reference image which is in turn acquired on ground truth information of field suspension basis.

Procedure for obtaining reference image array:

Step1: Digitization of input image and vector format generation.

Step2 : Conversion of vector format to raster format

Step3 : ASCII file generation using raster format

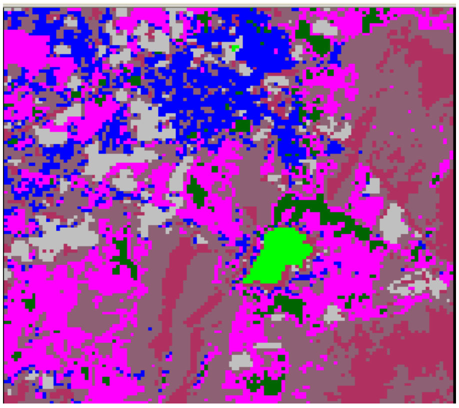

Fig 3: Minimum distance classified output

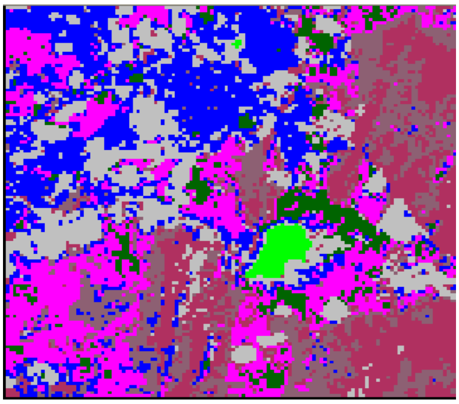

Fig 4: Parallelopiped classified output

This serves as the reference array and the classified images are used as the output array. These two arrays are used in the analysis. These results are required for comparison in the accuracy assessment process. The classified images using the different classifiers namely maximum likelihood, parallelepiped, minimum distance to means, mahalanobis are shown in Figures 3, 4, 5 and 6 respectively. All the seven classes are obtained with different colors. The seven classes are shown in Figure 2.

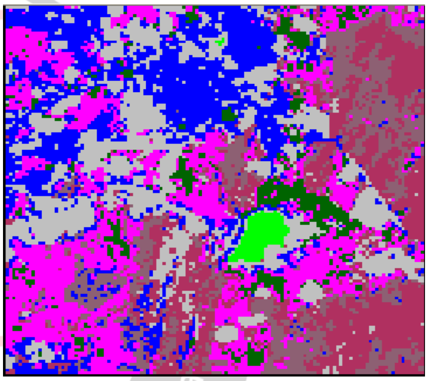

Fig 5: Maximum likelihood classified output

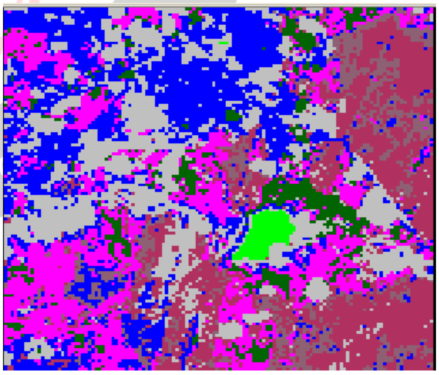

Fig 6: Mahalanobis classified output

VII. ACCURACY ASSESSMENT USING KAPPA ANALYSIS

The accuracy assessment is done by using the reference image and the classified images obtained using different supervised classifiers. The result of one classifier namely maximum likelihood is shown in Table 2 and 3. Table 2 shows the error matrix. In Table 3 the user's accuracy, producer's accuracy, commission error, omission error of

 DOI : 10.18231/2454-9150.2018.0424

all the seven classes is shown. Similarly the results of the other three classifiers is computed using the same reference image. Table 4 shows the overall accuracy and kappa coefficient of all the four classifiers.

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7
Class 1	1143	3	2	122	334	31	37
Class 2	0	208	0	0	0	0	0
Class 3	36	50	1362	213	51	0	353
Class 4	5	0	0	1569	318	18	115
Class 5	9	0	1	772	2010	94	1213
Class 6	26	0	0	73	194	408	35
Class 7	121	0	383	1148	418	4	3839

Table 2. Seven Class Error matrix

	Commission error	Omission error	Producer's accuracy	User's accuracy
Class1	14.57	14.70	85.30	85.43
Class2	0.00	20.31	79.69	100.00
Class3	34.04	22.08	77.92	65.96
Class4	22.52	59.74	40.26	77.48
Class5	50.96	33.49	66.51	49.04
Class6	44.57	22.14	77.86	55.43
Class7	35.08	31.35	68.65	64.92

Table 3. Different accuracies and errors

	Overall accuracy	Kappa coefficient
Maximum likelihood	64.32	0.54
Parallelopiped	58.59	0.50
Minimum distance to means	56.43	0.47
Mahalanobis	51.20	0.42

Table 4. Kappa coefficient

VIII. RESULTS - DISCUSSION

While computing error matrix all the total number of pixels in the image are considered. Then commission error, omission error, user's accuracy, producer's accuracy are computed using the error matrix row totals and column totals. Finally overall accuracy and kappa coefficient are found for every classification algorithm. In the proposed method the accuracy assessment is performed considering all the pixels in the image. Maximum likelihood classification method is observed to produce relatively higher overall accuracy than the other algorithms.

IX. CONCLUSION

If classification software's like erdas are used, classification accuracy entirely depends on few randomly selected pixels. 100% accuracy results in if these pixels are not classified correctly. 0% accuracy will be the result if these pixels are not classified correctly. Hence this serves as a limitation as the accuracy assessment parameters of the output classified result is dependent on few pixels that are randomly chosen.

Presented work involves developing a novel 'c' program as referred in the methodology. Following it, error matrix is first computed. Later commission error, omission error, user's accuracy, producer's accuracy, overall accuracy and kappa coefficient are calculated by taking into account image total number of N pixels. Hence the computed overall accuracy depends on all the pixels of image of the classified output image and the corresponding reference image data. So assuring this way of computation results in high accurate output results.

REFERENCES

[1] Claudio Persello, Lorenzo Bruzzone. 2010. "A Novel Protocol for Accuracy Assessment in Classification of Very High Resolution Images." IEEE Transactions on Geoscience and Remote Sensing. Vol. 48, No. 3, pp 1232-1244, doi 10.1109/TGRS.2009.202970.

[2] F.Okeke, A. Karnieli. 2006. "Methods for fuzzy classification and accuracy assessment of historical aerial photographs for vegetation change analyses. Part I: Algorithm development." International Journal of Remote Sensing. Vol. 27, No. 1-2, pp 153-176, doi 10.1080/01431160500166540.

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

[4] Dengxin Dai and Wen Yang, 2011."Satellite Image Classification via Two-Layer Sparse Coding With Biased Image Representation" IEEE Geoscience and Remote Sensing Letters, Vol. 8, No. 1, 2011.

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

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

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

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

[9] Luigi Boschetti, Stephane P. Flasse, Pietro A. Brivio. 2004. "Analysis of the conflict between omission and commission in low spatial resolution dichotimic thematic products: The Pareto Boundary", Remote Sensing of Environment 91 (2004) 280-292, doi:10.1016/j.rse.2004.02.015.

[10] National Remote Sensing Centre, https://bhuvan.nrsc.gov.in.

 DOI : 10.18231/2454-9150.2018.0424