Land use and Land Cover Classification using RGB&L Based Supervised Classification Algorithm

[1] M.Renuka Devi, [2] Dr.S. Santhosh Baboo,

[1] M.Renuka Devi, Asst.Prof in MCA Department, Sree Saraswathi Thayagaraja College, Pollachi, E-mail id:renuka_srk@yahoo.co.in

[2] Lt.Dr.S.Santhosh Baboo, Reader, Postgraduate and Research department of Computer Science, Dwaraka Doss Goverdhan Doss Vaishnav College, Chennai
E-mail id:santhosh2001@sify.com

Abstract:**After the Geometric correction and resampling the image should be classified. This paper introduced a new method for classifying the areas in a remotely sensed image under the category of supervised classification techniques. This classification technique describes how to classify the geographical areas in given image under supervised classification techniques conventions. So to tell the abstract of this technique in this technique from the top left corner of the image the analysis begins. Each pixel from the top left corner is taken for analysis and is examined for its resemblance with the training data. In this method training data is defined by the color values rather than by other attributes like distance from any mean points or corner of the image. This method utilizes the channeling techniques to split each pixels RGB channels from its color value and in addition the color density or the brightness of the pixel also grabbed for examination. This paper introduces the RGB&L algorithm and analyzing its performance with other classification algorithm.**

Keywords:RGB&L, Classification, Hard Classifier, IRS IA, IRS IC

I.INTRODUCTION

A procedure that use the remotely sensed image data to produce maps and/or tables showing the location and extent of various selected land cover types or earth surface feature is called **Image classification** [2]. This is the next step of the enhancement. This is the most common ways to use remotely sensed data is to create land cover maps. This technique requires minimal prior knowledge of the area where a map is needed and easily incorporates ancillary data.

Image classification is an important part of the remote sensing, image analysis and pattern recognition. In some instances, the classification itself may be the object of the analysis. For example, classification of land use from remotely sensed data produces a map like image as the final product of the analysis [2]. The image classification therefore forms an important tool for examination of the digital images. Using this classification tool we can extract our own representation of land use/land cover information.

The term classifier refers loosely to a computer program that implements a specific procedure for image classification [2]. The analyst must select a classification method that will best accomplish a specific task. At present, it is not possible to state which classifier is best for all situations as the characteristic of each image and the circumstances for each study vary so greatly. Therefore, it is essential that each analyst understand the alternative strategies for image classification so that he or she may be prepared to select the most appropriate classifier for the task in hand.

At present, there is different image classification procedures used for different purposes by various researchers ([1], [3] ,[4],[6], [7], [8], [9]). These techniques are distinguished in two main ways as supervised and unsupervised classifications. Additionally, supervised classification has different sub classification methods, which are named as parallelepiped, maximum likelihood and minimum distances. These methods are named as Hard Classifier. In this work used RGB& L Based supervised classification methods. It result and performance are discussed below.

II.STEPS AND DESCRIPTION OF RGB&L ALGORITHM

RGB&L Based Classification Algorithm

Step 1: Describe the training data
Step 2: Input the image to processed
Step 3: Grab pixels
Step 4: Examination hall
Step 5: Display the results

Step 1: Describe the training data: Before analyzing the pixels we must describes the each class or Groups that are to be mapped in the image. And we must also define the colour values by which the group is identified. We must specify distinguishably varying colour values for each groups thus we can get most appropriate and accurate results. We must also define the density of colour depth or the brightness level of the classing colour. And we must also define the tolerance level of each channel or of the composite colour by which the comparison can make an adjustment while examination at each pixel.

Training area process is called signature creation is shown in the following figure 1.a and figure 1.b. In the figure some of the classes like cropland, water body, barren land and hills were chosen as training area. In this process the red pixels are trained as cropland, block pixels are trained as water body or tank, the ash green pixels are trained as barren land and dark gray pixels selected as Hills continue this process according to our classification scheme. Thus in total under this method we will need to provide the following data as training data set.

- Name or label of the group

- Red, Green, Blue channel values of the colour for the group (the range of the each channel depends on the bit size taken to calculation .For example each channel will range 0-255 if the bit size of calculation is of 8 bit)
- Tolerance of the colour from the described colour (Range of tolerance will be depend on the range of a colour channel and will be equal to the range of the colour channel for instance the range of tolerance will be 0-255 if channels use 8 bit level)

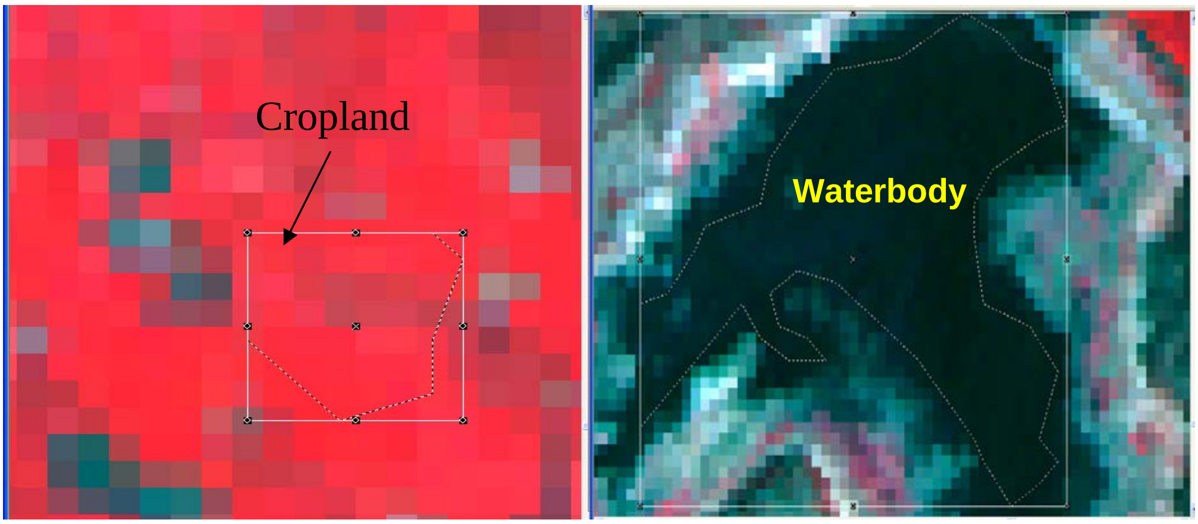

Figure 1.a. Training the red pixels as cropland and block texture as Waterbody

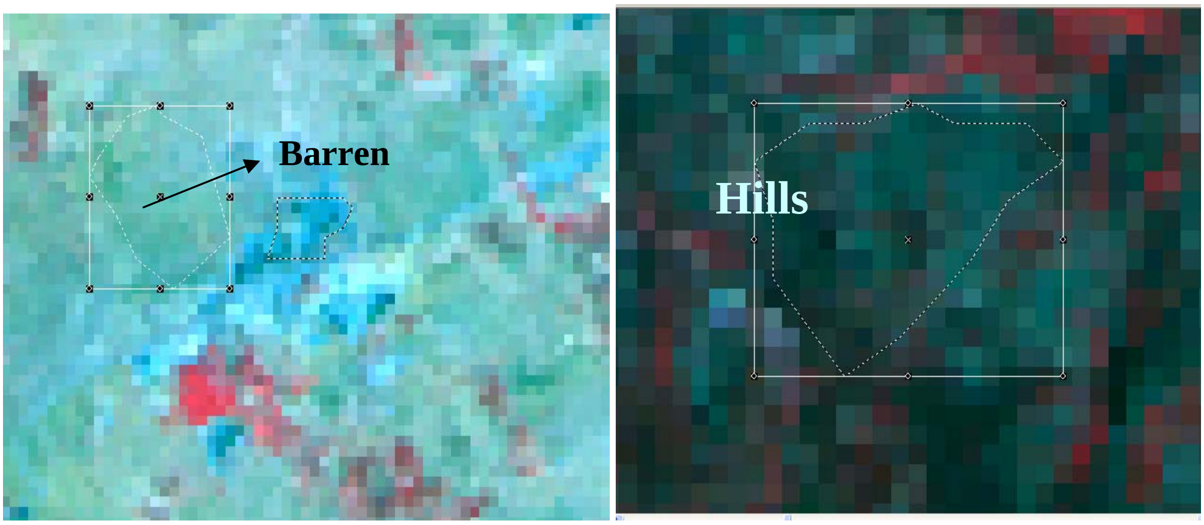

Figure 1.b. Training the pixel as Barren Land and Hills

Step 2:Input the image to process: An image for analysis should be given. As this method deals with the colour values, regardless of the type of the image (Whether it is of normal digital image or satellite image or infra-red or anything etc.) any image could be analyzed.

Step 3: Grab pixels: From the top left corner of the input image, the pixels are grabbed and are used for examination. While grabbing the pixel values, we transform the pixels colour values into the separate channel values and identify its density or the brightness level.

Step 4:Examination Hall: After picking up pixel and splitting up channels, each channel must be compared with the channels of each groups training data. On examining that the pixel value contains the value of training data, we consider the following constraints.

- If each channel contains the exact values as the training data then it must be labeled with the name of the group of the training data.
- If each channel or any of them fails to prove, to be the exact values in the training data then the pixel's values must be compared with the tolerated colour values form the colour of the group.(a tolerated value will be the colour, which is made up of deducting the brightness amount from the class's colour. Deducting from 0 to specified tolerance value).
- If pixel a value does not matches in any of the above circumstances then the pixel is labeled to be a "unknown" pixel and should start to compare again with the next class.
- Examination of a pixel must run until a group is found for a pixel or all the class found unmatched for the pixel.

Step5: Display the results: After all the pixels were examined, we can display the result. With the result, we can easily plot classified map image, according to the supervised classes. All the unknown pixels from this plotting will be the borders of the groups specified in any way and we can easily draw the borders of each group without any fail. The following figure 2 shows the implement of the new algorithm.

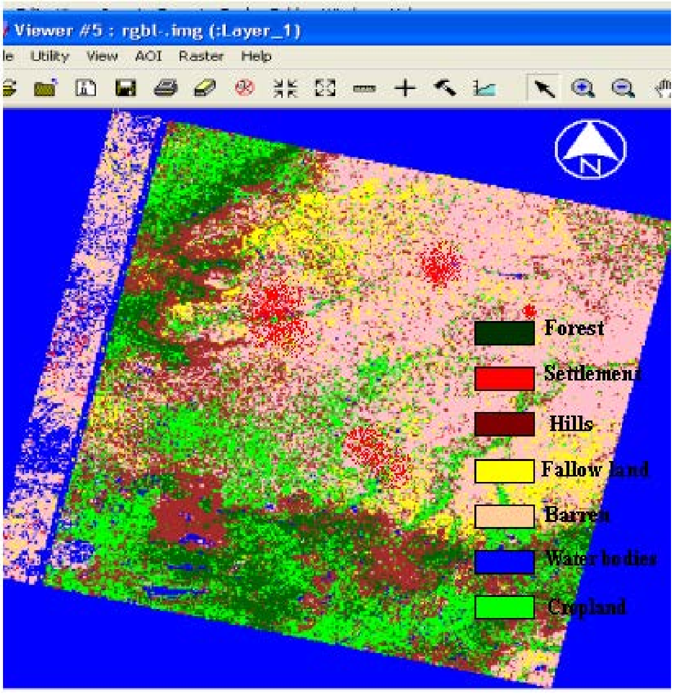

Figure 2. RGB&L Classification

III. EXISTING SUPERVISED CLASSIFICATION ALGORITHM

The three classifier tested with the IRS IA and IRS IC. These imageries projected by UTM and Everest as datum. This is shown in the figure 3. Before classification analyst couldn't find out this is crop, fallow hills and other class.

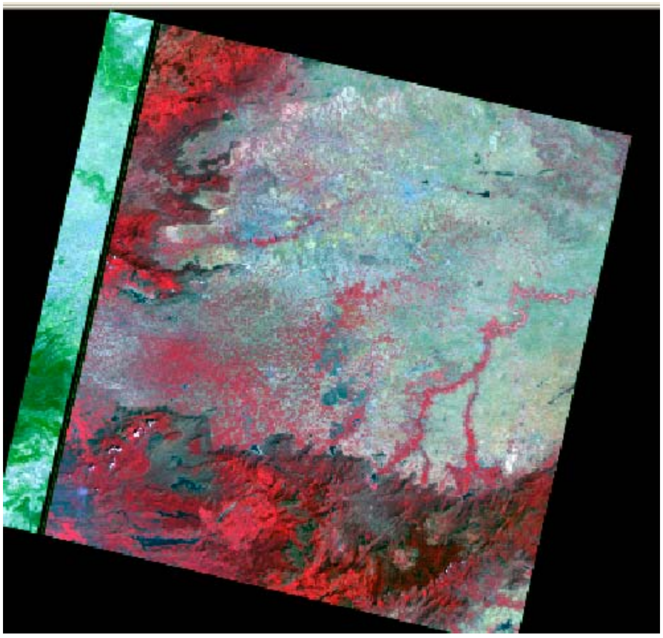

Figure 3. Unclassified Imagery

A. Parallelepiped Classification algorithm

First the original imagery was tested with parallelepiped algorithm. This is a widely used decision rule based on simple Boolean “and/or” logic. Training data in n spectral bands are used in performing the classification. Brightness values from each pixel of the multispectral imagery are used to produce an n-dimensional mean vector, Mc = (μ_{ck} , μ_{c2} , μ_{c3} , ... μ_{cn}) with μ_{ck} being the mean value of the training data obtained for class c in band k out of m possible classes, as previously defined. Sck is the standard deviation of the training data class c of band k out of m possible classes.

The decision boundaries form an n-dimensional parallelepiped in feature space. If the pixel value lies above the lower threshold and below the high threshold for all n bands evaluated, it is assigned to an unclassified category. Although it is only possible to analyze visually up to three dimensions, as described in the section on computer graphic feature analysis, it is possible to create an n-dimensional parallelepiped for classification purposes.

The parallelepiped algorithm is a computationally efficient method of classifying remote sensor data. Unfortunately, because some parallelepipeds overlap, it is possible that an unknown candidate pixel might satisfy the criteria of more than one class. In such cases it is usually assigned to the first class for which it meets all criteria. A more elegant solution is to take this pixel that can be assigned to more than one class and use a minimum distance to means decision rule to assign it to just one class.

The parallelepiped classifier uses the class limits and stored in each class signature to determine if a given pixel falls within the class or not. The class limits specify the dimensions (in standard deviation units) of each side of a parallelepiped surrounding the mean of the class in feature space. If the pixel falls inside the parallelepiped, it is assigned to the class. However, if the pixel falls within more than one class, it is put in the overlap class (code 255). If the pixel does not fall inside any class, it is assigned to the null class (code 0).

The parallelepiped classifier is typically used when speed is required. The draw back is (in many cases) poor accuracy and a large number of pixels classified as ties (or overlap, class 255. The result of this classifier is shown in the following figure 4. The IRS IC imagery projected with UTM projector and Everest was used as a datum. There are totally seven classes were used. Dark green pixel denoted as Forest, red pixel classified as settlement, brown shaded are hills, yellow pixels Fallow land, blue colour grouped as water bodies and light green pixel grouped in to crop land.

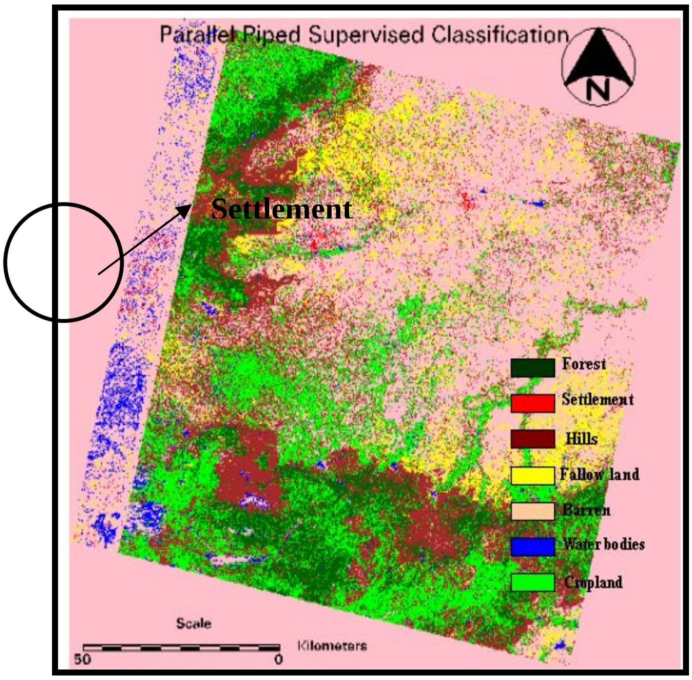

Figure 4. Parallelepiped classifier

B. Minimum Distance to Means Classification Algorithm

This decision rule is computationally simple and commonly used. When used properly it can result in classification accuracy comparable to other more computationally intensive algorithms, such as the maximum likelihood algorithm. Like the parallelepiped algorithm, it requires that the user provide the mean vectors for each class in each hand μ_{ck} from the training data. To perform a minimum distance classification, a program must calculate the distance to each mean vector, μ_{ck} from each unknown pixel (BV ijk). It is possible to calculate this distance using Euclidean distance based on the Pythagorean theorem. The computation of the Euclidean distance from point to the mean of Class-1 measured in band relies on the equation

$$\mathbf{Dist = SQRT\{ (BV_{ijk} - \mu_{ck}) + (BV_{ijl} - \mu_{cl}) \}}$$

----→ Eq.1

Where μ_{ck} and μ_{cl} represent the mean vectors for class c measured in bands k and l.
Many minimum-distance algorithms let the analyst specify a distance or threshold from the class means beyond which a pixel will not be assigned to a category even though it is nearest to the mean of that category.

Minimum distance classifies image data on a database file using a set of 256 possible class signature segments as specified by signature parameter. Each segment specified in signature, for example, stores signature data pertaining to a particular class. Only the mean vector in each class signature segment is used. Other data, such as standard deviations and covariance matrices, are ignored (though the maximum likelihood classifier uses this). The result of the minimum distance classifier shown in the above figure 5. The same imagery was used to classify.

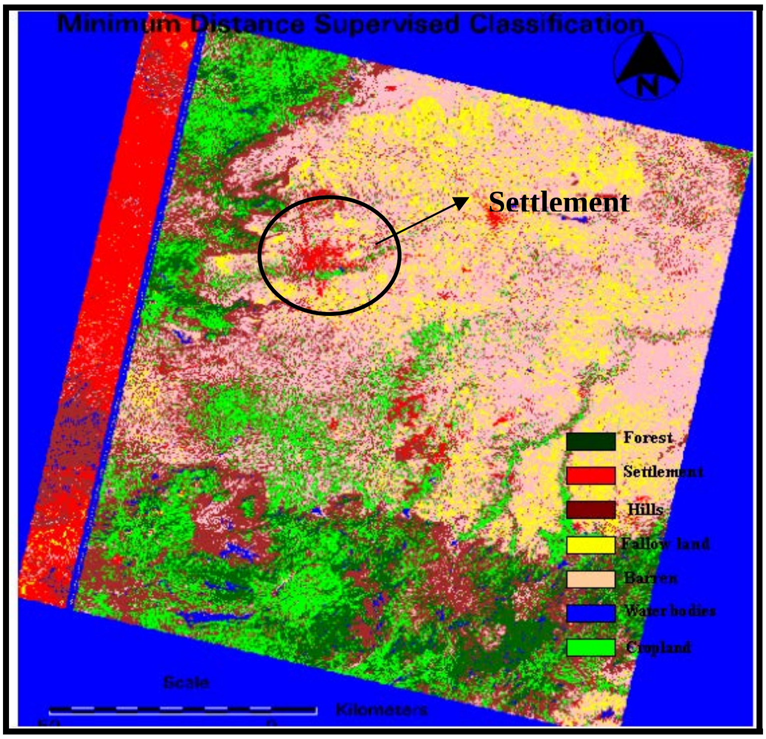

Figure 5. Minimum Distance supervised classification

C.Maximum Likelihood Classification Algorithm

The maximum likelihood decision rule assigns each pixel having pattern measurements or features X to the class c whose units are most probable or likely to have given rise to feature vector x. It assumes that the training data statistics for each class in each band are normally distributed, that is, Gaussian. In other words, training data with bi- or trimodal histograms in a single band are not ideal. In such cases, the individual modes probably represent individual classes that should be trained upon individually and labeled as separate classes. This would then produce unimodal, Gaussian training class statistics that would fulfill the normal distribution requirement.

The Bayes's decision rule is identical to the maximum likelihood decision rule that it does not assume that each class has equal probabilities. A priori probabilities have been used successfully as a way of incorporating the effects of relief and other terrain characteristics in improving classification accuracy. The maximum likelihood and Bayes's classification require many more computations per pixel than either the parallelepiped or minimum-distance classification algorithms. They do not always produce superior results.

Maximum likelihood Classification is a statistical decision criterion to assist in the classification of overlapping signatures; pixels are assigned to the class of highest probability. This is the most accurate of the classifiers of the ERDAS system (if the input samples/clusters have a normal distribution), because it takes the most variables into consideration.

The maximum likelihood classifier is considered to give more accurate results than parallelepiped classification however it is much slower due to extra computations. It is shown in the below figure 6. We put the word `accurate' in quotes because this assumes that classes in the input data have a Gaussian distribution and that signatures were well selected; this is not always a safe assumption. But the maximum likelihood equation is extensive, and takes a long time to compute. The computation time increases with the number of input bands. But compare with these three algorithm RGB& L algorithm gave more accuracy. The accuracy is assessed by using Error matrix. This is explained in the next section. Table 1 shows the Error matrix.

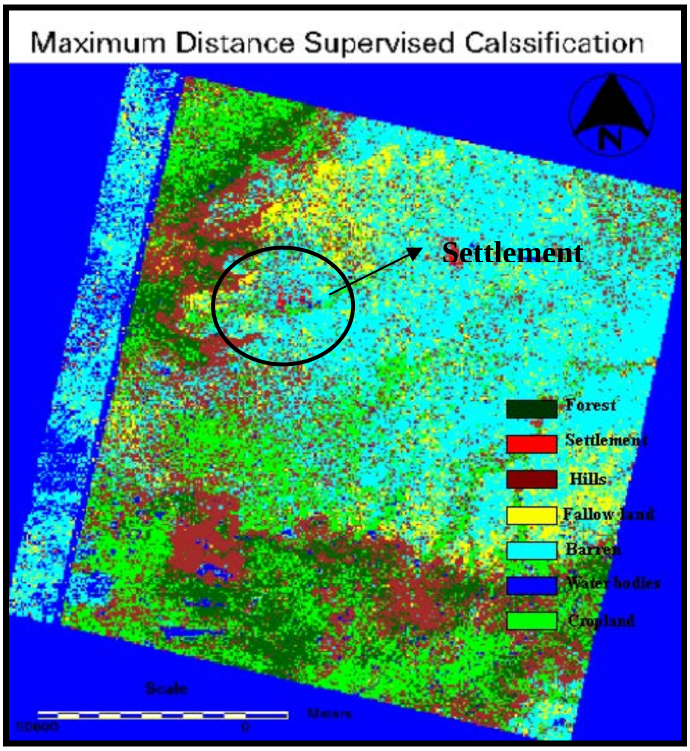

Figure 6. Maximum Distance supervised classification

Analyzing the above classification using three classifiers shown in the figure settlements were increased in some classifier decreased. This mistake happened in all classes such as cropland, fallow land, barren etc. This will make wrong estimate in the area calculation. This was rectified by new algorithm and its accuracy was tested and proved by Error Matrix shown in Table 1.

IV.ACCURACY ASSESSMENT

In this section accuracy of classifications were assessed by two methods Error matrix and visualization comparison.

A.Error Matrix

One of the most familiar means of expressing classification accuracy is the preparation of classification error matrix sometimes called confusion or a contingency table. Error matrices compare on a category-by-category basis, the relationship between known reference data (ground truth) and the corresponding results of an automated classification. Such matrices are square, with the number of rows and columns equal to the number of categories whose classification accuracy is being assessed. Table 1 is an error matrix that an image analyst has prepared to determine how well a Classification has categorized a representative subset of pixels used in the training process of a supervised classification. This matrix stems from classifying the sampled training set pixels and listing the known cover types used for training (columns) versus the Pixels actually classified into each land cover category by the classifier (rows).

Table 1. Error Matrix resulting from classifying training Set pixels

Classes	W	S	F	C	B	FA	H	Row Total
W	480	0	5	0	0	0	0	485
S	0	52	0	0	0	20	0	72
F	0	0	313	0	15	20	15	363
C	0	25	0	252	20	0	38	335
B	0	0	0	70	126	10	0	206
FA	0	0	0	0	2	342	65	407
H	0	0	38	20	43	15	347	463
Column Total	480	77	356	342	206	407	463	2331

W-Water body,S-Settlement,F-Forest, C-Crop,B-Barren,FA-Fallow, H-Hills

Producer's Accuracy	*User's Accuracy*
W=480/480=100%	W=480/485=99%
S=52/77=68%	S=52/72=72%
F=313/356=88%	F=313/363=86%
C=252/342=74%	C=252/335=75%
B=126/206=65%	B=126/206=65%
FA=342/407=84%	FA=342/407=84%
H=347/463=75%	H=347/463=75%

Overall Accuracy= (480+52+313+252+126+342+347)/2331=*83%*

An error matrix expresses several characteristics about classification performance. For example, one can study the various classification errors of omission (exclusion) and commission (inclusion). Note in Table1 the training set pixels that are classified into the proper land cover categories are located along the major diagonal of the error matrix (running from upper left to lower right). All non-diagonal elements of the matrix represent errors of omission or commission. Omission errors correspond to non-diagonal column elements (e.g. 25 pixels that should have classified as “settlement” were omitted from that category). Commission errors are represented by non-diagonal row elements (e.g 2 barren pixels plus 65 hills pixels were improperly included in the corn category).

Several other measures for e.g. the overall accuracy of classification can be computed from the error matrix. It is determined by dividing the total number correctly classified pixels (sum of elements along the major diagonal) by the total number of reference pixels. Likewise, the accuracies of individual categories can be calculated by dividing the number of correctly classified pixels in each category by either the total number of pixels in the corresponding rows or column. Producers accuracy which indicates how well the training sets pixels of a given cover type are classified can be determined by dividing the number of correctly classified pixels in each category by number of training sets used for that category (column total). Users accuracy is computed by dividing the number of correctly classified pixels in each category by the total number of pixels that were classified in that category (row total). This figure is a measure of commission error and indicates the probability that a pixel classified into a given category actually represents that category on ground.

Note that the error matrix in the table indicates an overall accuracy of 83%. However there is no difference between producer’s accuracy ranges users accuracy ranges. This error matrix is based on training data. If the results were good it indicates that the training samples were spectrally separable and the classification works well in the training areas. This aids in the training set refinement process, but indicates little about classifier performance else where in the scene.

V.VISUALIZATION COMPARISONS AND DISCUSSION

The following result was received in the experiment. Figure 7.a., 7.b show the difference between RGB&L supervised and other classification. when compare result of parallel piped and minimum distance with new algorithm gave 83% accuracy. We can observe that the settlement, crop land barren almost same in the user’s result.

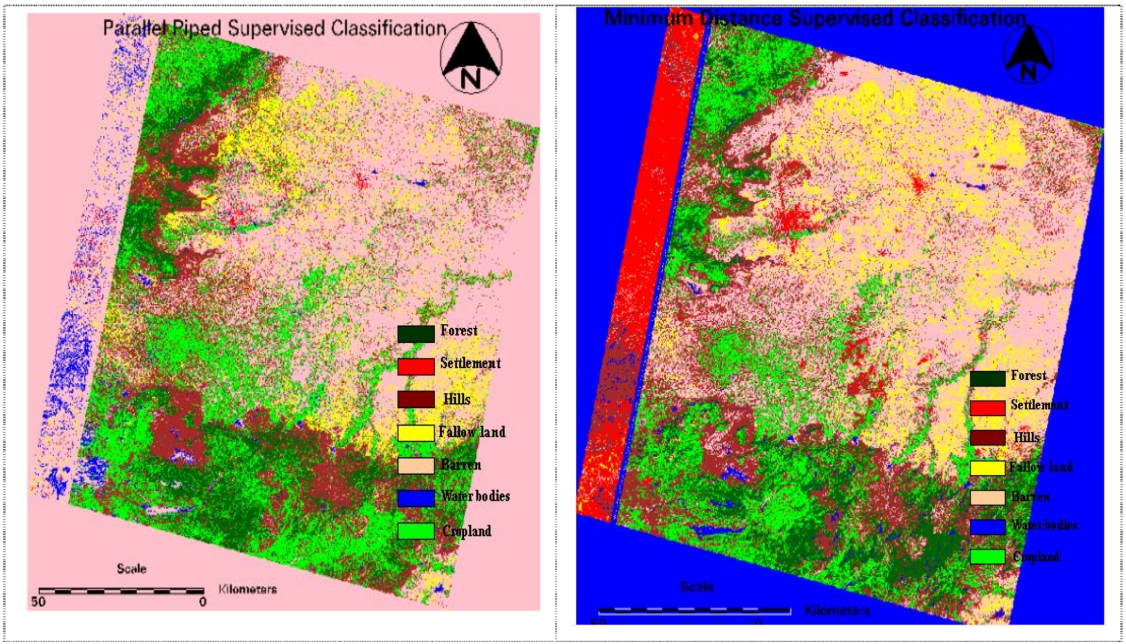

Figure 7.a. Comparison of Parallel Piped and Minimum Distance

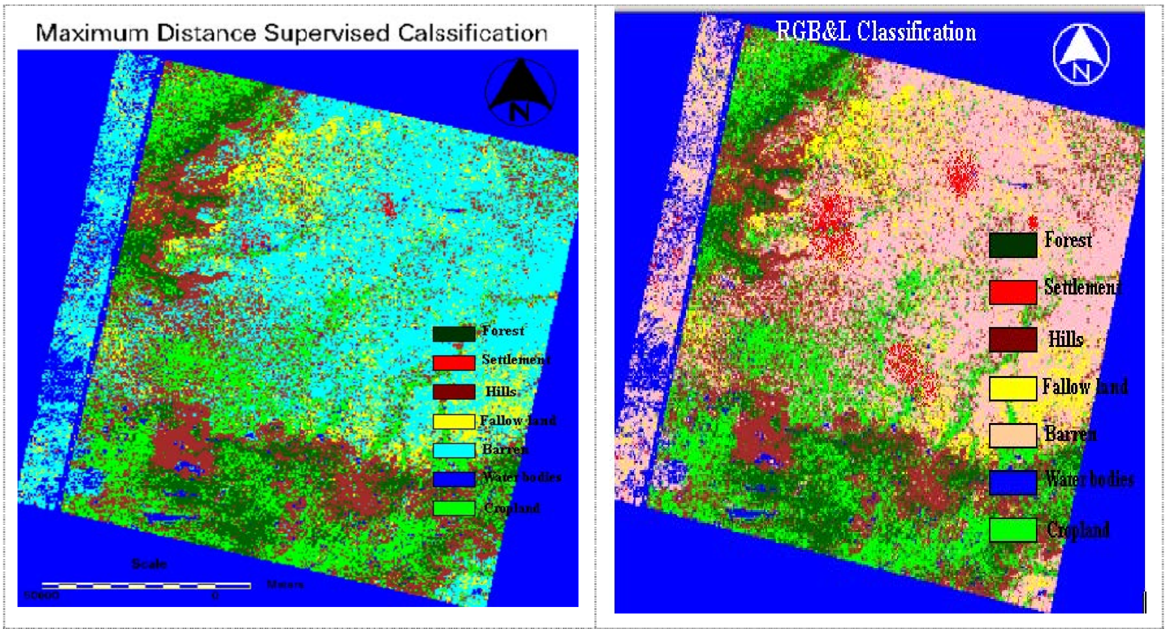

Figure 7.b. Comparison of Maximum Distance

When visually compared with existing method minimum distance and RGB&L appeared same. But we could observe that the variation in classification. Figure 8.a shows the signature and its pixel values of RGB&L and figure 8.b. shows the mean of the class. In the unsupervised classification we can extract minimum features but in the RGB&L supervised classification we define all features of the image. I have done these by using IRS IC List III data for the year 2002.

Class #	>	Signature Name	Color	Red	Green	Blue	Value	Order	Count	Prob.	P	I	H	A	FS
1	>	Water bodies		0.000	0.000	1.000	20	239	7065	1.000	X	X	X	X	
2		FallowLand		1.000	1.000	0.000	17	277	15751	1.000	X	X	X	X	
3		CropLand		0.000	1.000	0.000	36	319	151070	1.000	X	X	X	X	
4		Hills		0.647	0.165	0.165	30	345	168525	1.000	X	X	X	X	
5		Barren Land		1.000	0.753	0.796	40	405	58612	1.000	X	X	X	X	
6		Forest1		0.000	0.392	0.000	3	443	100560	1.000	X	X	X	X	
7		Settlement		1.000	0.000	0.000	5	447	56	1.000	X	X	X	X	

Figure 8.a. Pixel values of supervised

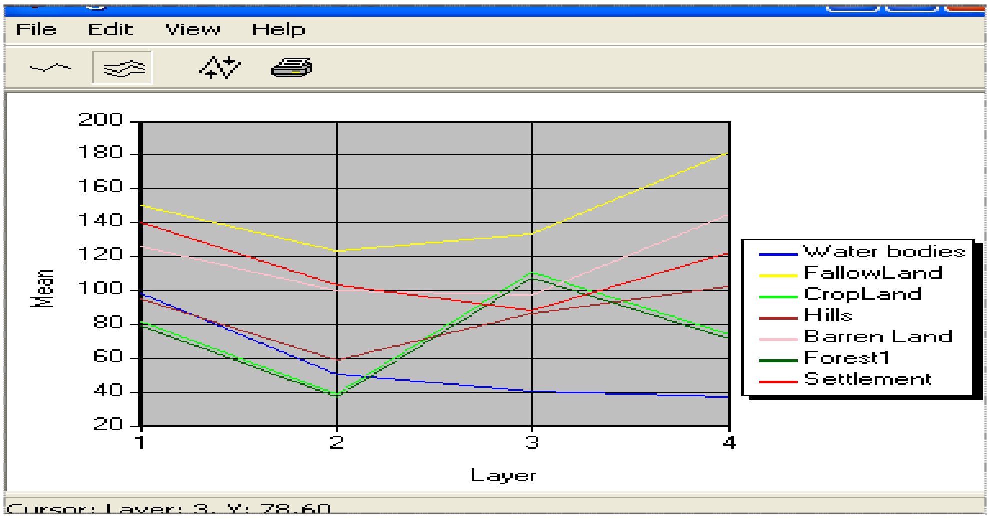

Figure 8.b. Spectral means of the classes in every band Classification

VI. CONCLUSION

The RGB& L supervised classification method gave 83 % of accuracy than the other. The result of proposed algorithm compared with existing method such as nearest neighbor, bilinear and parallel piped line. It observed that the spectral means of the classes in every band was good. If the results were good it indicates that the training samples were spectrally separable and the classification works well in the training areas. This aids in the training set refinement process, but indicates little about classifier performance else where in the scene.

REFERENCES

[1] Butera. M.K., Remote sensing of wetlands, IEEE Transactions on Geoscience and Remote Sensing GE-21,pp. 383-392, 1983

[2] Campbell, Introduction to Remote Sensing, CORINE Land Cover Technical Guide, European Commission Luxemburg, pp. 21-53. 2002.

[3] Dean, A.M , Smith, G.M., An evaluation of per parcel land cover mapping using maximum likelihood class probabilities, International Journal of Remote Sensing, 24 (14), pp. 2905-2920, 2003

[4] Ernst, C.L., Hoffer R.M., Digital processing of remotely sensed data for mapping wetland communities. LARS Technical Report 122079. Laboratory for Applications of Remote Sensing. Purdue University, West Lafayette, 119 pp. 1979.

[5] Kiyonari Fukue, Haruhisa Shimoda, Yoshiaki, Evaluations of unsupervised methods for land-cover/use classifications of landsat TM data,Geocarto International, Volume 3, Issue 2, Pages 37 – 44, 1988

[6] Lu ab, D. Batistella, C. M. Moran ab, E., Land-cover classification in the Brazilian Amazon with the integration of Landsat ETM+ and Radarsat data; International Journal of Remote Sensing. Vol.28, Issue. 24, pp. 5447 – 5459,2007

[7] Lillesand, T.M., Kiefer R., Remote Sensing Image Interpretation. John Wiley, NewYork. 1993.

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

[9] Lambin E.F.,A.H. Strahler.,Change vector analysis in multitemporal space: a tool to detect and categorize land cover change processes using high temporal resolution satellite data. Remote Sensing of the Environment, 48:231-244. 1994.

[10] Story, M , R. Congalton, Accuracy assessment: a user’s perspective. Photogrammetric Engineering and Remote Sensing, 52(3):397-399. 1986

BIBLIOGRAPHY

Mrs. M.Renuka Devi , has nearly 10 years of post graduate teaching experience in Computer Science. She has indulged in training the post graduate students to complete real time projects and also guides research scholars in Computer Science. Currently she is working as Asst.Prof in the Department of MCA at Sree Saraswathi Thyagaraja College (Autonomous),and An ISO 9001 Certified / NAAC Accredited Institution, Pollachi, Coimbatore (Dt), Tamil Nadu,India.

Lt.Dr.S.Santhosh Baboo, has around Seventeen years of postgraduate teaching experience in Computer Science, which includes Six years of administrative experience. He is a member, board of studies, in several autonomous colleges, and designs the curriculum of undergraduate and postgraduate programmes. He is a consultant for starting new courses, setting up computer labs, and recruiting lecturers for many colleges. Equipped with a Masters degree in Computer Science and a Doctorate in Computer Science, he is a visiting faculty to IT companies. It is customary to see him at several national/international conferences and training programmes, both as a participant and as a resource person. He has been keenly involved in organizing training programmes for students and faculty members. His good rapport with the IT companies has been instrumental in on/off campus interviews, and has helped the post graduate students to get real time projects. He has also guided many such live projects. Lt.Dr. Santhosh Baboo has authored a commendable