Land Cover Classification of Remote Sensing Images using Texture and Entropy Features

[1]Sridhar E,
Research Scholar, Department of Computer Applications, Bharathiar University, Coimbatore, Tamil Nadu, India

Abstract— **This This Classification of land cover is the basic building block for spatial planning and management. The spectral based methods that are commonly used are likely to fail due to the fact that each land cover type or feature is presented in several adjacent pixels of different spectral value. The features are much larger than the pixel resolution and there is a need for additional information such as shape, size, association, and tonal / textural relations of the features. In the present work, texture pattern variations and entropy have been used as one of the important characteristics for identifying features from high resolution satellite data A methodology has been proposed for the identifying the land cover regions. The major steps involved in the proposed method are (a) Pre-Processing, consisting of histogram equalization (b) extraction of linear binary patterns and entropy as features from the image (c) classification of the land cover regions using K-means Clustering. A part of Pune from the LANDSAT database is used for the experiments. The experimental results show that the quality measures obtained viz., overall classification accuracy and kappa coefficient are promising for the proposed method**

Keywords: **Remote Sensing, Classification, Local Binary Pattern, Image Entropy.**

I. Introduction

Over the years, remotely sensed image data have been increasingly used in a variety of domains and extensively being employed in a diversity of Earth surface, oceanographic and atmospheric applications such as environmental modeling and monitoring, updating of geographical databases and land cover mapping. Appropriate classification of land land cover is the most crucial factor in planning, utilization, management and monitoring of resources for sustainable development. Analysis of remote sensing images and extraction of digital information can provide faster and more accurate results. It will provide enhanced interpretation approaches, leading to increased efficiency in the processes of automation and offers reduced time in field data gathering, costs and update frequency. This present work intends to utilize a combination of features viz., local binary patterns and entropy of the remote sensing image for the land cover classification. The main objectives of this work are 1) to perform an extensive review and comparison of various existing methods for land cover classification 2) propose an improved land cover classification based on a combination of local features viz., local binary pattern and entropy 3) design and implement the proposed land cover classification method 4) evaluate the proposed land cover classification method using LANDSAT image. The major steps involved in the proposed method are (a) Pre-Processing, consisting of histogram equalization (b) extraction of linear binary patterns and entropy as features from the image (c) classification of the land cover regions using K-means Clustering.

A part of Pune from the LANDSAT database is used for the experiments. The experimental results show that the quality measures obtained viz., overall classification accuracy; expected classification accuracy and kappa coefficient are promising for the proposed method. Hence, it is demonstrated that high accuracy can be achieved land cover classification of remote sensing image using combined texture and entropy-based features.

II. LITERATURE SURVEY

A novel semi supervised SVM model that uses self- training approach is proposed in [1] to address the problem of remote sensing land cover classification. The key characteristics of this approach are that (1) the self-adaptive mutation particles warm optimization algorithm is introduced to get the optimum parameters that improve the generalization performance of the SVM classifier, and (2) the Gustafson–Kessel fuzzy clustering algorithm is proposed for the selection of unlabeled points to reduce the impact of ineffective labels. The effectiveness of this proposed technique is evaluated firstly with samples from remote sensing data and then by identifying different land cover regions in the remote sensing imagery. Experimental results show that accuracy level is increased by applying this learning scheme, which results in the smallest generalization error compared with the other schemes. Artificial neural networks and support vector machines can be used for remotely sensed image classification applications. The major drawback in applying these models is that the user cannot readily realize the final rules. A rule-based classifier derived from improved genetic algorithm (GA) approach is proposed in [2] to determine the knowledge rules for land-cover classification done automatically from remote sensing image datasets. This algorithm is demonstrated for two image datasets classification problems. Results are compared to other approaches in the literature. The preliminary results indicate that the GA rule-based approach for land-cover classification is promising. The Structural Neural Network method was proposed in [3] to land cover classification in remote sensing images. The purpose of this approach is to give a criterion for network architecture definition and to allow the interpretation of the" network behavior". The first result aims to avoid a cumbersome trail-and-error process; the later one can be used to obtain information about the relevance of sensors and related bands to land cover classification. The architecture of structured network was tailored and transformed into" simplified networks" which allows one to evaluate the relevance of sensors and related bands. The experimental results on a multi sensor data set of an agricultural area were reported. The method has been applied on another multispectral image of a forest area. The Bayesian classifier is also used in the proposed method to provide effective result. A new Binary Partition Tree (BPT) method is proposed in [4] along with an unsupervised evaluation of image segmentations by energy minimization. For building extraction, they had applied fuzzy sets to create a fuzzy landscape of shadows which in turn involves a two-step procedure. The first step is a preliminarily image classification at a fine segmentation level to generate vegetation and shadow information. The second step models the directional relationship between building and shadow objects to extract building information at the optimal segmentation level. The results show that this method of classification produced the highest overall accuracies and kappa coefficients, and the smallest over-classification and under-classification geometric errors. This method which integrates BPT with energy minimization offers an effective means for image segmentation. Secondly, the study suggests that the directional relationship between building and shadow objects represented by a fuzzy landscape is important for building extraction. A more effective supervised classification algorithm was proposed in [5] on remote sensing satellite image that uses the average fuzzy intracluster distance within the Bayesian algorithm. The suggested algorithm establishes the initial cluster centers by selecting training samples from each category. It executes the extended fuzzy c-means which calculates the average fuzzy intracluster distance for each cluster. The membership value is updated by the average intracluster distance and all the pixels are classified. The average intracluster distance is the average value of the distance from each data to its corresponding cluster center, and is proportional to the size and density of the cluster.

III. PROPOSED METHOD

This section describes the methodology of the proposed land cover classification system for remote sensing images using the color and texture features. It also describes the implementation of the proposed land cover classification-based system in MATLAB. Section 3.1 gives an overview of the proposed land cover classification system, Section 3.2 describes the implementation in MATLAB and Section 3.3 describes classification accuracy.

3.1. Overview of Proposed Land Cover Classification System

Figure 1 gives the block diagram of the proposed land cover classification method. The three steps involved in this method are 1) image preprocessing 2) feature extraction and 3) classification of the land cover regions. These steps are described in detail in the following sections.

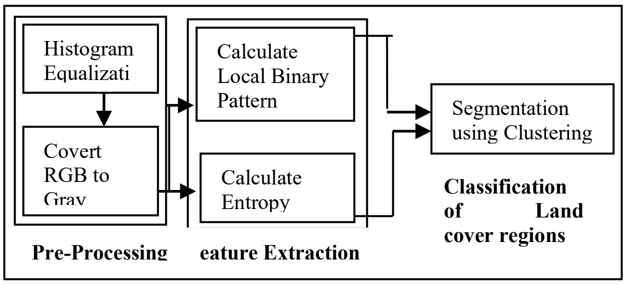


Fig 1. Overview of proposed land cover classification system

3.1.1 Image Pre-Processing

Prior to data analysis, initial processing on the raw data is usually carried out to correct for any distortion due to the characteristics of the imaging system and imaging conditions. Depending on the user's requirement, some standard correction procedures may be carried out by the ground station operators before the data is delivered to the end-user. These procedures include radiometric correction to correct for uneven sensor response over the whole image and geometric correction to correct for geometric distortion due to Earth's rotation and other imaging conditions such as oblique viewing.

A color histogram is a representation of the distribution of colors in an image. A color histogram of an image represents the distribution of the composition of colors in the image The color histogram can be built for any kind of color space, although the term is more often used for three-dimensional spaces like RGB or HSV. If the set of possible color values is sufficiently small, each of those colors may be placed on a range by itself; then the histogram is merely the count of pixels that have each possible color. Histogram equalization is a technique for adjusting image intensities to enhance contrast [6]. Let f be a given image represented as a mr by mc matrix of integer pixel intensities ranging from 0 to L − 1. L is the number of possible intensity values, often 256. Let p denote the normalized histogram of f with a bin for each possible intensity. So

$$P_n = \frac{\text{Number of pixels with intensity n}}{\text{Total number of pixels}} \qquad n = 0, 1, 2, ...L - 1 \tag{1}$$

3.1.2 Local Binary Pattern

The original LBP operator labels the pixels of an image with decimal numbers, called Local Binary Patterns or LBP codes, which encode the local structure around each pixel [8]. It proceeds thus, as illustrated in Fig.4.3: Each pixel is compared with its eight neighbors in a 3x3 neighborhood by subtracting the center pixel value; The resulting strictly negative values are encoded with 0 and the others with 1; A binary number is obtained by concatenating all these binary codes in a clockwise direction starting from the top-left one and its corresponding decimal value is used for labeling. The derived binary numbers are referred to as Local Binary Patterns or LBP codes. Formally, given a pixel at (*xc*, *yc*), the resulting LBP can be expressed in decimal form as:

$$LBP_{P,R}(X_c, Y_C) = \sum_{P=0}^{P-1} S(i_p - i_c) 2^P \tag{2}$$

3.1.3 Image Entropy

Image entropy is a quantity which is used to describe the `business' of an image, i.e. the amount of information which must be coded for by a compression algorithm. Low entropy images, such as those containing a lot of black sky, have very little contrast and large runs of pixels with the same or similar values. An image that is perfectly flat will have an entropy of zero. Consequently, they can be compressed to a relatively small size. On the other hand, high entropy images such as an image of heavily cratered areas on the moon have a great deal of contrast from one pixel to the next and consequently cannot be compressed as much as low entropy images.

The formula for calculating entropy as follows

$$\text{Entropy} = -\sum_i P_i \text{ Log}_2 P_j \tag{3}$$

In the equation 9, P i is the probability that the difference between 2 adjacent pixels is equal to i, and log 2 is the base 2 logarithm.

3.1.4 K-means Clustering

K-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining [9]. K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells.

Algorithmic steps for k-means clustering

Let $X = \{x_1,x_2,x_3,\ldots\ldots,x_n\}$ be the set of data points and $V = \{v_1,v_2,\ldots\ldots,v_c\}$ be the set of centers.

1. Randomly select *'c'* cluster centers.
2. Calculate the distance between each data point and cluster centers.
3. Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.
4. Recalculate the new cluster center using above equation:

$$V_i = (1/C_i)\sum_{j=1}^{C_i} X_i \quad (4)$$

where, *'c_i'* represents the number of data points in i^{th} cluster.

5. Recalculate the distance between each data point and new obtained cluster centers.
6. If no data point was reassigned then stop, otherwise repeat from step 3.

3.2 Implementation

This proposed land cover classification implementation work is carried out using Matlab. The figure 2 shows the execution of the main module of the proposed method.

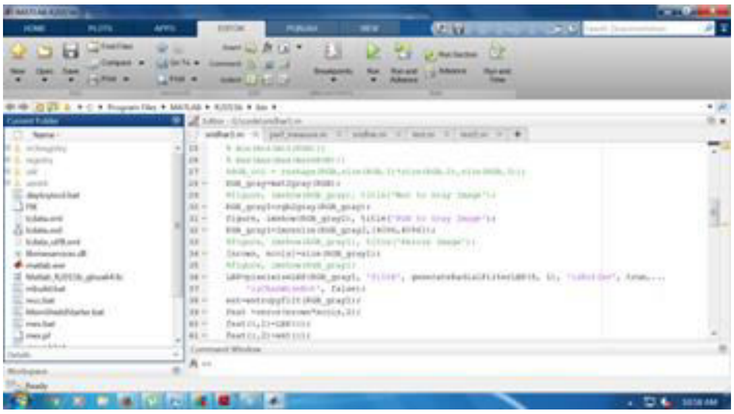


Fig 2. Implementation using Matlab

3.2.1. Classification accuracy

Figure 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. Acknowledgment *(Heading 5)*

Overall Classification Accuracy

Total accuracy is simply the sum of true positive and true negatives, divided by the total number of items, that is:

$$P_0 = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

3.2.2 Kappa Coefficient

Cohen's kappa coefficient is a statistic which measures inter-rater agreement for qualitative or categorical items. It is generally thought to be a more robust measure than simple percent agreement calculation, since κ takes into account the possibility of the agreement occurring by chance. Cohen's kappa measures the agreement between two raters which classify *N* items into *C* mutually exclusive categories.

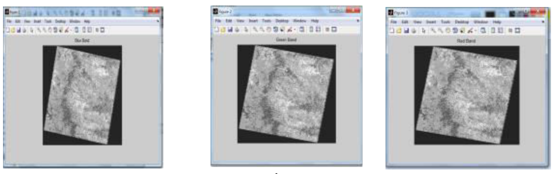
The equation for

$$K = \frac{p_0 - p_e}{1 - p_e} = 1 - \frac{1 - p_0}{1 - p_e} \quad (6)$$

where p_0 is the overall classification accuracy, and p_e is the expected classification accuracy. If the raters are in complete agreement then $\kappa = 1$. If there is no agreement among the raters other than what would be expected by chance (as given by p_e), $\kappa \leq 0$.

IV. RESULTS

After In Figure 3 there are 3 multispectral bands displayed namely blue, green and red bands. The blue band represents a deep water area, green band represents a forestry and agriculture and the red band represents a manmade objects. These bands are read using histogram equalization function. Because the normal imread function of MATLAB cannot properly read big graphics file, histogram equalization is used to read geo graphical image data



a b c

Fig 3. (a) Blue Band, (b) Green Band, (c) Red BanTABLE 5.1 BAND INFORMATION

In Figure 3 there are 3 multispectral bands displayed namely blue, green and red bands. The blue band represents a deep water area, green band represents a forestry and agriculture and the red band represents a manmade objects. These bands are read using histogram equalization function. Because the normal imread function of MATLAB cannot properly read big graphics file, histogram equalization is used to read geo graphical image data.

The Figure 3 (a) shows a RGB color composite image. The composite image is a combination of red, green and blue bands. The bands are combined by the descending order of band 3 (red band), Band 2 (green band) and band 1(blue band). These bands are concatenated along the dimensions. Figure 3 (b) is a Conversion of RGB to True Color Composite Image. The three bands viz., red, green and blue can be combined together to for true color composite image. The colors of the resulting color composite image resembles closely to the colors observed by the human eyes

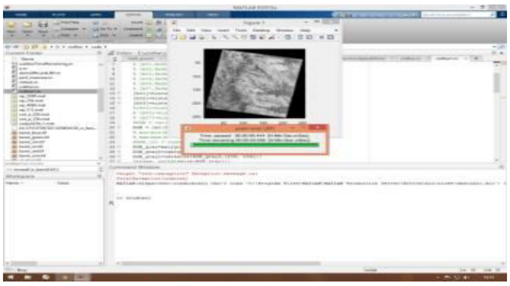
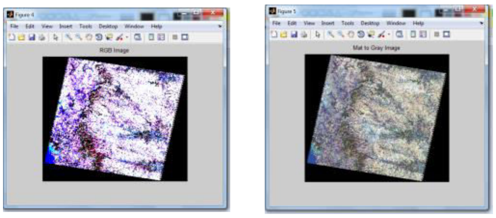


Fig 4. Pixel wise LBP process



a b

Fig 5. (a) RGB Image, (b) True Color image

The figure 4 shows the local binary pattern process for given input image. The grayscale image is the input for LBP operation. Each pixel is compared with its eight neighbors in a 3x3 neighborhood by subtracting the center pixel value; The resulting strictly negative values are encoded with 0 and the others with 1. A binary number is obtained by concatenating all these binary codes in a clockwise direction starting from the top-left one and its corresponding decimal value is used for labeling. The land cover regions are clustered using K-means clustering of the local binary pattern and entropy features. The number of land cover regions considered is 6. Four different sizes of the input image is considered viz., 256 X 256, 512 X 512, 2048 X 2048 and096 X 4096. The proposed method considers the local binary pattern and entropy features for land cover classification. The results of the proposed method are compared with the results of existing method which uses only the intensity values for land cover classification

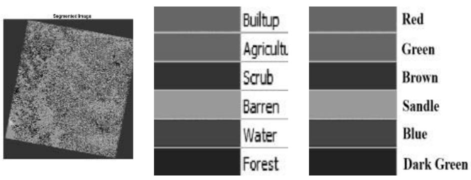


Fig 6. Classified image with Land cover regions and its identical class Informa

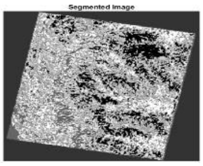
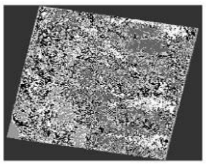
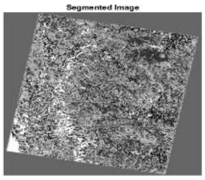
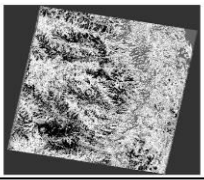
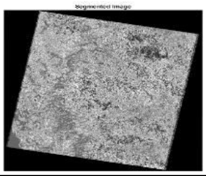
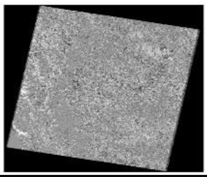
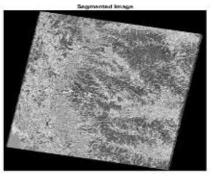
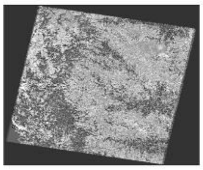
Image Size	Existing Method	Existing Method
Image Size 256 × 256		
Image Size 512x512		
Image Size 2048 x 2048		
Image Size 4096 x 4096		

Fig 7. Classified land cover regions for Proposed method and Existing method using different size of images

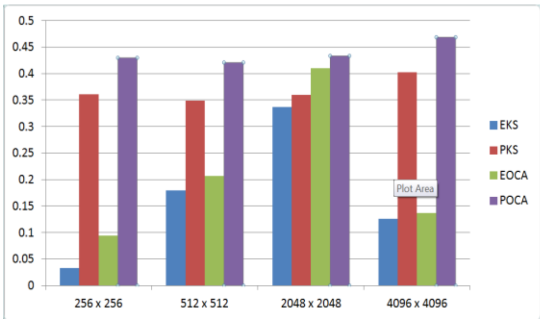


Fig 8. Kappa Statistic and Overall Classification Accuracy of Existing and Proposed Methods using Different Image Size

The Figure 6 shown the Classified land cover regions for Proposed method and Existing method using different size of images. The Figure 7 show the existing and proposed method results using different sized images. The Overall Classification accuracy obtained from the proposed method for input image of size 256 X 256 is 0.4298. Kappa coefficient obtained from the proposed method is 0.3607 and from the existing method is 0.0333. The Overall Classification accuracy obtained from the proposed land cover classification method for the input image of size 512 X 512 is 0.4208. Proposed land cover classification method gives the kappa coefficient as 0.3485 and 0.1795 is the kappa coefficient obtained for existing method. The Overall Classification accuracy of proposed land cover classification method for input image of size 2048 X 2048 is 0.4335 and the Overall Classification accuracy of existing land cover classification method for input image of size2048 X 2048 is 0.4105. Proposed land cover classification method gives kappa coefficient as 0.3595 and 0.3363 is the kappa coefficient for existing method. The Overall Classification accuracy of proposed land cover classification for input image of size 4096x4096 is 0.4686 and the Overall Classification accuracy of existing land cover classification method for input image of size 4096 x 4096 is 0.1362. The Overall Classification accuracy of proposed land cover classification for input image of size 4096x4096 is 0.4686 and the Overall Classification accuracy of existing land cover classification method for input image of size 4096x4096 is 0.1362. These results are plotted as graph in figure 8. It can be observed that the proposed method is better compared to existing method in terms of classification accuracy and kappa coefficient.

V. CONCLUSION

High resolution satellite data split the real-world features in several pixels and thereby limiting the use of classification methods in extracting features. Therefore, various urban land cover categories are more densely packed in small areas than the rural land uses. Feature delineation from high resolution data for an urban area is a challenging task. Thus, information such as size, shape, texture and association are often used for the classification of high-resolution satellite data from LANDSAT images. The main aim of the present work is to make use of texture pattern variations and entropy as features from the very high-resolution satellite data to classify the land cover regions. The study area comprised of a smaller part of Pune obtained from the LANDSAT database. It contains a number of land cover classes at various levels. Six land cover classes viz., builtup area, agricultural, scrub, forestry, barren area, water and forestry were considered for the experiments in this research work. Seven bands from the remote sensing image were considered. The experiments conducted on different resolutions viz., 256 × 256, 512 × 512, 2048 × 2048 and 4096 × 4096 have produced better classification accuracy against an existing method based on only the intensity values. Texture features combined with entropy features helped in better classifying the land cover regions.

REFERENCES

[1] Ajdeep singh chauhan, Manpreet singh, “Image contrast enhancement using histogram Equalization,” *International journal of computing & business research,* ISSN (online): 2229 -6166, 2012.

[2] Tuia .D, Muñoz-Marí .J, Kanevski .M, and Camps-Valls .G, “Structured output SVM for remote sensing image classification,” International Journal of Signal Processing Systems, vol. 65, no. 3, pp. 301–310, 2011.

[3] Dhodhi. M. K, Saghri. J, Ahmad. I, and Ul-mustafa. R, “D-ISODATA : A Distributed Algorithm for Unsupervised Classification of Remotely Sensed Data on Network of Workstations,” *Journal of Parallel and Distributed Computing*, vol. 59, no. 2, pp. 280–301, 1999.

[4] Joshi. C, de Leeuw. J and van Duren. L.C, “Remote Sensing and GIS Applications for Mapping and Spatial Modelling of Invasive Species *International journal of Computer and Geosciences*, vol. 2, no. Graph 1, pp. 669–677, 2002.

[5] Kasetkasem .T, Arora .M, and Varshney .P, “Super-resolution land cover mapping using a Markov random field based approach,” *Journal of Remote Sensing Environment*, vol. 96, no. 3–4, pp. 302–314, 2005.

[6] Tooke T. R, Coops N. C, Goodwin N. R, and Voogt J. A, “Extracting urban vegetation characteristics using spectral mixture analysis and decision tree classifications,” *International Journal of Remote Sensing Environment*, vol. 113, no. 2, pp. 398–407, 2009.

[7] Du .P, Chen .Y, " A Novel Remote Sensing Image Classification Scheme Based on Data Fusion, Multiple Features and Ensemble Learning "*Journal of Indian Society of Remote Sensing*, pp 41-213. doi:10.1007/s12524-012-0205-8, 2013.

[8] Zhu .C, Yang .S, "Robust Semi-supervised Kernel-FCM Algorithm Incorporating Local Spatial Information for Remote Sensing Image Classification," *Journal of Indian Society of Remote Sensing*, vol 42: 35.ISSUE 1, pp 35-49, 2014.

[9] Benz U.C, Hofmann .P, Willhauck .G, Lingenfelder .I, and Heynen .M, “Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information,” *International Journal of Photogrammetry and Remote Sensing*, vol. 58, no. 3–4. pp. 239–258, 2004.

AUTHOR(S) BIOGRAPHY

Sridhar.E, Research Scholar, Bharathiar University, Coimbatore, Tamil Nadu, India. He has received his B.C.A (2012), M.C.A (2015) and M.Phil (2017) in Computer Applications from Bharathiar University, Coimbatore. He has 3 years’ experience as Technical assistant at Tamil Nadu Agricultural University and 1 year Experience as Teaching Assistant for under graduate students in the same University. His broad area of research is Remote Sensing Images under Image Processing. He have published his research work previously in IJARBEST journal. He is an YouTube lecturer, Graphics Designer and Tamil Poet writer. He has won “Kavithendral Virudhu” award during 2011,2015 and 2017.