Level II Vegetation Classification of Satellite Image Data using PCA Method

Mayur Kolhe[1], Yogesh Kadam[2], Ninad Chousalkar[3], Sunny Chordiya[4]

Student at Sinhgad College of Engineering, Computer Department, Pune[1-4]

Abstract: Satellite image processing plays an important role to extract useful information about earth. Satellite image of earth can be used in various applications such as detecting land cover, ship detection, forestry, geology, meteorology etc. The traditional surveying method which requires a huge amount of manpower, money and time can be reduced by applying various image processing methods on satellite image. Our project aims in using two levels of classifications of which the first level of classification includes classification of land, water bodies and urban areas using unsupervised approach. The level two classification aims to divide the land data into vegetated and non-vegetated regions using PCA method. The result of above classification would give us the percent ratio of land under vegetation and non-vegetation

Keyword: PCA, Classification system

I. INTRODUCTION

The use of satellite images is rapidly growing nowadays. First satellites for imaging land areas were placed in orbit in 1972. They have various applications in day-to-day life. More and more satellites are being sent to the space and various everyday home appliances are connected via satellite using computer technology. All electronic gadgets such as mobile phones, laptops etc. are connected with the satellites. The use of satellite image processing to extract various useful information from the satellite images. Various predictions related to different scientific fields can be done using satellite images .Various models are created and their analysis can be done. Satellite image processing is a powerful tool in surveying as large amount of area can be surveyed using one image. Land management, monitoring and planning can be efficiently done using satellite image processing systems. For the project we are using Resource Sat 1 which gives IRS P6 – LISS IV as input image. ISRO is the organization of satellite. The resolution of the satellite is 5.8m along with swat width 23.5km * 23.5 km. National Remote Sensing Centre, Hyderabad is the source of the satellite image.

II. RELATED WORK

Various research has been carried out on land cover classification which includes level-1 classification using K-means algorithm which is a very efficient technique and good results are obtained for all datasets [1]. ERDAS and ARC GIS[5].Also classification based on confusion matrix and kappa coefficient is carried out in which around 97% accuracy is obtained but as it requires features to be very accurate or more number to get correct results[3].The level-2 classification which is used to classify the green land into forest and crop is done using the PCA algorithm and very accurate results are found[2].Also level-2 classification is done on the basis of phonological patterns which has drawbacks as it requires very specific patterns[4]

III. METHODOLOGY

The main purpose of the project is to identify, label and percentage coverage of various land cover/ land use pattern namely Urban, Green Fields and Water Bodies from multispectral optical satellite image which helps in identifying nature of terrain, building areas and green lands, also, to categorize different types of vegetation available in green elds.
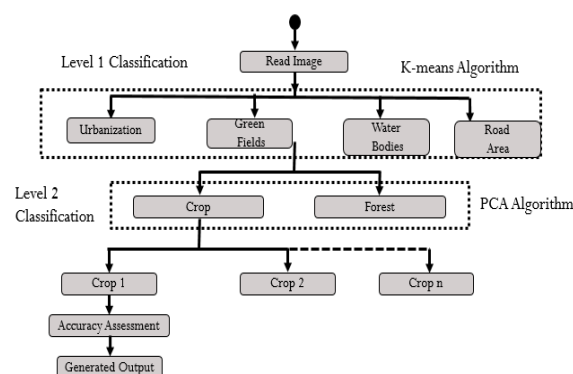

Fig.1 Architecture of Veg. Classification System

The satellite image is accepted as input for the processing. The pixel values of the image are stored in the matrices for further Overview Classification processing. The threshold is maintained in order to categorise the image into various sub-parts. If the pixel value of the image comes under the range of given value of threshold then it is considered for forming clusters otherwise neglected. The group of formed such pixels is collected and shaded as per respective field. Following are the steps that need to be considered for the classification of the image.

A. Step1: Feature extraction

Satellite image is given as an input for the image processing. The pre-processing is done on the image as a first step and proceeded for the L1 classification. It includes feature extraction and removing the unnecessary properties of the satellite image.

The feature extraction also deals with the extraction of all the pixel values of the image given as an input.

The pixel values are stored in the matrix for further processing.

B. Step2: Level-I Classification using K-means Algorithm

Clustering is the classification of objects into different groups, or more precisely, the partitioning of a data set into subsets (clusters), so that the data in each subset (ideally) share some common trait - often according to some defined distance measure.

K-means clustering is a method of classifying/grouping items into k groups (where k is the number of pre-chosen groups). The grouping is done by minimizing the sum of squared distances between items and the corresponding centroid. A centroid is "the centre of mass of a geometric object of uniform density", though here, we'll consider mean vectors as centroids. Clustering is the classification of objects into different groups, or more precisely, the partitioning of a data set into subsets (clusters), so that the data in each subset (ideally) share some common trait - often according to some defined distance measure.

K-means clustering is a method of classifying/grouping items into k groups (where k is the number of pre-chosen groups). The grouping is done by minimizing the sum of squared distances between items and the corresponding centroid. A centroid is "the center of mass of a geometric object of uniform density", though here, we'll consider mean vectors as centroids.

K-means algorithm is useful for undirected knowledge discovery and is relatively simple. K-means has found wide spread usage in lot of fields, ranging from unsupervised learning of neural network, Pattern recognitions, Classification analysis, Artificial intelligence, image processing, machine vision, and many others.

Choosing K:

The algorithm described above finds the clusters and data set labels for a particular pre-chosen K. To find the number of clusters in the data, the user needs to run the K-means clustering algorithm for a range of K values and compare the results. In general, there is no method for determining exact value of K, but an accurate estimate can be obtained using the following techniques.
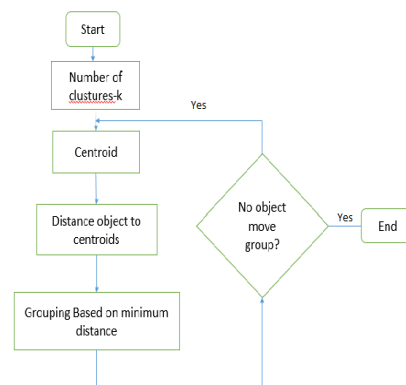

Fig.3 Working of K-means algorithm

One of the metrics that is commonly used to compare results across different values of K is the mean distance between data points and their cluster centroid. Since increasing the number of clusters will always reduce the distance to data points, increasing K will always decrease this metric, to the extreme of reaching zero when K is the same as the number of data points. Thus, this metric cannot be used as the sole target. Instead, mean distance to the centroid as a function of K is plotted and the "elbow point," where the rate of decrease sharply shifts, can be used to roughly determine K.

A number of other techniques exist for validating K, including cross-validation, information criteria, the information theoretic jump method, the silhouette method, and the G-means algorithm. In addition, monitoring the distribution of data points across groups provides insight into how the algorithm is splitting the data for each K.

The processed image is sent for the further classification as pixel values in the matrix using K-means classification. This step will create the numbers of clusters and each cluster is assigned to its closest mean vector.

Then the revised mean clusters are computed. The reclassification will continue till there is no change in the location of class mean vectors. A threshold value is maintained in the order to ensure that the pixels under threshold should belong to the specific field. The threshold is decided by working on different images and taking a mean value for respective field. This step will divide the image into different fields like Green Field, Water Bodies, and Bare Land etc. We will be focusing on the Green Field for further processing.
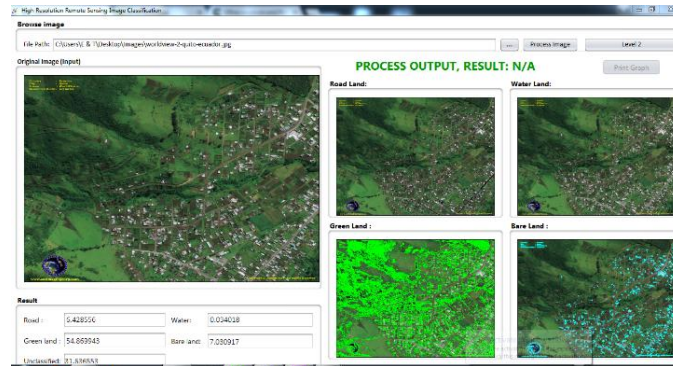

Fig. 2 Output of K-means algorithm

The fig.2 shows that given satellite image is divided into four fields as Green, Water, Bare Land and Road Land. The highlighted green portion indicates that the Green Fields are separated from the input image similarly Dark Blue area formed clusters highlighted indicates the Road area, Black colour indicates the Water area is separated and Sky blue colour indicates the Bare Land present in the image.

The Result field indicates that the percentage of image contains the respective field. The unclassified field for the Result shows that the percent of figure part is not under any of the respective fields.

C. Step3: Level-II Classification   using PCA method:

PCA belongs to linear transforms based on the statistical techniques. PCA takes advantage of eigenvectors properties for determination of selected object orientation. Principal Component Analysis is the general name for a technique which uses sophisticated underlying mathematical principles to transforms a number of possibly correlated variables into a smaller number of variables called principal components. PCA is a statistical method under the broad title of factor analysis. The purpose of PCA is to reduce the large dimensionality of the data space (observed variables) to the smaller intrinsic dimensionality of feature space (independent variables), which are needed to describe the data economically. This is the case when there is a strong correlation between observed variables.

The jobs which PCA can do are prediction, redundancy removal, feature extraction, data compression, etc. Because PCA is a classical technique which can do something in the linear domain, applications having linear models are suitable, such as signal processing, image processing,
 etc. Image processing has many applicable areas.

After the L1 classification the pixel values of Green Field image are used for L2 classification with PCA algorithm. L2 classification deals with classification of Green Fields into further sub fields as Crop Fields and Forest Field. After finding covariance matrix, we apply linear transformation matrix using Eigen values and Eigen vectors on the matrix. In PCA all relevant Information is present in B and I bands. Therefore R and G Band is ignored for Accuracy. In this classification level PCA components Crop and Forest fields are compared with the trained datasets for supervised classification which will classify the Green Fields.
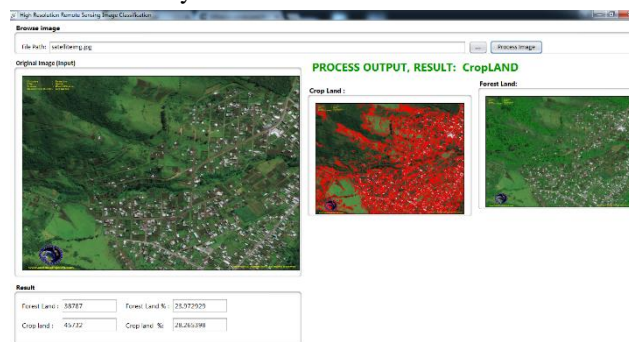

Fig.4 Output of PCA algorithm

The fig.4 shows that the Classified Green Field image in Level 1 Classification is given as an input in this step. The image is processed using PCA algorithm which divides the image in Crop and Forest Field as shown in fig.4. The Dark Green portion in the image indicates the Crop fields and Dark Red colour indicates the area under Forest fields. The result section in the figure indicates the percentage of image is classified using PCA. Similarly the unclassified section indicates that the percentage of image that didn't classified in neither Green nor Crop field.

IV. RESULT

The paper satisfactorily explains the total pixel values and their percentage that are divided into various fields as shown in the table.

L1: Total Numbers of pixels= 164381			
Sr.No.	Class	Nos. of pixels classified	% of total image
1.	Green	88766	54.869943
2.	Water	49	0.034018
3.	Road	10560	6.428556
4.	Urban	11550	7.030917
5.	Unclassified	51972	31.636553

L2: Total Numbers of pixels= 88842			
Sr.No.	Class	Nos. of pixels classified	% of total image
1.	Forest Land	38787	23.972929
2.	Crop Land	45732	28.265398
3.	Unclassified	4323	2.63

Fig. 5 Percentage of classified image

From the above mentioned figure the total numbers of pixels in L1 classification are classified with highest percentage of Green Fields i.e. 54.8% and accordingly Water field with 0.03%, Road area containing 6.42%, urban area containing 7.03%. However 31.6% of image pixel values are not classified because they don't come under any field.
The Green field is further classified into Forest land and Crop land with percentage 23.97 and 28.26 of the total image. The unclassified field is also present in the L2 classification as some pixels don't come under the given fields. Hencefore 2.63% of total Green Field is unclassified.

V. CONCLUSION

Analyzing the classification results of these two methods it can be said that our proposed method is more accurate than the conventional method. The proposed method gives 84 percent on detecting the two diseases whereas the conventional method gives only 82 percent accuracy. This method can also be used for other rice diseases detection and also for other crop diseases too.

ACKNOWLEDGMENT

REFERENCES

[1] Classification of Basmati Rice Grain Variety using Image Processing and Principal Component Analysis, Rubi Kambo, Amit Yerpude International Journal of Computer Trends and Technology (IJCTT) – volume 11 number 2 – May 2014 ISSN: 2231-5381 http://www.ijcttjournal.org
[2] Vegetation Extraction from Multi Sensor IRS-1C LISS III Image by Principle Component analysis and Comparative Assessment, Rubina Parveen, Subhash Kulkarni, V.D. Mytri International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB16).
[3] Image Processing using Principle Component Analysis, Pramod Kumar Pandey, Yaduvir Singh, Sweta Tripathi InternationalJournalofComputerApplications(0975–8887)Volume15–No.4,February 2011.
[4] Classification of Multispectral Satellite Images, Shivali A. Kar, and Vishakha V. Kelkar ICATE 2013 Paper Identification Number-115.
[5] Impact of Principal Component Analysis in the Application of Image Processing, Abhishek Banerjee Information Technology Department, Pailan College Of Management Of Technology ,Pailan, Joka, Kolkata-700104, West Bengal, India.

[6]   Forest Classification Using IRS Satellite Data Somaye Zangiabadi, Ali Ahmadimoghadam and Farzin Naseri World Applied Sciences Journal 15 (10): 1409-1413, 2011 ISSN 1818-4952 © IDOSI Publications, 2011

[7]   PRELIMINARY RESULTS OF THE METASENSING FULLY POLARIMETRIC LBAND AIRBORNE SAR FOR VEGETATION CLASSIFICATION L. Corucci1, A. Meta1, A. Coccia1 MetaSensingBV,Huygensstraat,44,2201DK,Noordwijk,TheNetherlandslinda.corucci; adriano.meta; alex.coccia@metasensing.com

[8]   Study on Classification for Vegetation Spectral Feature Extraction Method Based on Decision Tree Algorithm Weiwei Li and Jian Du Dept. of Computer Science, Huazhong Normal University, Wuhan, P. R. China BaolinYiDept.ofNationalDigitalLearningEngineeringResearchCenter,HuazhongNormal University, Wuhan, P. R. China

[9]   Urban Vegetation Classification Based on Phenology using HJ-IA/B time series imagery Feng Li, Zhu Liujun, Liu Han, Huang Yinyou, Du Peijun, Ebenezer Adaku Key Laboratory for Satellite Mapping Technology and Applications of State Administration of Surveying, Mapping and Geoinformation of China, Nanjing University Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Nanjing University ermaI014@126.com Ghana Institute of Management and Public Administration (GIMPA), P. O. Box AH SO, Achimota, Accra, Ghana

[10]   Urban Vegetation Classification Based on Phenology using HJ-IA/B time series imagery Feng Li, Zhu Liujun, Liu Han, Huang Yinyou, Du Peijun, Ebenezer Adaku Key Laboratory for Satellite Mapping Technology and Applications of State Administration of Surveying, Mapping and Geoinformation of China, Nanjing University Jiangsu Provincial Key Laboratory of Geographic Information Science and Technology, Nanjing University ermaI014@126.com Ghana Institute of Management and Public Administration (GIMPA), P. O. Box AH SO, Achimota, Accra, Ghana

[11]   IMAGE CLASSIFICATION AND PERFORMANCE EVALUATION OF IFS 1C LISS I11 DATA I. V. Muralikrishna, Member IEEE Center for Remote Sensing, JNT University, Hyderabad - 500 028 A. P. INDIA

[12]   Unsupervised Land-Use Classification of Multispectral Satellite Images. A Comparison of Conventional and Fuzzy-Logic Based Clustering Algorithms. Tanja Duda and Morton Canty Research Centre Jiilich, Programme Group Technology Assessment D-52425 Jiilich, Germany