International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

Tree Identification Using K-Mean Clustering Algorithm

Archana Singh[1], Anuj Khuttan[2]

[1, 2]School of Computer Science and Engineering, Galgotias University, Greater Noida, U.P., India

Abstract: ***Trees occupy an important place in the life of man. Tree plays an important role in purifying air around us. Trees are also very necessary for having good rainfall. So that to classify the tree type is very important for the forest maintenance. With the beginning of high spatial resolution remote sensing sensors, our capacity has greatly improved for tree type identification. Considering the amount of data in need of processing and the high computational costs required by image processing algorithms, predictable computing environments are simply not practical. Therefore, it is necessary to develop techniques and models for resourcefully processing large volume of remote sensing images. In this study, a cluster computing environment was adopted to speed up the calculation time. The test image was first partitioned into hundreds of manageable sub-images. Use of K-mean Clustering algorithm we identifying trees.***

Keywords: Image Processing Algorithm, Computing environments, K-mean Clustering, Remote Sensing Sensors

1. Introduction

Details of forest can be obtained by field investigation; however it is almost impossible to constantly measure large area, especially on mountain area where field investigations are usually costly and laborious. On the other hand, remote sensing is helpful for investigation of wide area of forest. In this study, identification of forest types in a high spatial resolution image is demonstrated with the revolutionizing advents in sensor technology, [2] high spatial and/ or spectral. Resolution remote sensing images are increasingly produced.

For example, high-resolution images present a new challenge over, in that a huge amount of data must be analyzed for investigation of wide area of interest [7]. Considering the amount of data in need of processing and the high computational costs required by image processing algorithms, conventional computing environments are simply impractical. Therefore, it is necessary to develop techniques and models for efficiently processing large volume of remote sensing images. High performance computing techniques such as cluster computing, grid computing, and parallel algorithm on a multi-core processor are very good answers to increase the computational performance [6]. Cluster computing provides an advanced computing and sharing model to solve large and computationally intensive problems [6]. In contrast to grid computing, the nodes on a cluster are homogeneous and are networked in a tightly-coupled fashion. Generally, the nodes are configured identically and are all on the same subnet of the same domain. Only the cluster application run on a cluster node, so each node on a cluster is a dedicated resource. It is very difficult to identify a spatially small target such as tree in coarse spatial resolution imagery because the spectral signature of each pixel is a mixture of spectral signatures of targets present in the image. High spatial resolution imaging provides more information than coarse resolution imagery for detailed observation. Hence, with the advent of high spatial resolution remote sensing sensors, our ability has greatly increased for spatially tree type identification [7-9]. A small target in a high spatial resolution image could still occupy more than one pixel. Sometimes, spectral image analysis methods fail to produce satisfying results because objects with similar spectral signatures cannot be spectrally separated. On the other hand, in a high spatial resolution image the spectral features vary a lot within each class also making it difficult to be identified. As a result, information captured in neighboring pixels may provide useful supplementary information to distinguish those spectrally similar objects. Therefore, integration of information from spectral and spatial domain presents the potential for increased classification performance for remote sensing image classification. Komura and Muramoto [7] developed a method that considering shapes and sizes of tree crown calculated from high spatial resolution satellite image for forest stand classifications. By digitizing of crown shapes, Meyer *et al.* [8] presented a semi-automated procedure for tree type identification in high spatial resolution data from Digitized color infrared-aerial photography. By digitizing of crown shapes, Yu *et al.* [9] proposed an object-based Classification approach for detailed vegetation classification with airborne high spatial resolution remote sensing imagery. By captivating the advantage of the rich amount of local spatial information present in an image, the method overcame the problem of salt-and-pepper effect found in classification results from traditional pixel-based approach.

2. K-Mean Clustering Algorithm

In class, we discussed how images could be segmented by grouping the pixels into clusters. In this problem, you will use the k-means algorithm to group pixels into segments.

1. Implement a function im2featurevec that takes an image with NP pixels and returns a matrix (2D array) with NP rows and M columns per row. Each of the M columns will denote a single feature response. Initially, there will be five features: R, G, B, x, and y [3].
2) Implement the k-means algorithm. Assuming that there are NC clusters, the basic steps of the algorithm are:
3) Initialize the algorithm by either initializing the cluster assignments or the cluster locations
4) Until convergence criterion is satisfied:
   - Assign each point to the nearest cluster center

International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

- Ensure that each center has at least one point assigned to it. One possible heuristic for filling in empty clusters is to choose a random point that is far from the cluster center.
- Set each cluster center to the average of the points assigned to it. You will average each dimension of the vectors separately.

5) The input to your function should be an array of points and the number of cluster centers. The output should be the cluster to which each point is assigned and the cluster centers.
6) Write a function that takes an image and returns an image where each pixel contains the cluster number assigned to that cluster. Write a second function that takes this image and displays the boundaries. A simply method for finding the boundaries is to filter the image with vertical and horizontal derivative filters, compute the gradient magnitude, then threshold that value.
7) Test your function on the tree images provided.
8) Problem 1.1 does not ask you to normalize the feature vectors. To see the effect of normalizing the vectors, write a new function that creates the feature matrix as before, but now subtracts the mean from each column, so that each column has a zero mean. In addition, the variance of each column should be divided by a constant so that the quantity:

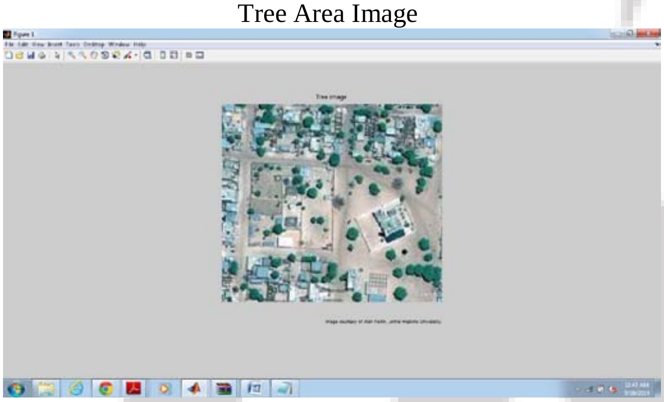
$$\frac{1}{N_P}\sum_{n=1}^{N_P}(F_n - \mu)^2$$

is equal to 1. The term $F_n$denotes the feature at row n in a column. Compare the new results to the old results [1-3].

3. Experimental Results

3.1 Read and Show Image

Tree Area Image



3.2 Convert Image from RGB Color Space to L*a*b* Color Space

How many colors do you see in the image if you ignore variations in brightness? There are three colors: white, blue, and pink. Notice how easily you can visually distinguish these colors from one another. The L*a*b* color space (also known as CIELAB or CIE L*a*b*) enables you to quantify these visual differences.

3.3 Classify the Colors in 'a*b*' Space Using K-Means Clustering

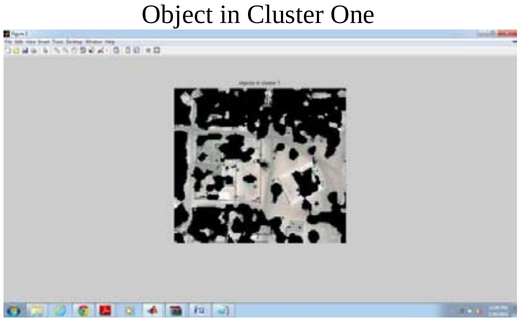
Clustering is a way to separate groups of objects. K-means clustering treats each object as having a location in space. It finds partitions such that objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. K-means clustering requires that you specify the number of clusters to be partitioned and a distance metric to quantify how close two objects are to each other [9].

3.4 Every Pixel in the Image Using the Results from KMEANS [3]

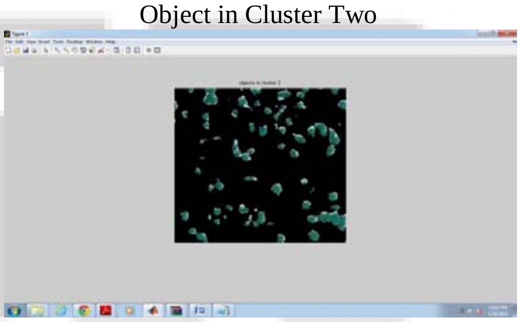
Image labeled by cluster index

3.5 Create Images that Segment the Tree Image by Color

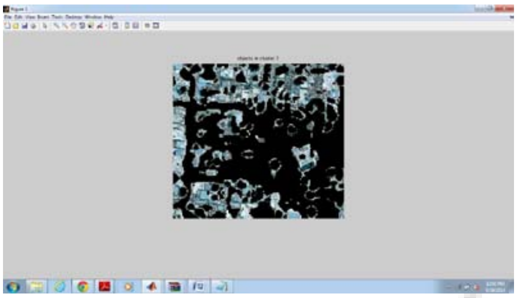
Object in Cluster One



Object in Cluster Two



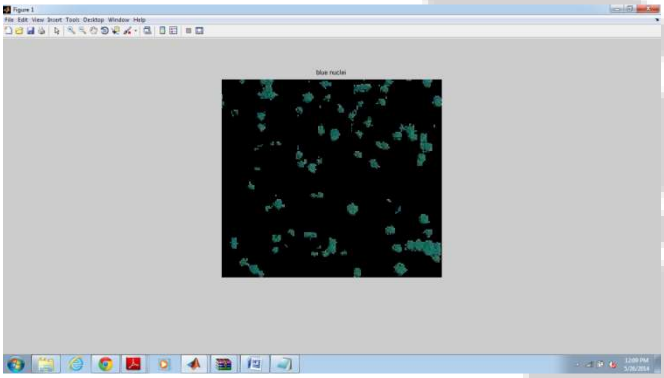
Object in Cluster Three



3.6 Segment the Nuclei into a Separate Image

Notice that there are dark and light green objects in one of the clusters. You can separate dark green from light light using the 'L*' layer in the L*a*b* color space. The cell nuclei are dark green. By using this we are identifying trees.

Green Nuclei



4. Conclusion

In this paper we successfully implemented k-mean clustering algorithm for image processing. Use of this algorithm we identifying trees. Trees are identify by K-mean clustering, it is easy to implement and also not coastally as like remote sensing system. Implemented on a clustering environment, the K-mean algorithm was applied to a high spatial resolution image for tree type identification in this study.

5. Future Work

Trees are identifying by the clustering algorithm. In this paper we are identifying trees by using k-mean clustering algorithm but we also use other portioning based clustering algorithm as like K-medoid, clara and clarans. K-meadoid is not good for this because it is so costly but clearance is good for identifying trees type or spatial data type.

References

[1] Jiawei Han and Micheline Kamber, Data mining: Concept and Techniques [M], Morgan Kaufmann Publishers. Inc 2004-8.

[2] K.T Fan, Y.C Tzeng, Y.F Lin, Y. J. Su, and K. S Chen, "Tree Identification Using a Distributed K-Mean Clustering Algorithm", IEEE, 2010.

[3] Suman Tatiraju, Avi Mehta, " Image Segmentation using K-Means clustering , EM and Normalized Cuts", Report

[4] Hailiang Jin,. Baoliang Miao,"The Research Progress of Spatial Data Mining Technique", IEEE Trans.. pp 81-84, Key Laboratory of Mine Spatial Information Technologies. 2010.

[5] Mengyue Wang, Changlin Zhang, Yan Song," Extraction of image Semantic Features with Spatial-rang Mean Shift Clustering Algorithm", IEEE, 2010.

[6] Z. Shen, Luo J., Huang G., Ming D., Ma W., and Sheng H. , 2007, "Distributed Computing Model for Processing Remotely Sensed Images Based on Grid Computing," *Information Science*, Vol. 117, pp. 504-518, 2007.

[7] R. Komura and Muramoto K. , 2007, "Classification of Forest Stand Considering Shapes and Size of TreeCrown Calculated from High Spatial Resolution SatelliteImage," *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, (IGARSS 2007), July 23-27,Barcelona, Spain, pp. 4356-4359, 2007.

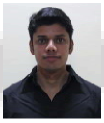
[8] P. Meyer, Staenz K., and Itten K. I. , 1996, "Semi automated Procedures for Tree Species Identification in High Spatial Resolution Data from Digitized Colour Infrared-Aerial Photography," *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 51, pp. 5-16, 1996.

[9] Q. Yu, Gong P., Clinton N., Biging G., Kelly M., and Schirokauer D. , 2006, "Object-based Detailed Vegetation Classification with Airborne High Spatial Resolution Remote Sensing Imagery," *Photogrammetric Engineering & Remote Sensing*, Vol. 72, pp. 799-811, 2006.

[10] Amiya halder, Avijit Dasgupta, Susmita Ghosh," Image Segmentation using Rough- Fuzzy K-medoid Algorithm", IEEE, 2012

Author Profile

Archana Singh, She is pursuing M.tech, in Galgotias University Greater Noida, Delhi NCR, She is completed M.Sc in Information Technology from Punjab Technical University, Jalandhar, Area of interest is Data Mining and Software Development.



Anuj Khuttan, He is pursuing M.tech, in Galgotias University Greater Noida, Delhi NCR, He is completed B. Tech in Electronics And Communication from Guru Gobind Singh Indraprastha University, New Delhi, Area of interest is Data Mining