check for updates

COLOR-BASED FOREST COVER TYPE IMAGE SEGMENTATION USING K-MEANS CLUSTERING APPROACH

🔗 **Yeong Nain Chi**

Department of Agriculture, Food and Resource Sciences University of Maryland Eastern Shore Princess Anne, MD, USA.
Email: ychi@umes.edu Tel: (410)651-8186

ABSTRACT

In order to understand forest composition, classifying forest cover type can help research regarding forest resilience, carbon sequestration, and climate change concerns. The purposes of this study were to develop and implement some image processing functions based on the histogram of forest cover type color image, and to classify forest cover type using its color feature sets of image pixels. Color-based image segmentation that is based on the color feature of image pixels assumes that homogeneous colors in the image correspond to separate clusters and hence meaningful objects in the image. The Image Processing Toolbox of MATLAB R2019a was used to convert the original forest cover type image to the enhance contrast image, including histogram of enhance contrast image. Furthermore, It was also used to analyze color-based forest cover type image segmentation using the enhance contrast image for this study. Using K-Means clustering analysis, a three-cluster solution was developed, labeled as *Hardwoods (Yellow Color) Cover Type*, *Hardwoods (Gray Color) Cover Type*, and *Loblolly Pines Cover Type*. There was a significant difference among three different forest cover type clusters in terms of histograms and L*a*b* color space features visually.

Contribution/Originality: This study is one of very few studies which have classified forest cover type using its color feature sets of image pixels in order to understand forest composition. This study also addresses that K-Means clustering analysis can be utilized to develop and implement the classification of forest cover type image.

1. INTRODUCTION

The forests in the Lower Eastern Shore, the southern part of Maryland's Eastern Shore, are characterized by large areas of loblolly pine (*Pinus taeda*), mixed pine-hardwood, bottomland hardwood, and bald-cypress forests. In general, the mixed pine-hardwood, hardwoods, and bald cypress stands are older, mature forests, while loblolly pine stands are more evenly distributed across all age classes.

Understanding forest composition is a valuable aspect of managing the health and vitality of forest resources and ecosystems. In order to map different forest cover types (the predominant type of tree cover) in a small patch of wooded area, classifying forest cover type can help research regarding forest resilience, carbon sequestration, and climate change concerns. Hence, the primary purposes of this study were to develop and implement some image processing functions based on the histogram of forest cover type image, and to classify forest cover type using its color feature sets of image pixels.

Forest cover type classification can be done by visual interpretation, but this can be very human resource intensive because the number of pixels may be very large and interpretations can vary due to human judgement.

This may be overcome by using automated algorithms in either non-supervised or supervised approaches to give results consistent with human interpreters in allocating a pixel to one forest cover type or another [1].

The image of forest cover type can be displayed as arrays of numeric data consisting of a data matrix and a color map matrix for further image analysis using an appropriate tool (e.g., MATLAB) [2]. Image processing is the use of computer algorithms to create, process, communicate, display, and analyze images. It covers various techniques that are applicable to a wide range of applications. Among various image processing tasks, segmentation can be viewed as the first essential and important step of image performance [2].

Image segmentation is an important step in image processing, and it is used to separate objects and analyze each object individually to check what it is. Image segmentation is the process of partitioning an image into multiple different segments, which means assigning a label to each pixel (image object) in the image such that pixels with same labels share common visual characteristics. It makes an image easier to analyze in the image processing tasks [2].

In most classifications, the criteria used to derive classes are not systematically applied. Often, the use of different ranges of values depends on the importance given by the user to a particular feature (e.g., color based, region based, or edge based). Color-based image segmentation, in this study, that is based on the color feature of image pixels assumes that homogeneous colors in the image correspond to separate clusters and hence meaningful objects in the image. In other words, each cluster defines a class of pixels that share similar color properties. As the segmentation results depend on the used color space, there is no single color space that can provide acceptable results for all kinds of images [2].

2. STUDY SITE

The University of Maryland Eastern Shore (UMES), the state's historically black 1890 land-grant institution, has its purpose and uniqueness grounded in distinctive learning, discovery and engagement opportunities in the arts and sciences, education, technology, engineering, agriculture, business and health professions (https://www.umes.edu/About/Pages/Mission/). In addition to 745 acres on its main campus in Princess Anne, UMES also operates a 385-acre research farm in southern Somerset County.

The Stewart Neck Farm is located in Somerset County, Maryland (38°10'31"N 75°42'42"W), 5 miles from UMES campus Figure 1. The Stewart Neck Farm was recently acquired by the University of Maryland System and includes a 240-acre contiguous forest fragment. Roughly 210 acres of this fragment represents various plantings of mostly even-aged loblolly pine (*Pinus taeda*), and with some various sized "islands" disbursed among the cultured pine with what appears to be volunteer, mixed hardwoods. All the forested areas are also adjacent to arable plots with seasonal crops. The interior of the forest is easily accessible by way of the interconnected logging roads, as well as by several firebreaks that also serve to delineate some of the stands of loblolly pine.

3. MATERIALS

A mixed pine-hardwood forest cover type image Figure 2 in UMES's Stewart Neck Farm was extracted from World Map/Satellite (https://satellites.pro/USA_map#38.164533,-75.702329,19). The Image Processing Toolbox of MATLAB R2019a was used to process forest cover type image, including histogram Figure 3, RGB (Red, Green, Blue) Color Space Figure 4, HSV (Hue, Saturation, Value) Color Space Figure 5, and L*a*b* Color Space (L*: lightness between black--white, a* : green--red, and b*: blue--yellow) Figure 6 of forest cover type image.

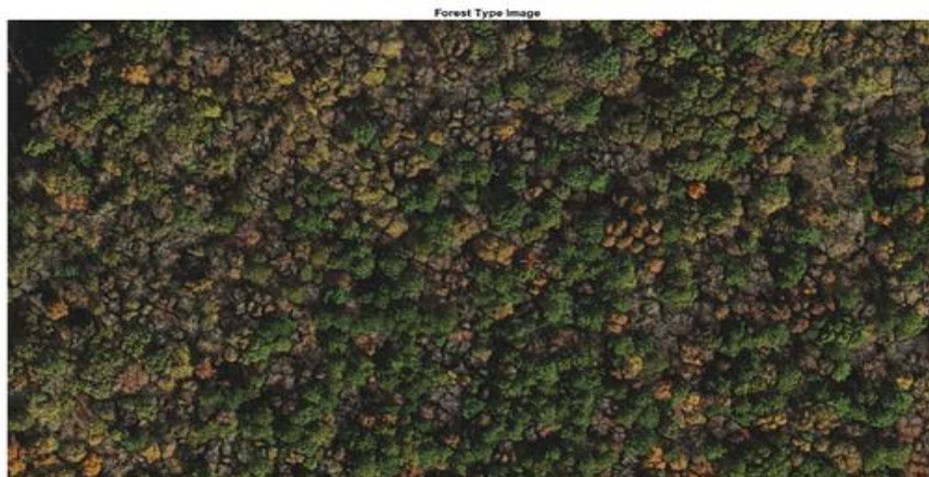**Figure-1.** UMES campus and stewart neck farm.
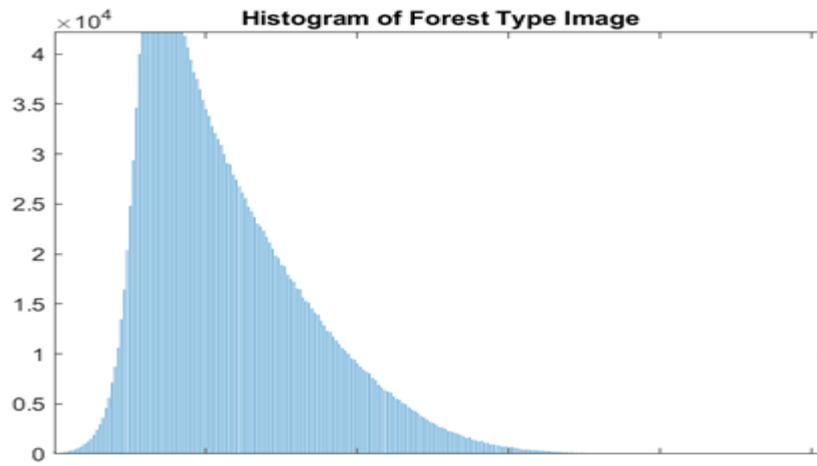

Figure-2. Forest cover type image.

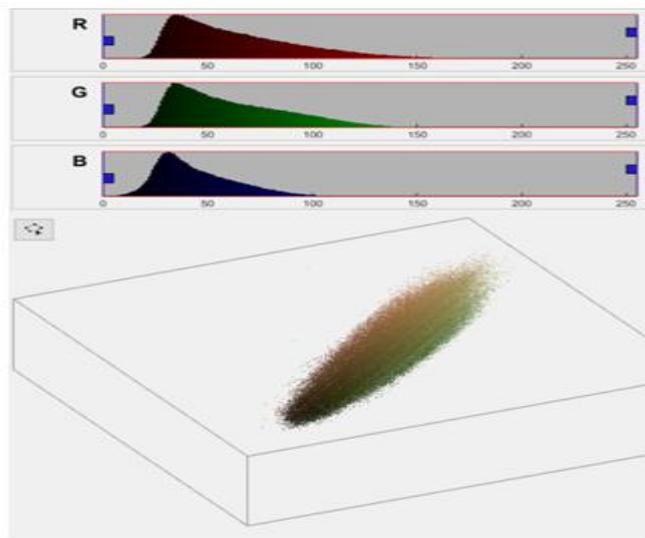**Figure-3.** Histogram of forest cover type image.

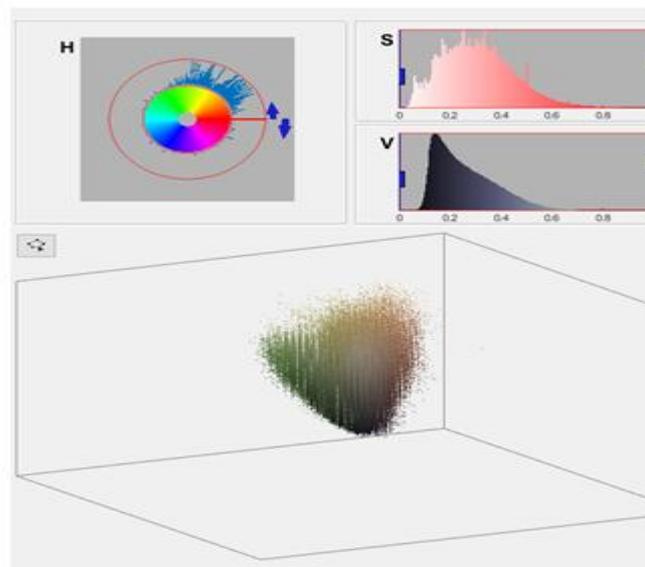**Figure-4.** RBG color space.
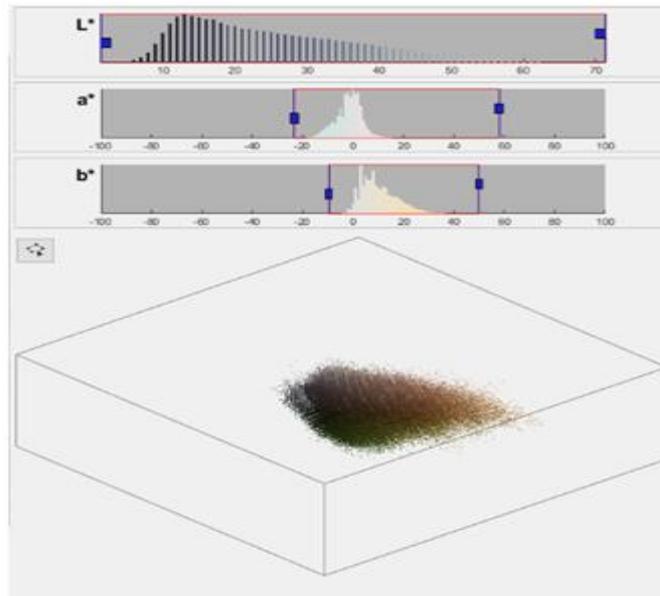

Figure-5. HSV color space.

Figure-6. L*a*b* color space.

4. METHODS

4.1. Histogram Approach

Image analysis is the process of extracting meaningful information from images such as identifying colors, finding shapes, counting objects, measuring object properties, detecting edges, removing noise, and calculating statistics for image quality. Basically, a histogram of an image provides a vast description about an image. In an image processing context, an image histogram is a graph to show how many pixels are at each scale level or at each index for the indexed color image. The histogram contains information needed for image equalization, where the image pixels are stretched to give a reasonable contrast [2].

Histogram modeling (e.g. histogram equalization) provides a sophisticated method for modifying the dynamic range and contrast of an image by altering that image such that its intensity histogram has a desired shape. Histogram equalization, a technique for adjusting image intensities to enhance contrast, employs a monotonic, non-linear mapping which re-assigns the intensity values of pixels in the input image such that the output image contains a uniform distribution of intensities (i.e. a flat histogram). This technique is used in image comparison processes and in the correction of non-linear effects introduced by a display system [2].

Histogram approach is very efficient because it typically requires only one pass through the pixels. In this technique, a histogram is computed from all of the pixels in the image, and the peaks and valleys in the histogram are used to locate the clusters in the image [3]. Histogram approach can also be applied on a per-pixel basis where the resulting information is used to determine the most frequent color for the pixel location. Histograms of images have proved effective in tasks such as image classification and object class recognition. Classification is then achieved by k-means clustering analysis [4].

4.2. K-Means Clustering Analysis

Clustering is the process of grouping object attributes and features such that the data objects in one group are more similar than data objects in another group. Clustering has been widely used in different fields such as engineering, biology, medicine, psychology, economics, and also has been found in a number of applications including data mining, search engines, recommendation systems, knowledge discovery, bioinformatics and documentation, information retrieval, computer vision and pattern recognition, and image processing.

The most popular clustering algorithm used is K-Means clustering since it is very simple, fast and efficient. It is an effective method to classify data into different groups based on input parameters and their convergence trends.

It is a type of unsupervised learning that is used to classify unlabelled data into clusters. The objective of K-Means clustering is to minimize an objective function know as squared error function [5] given by:

$$J = \Sigma^k_{j=1} \Sigma^n_{i=1} \,||X_i - C_j||^2$$

Where J = objection function, n = the number of objects, k = the number of clusters, x_i = object i, c_j = centroid for cluster j, and $||X_i - C_j||$ = the Euclidean distance between x_i and c_j.

K-Means clustering intends to partition n objects into k clusters in which each object belongs to the cluster with the nearest mean. This method produces exactly k different clusters of greatest possible distinction. It is an iterative method which assigns each object to the cluster whose centroid is the nearest. Then it again calculates the centroid of these groups by taking its average [6].

5. RESULTS

MATLAB is a high-performance language for technical computing with powerful commands and syntax. The Image Processing Toolbox provides a comprehensive suite of reference-standard algorithms and visualization functions for image analysis tasks such as image processing, image histogram modeling, and pattern recognition [2].

The Image Processing Toolbox of MATLAB R2019a was used to convert the original forest cover type image to the enhance contrast image Figure 7, including histogram Figure 8, RGB Color Space Figure 9, HSV Color Space Figure 10, and L*a*b* Color Space Figure 11 of enhance contrast image. Furthermore, It was also used to analyze color-based forest cover type image segmentation using the enhance contrast image for this study.
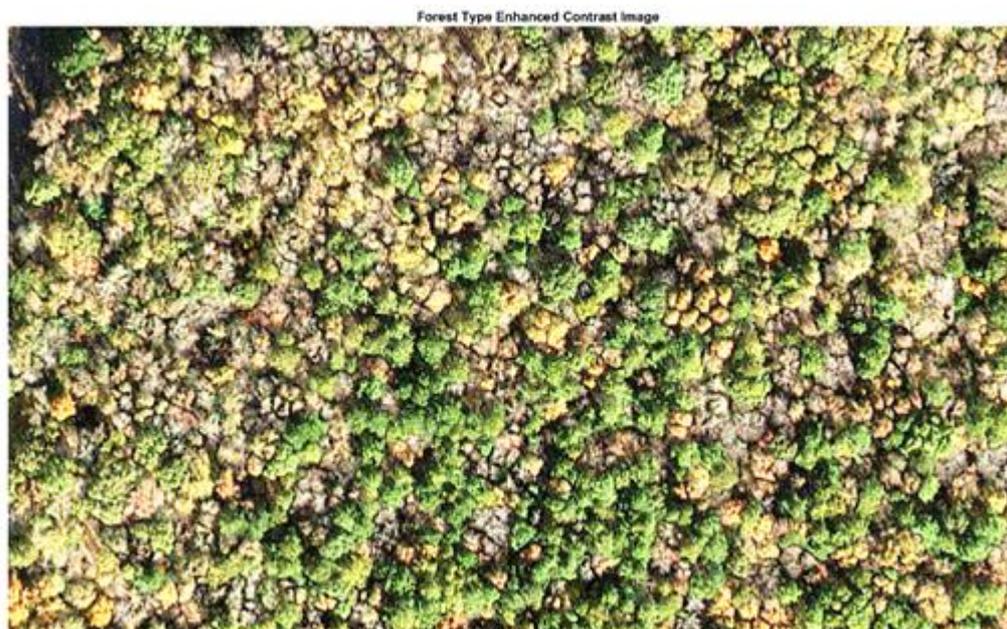

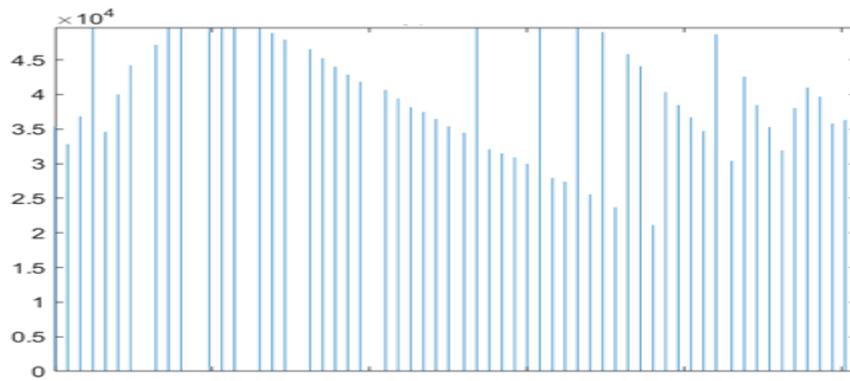**Figure-7.** Enhanced contrast image.

23

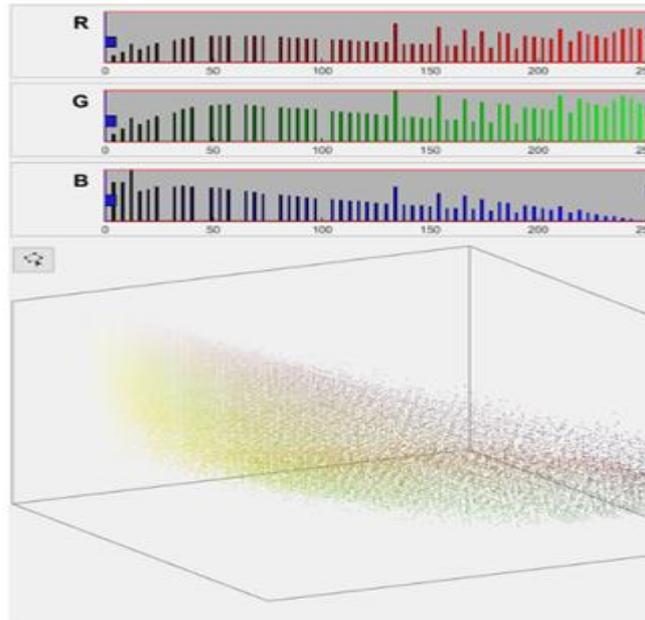**Figure-8.** Histogram of enhanced contrast image.
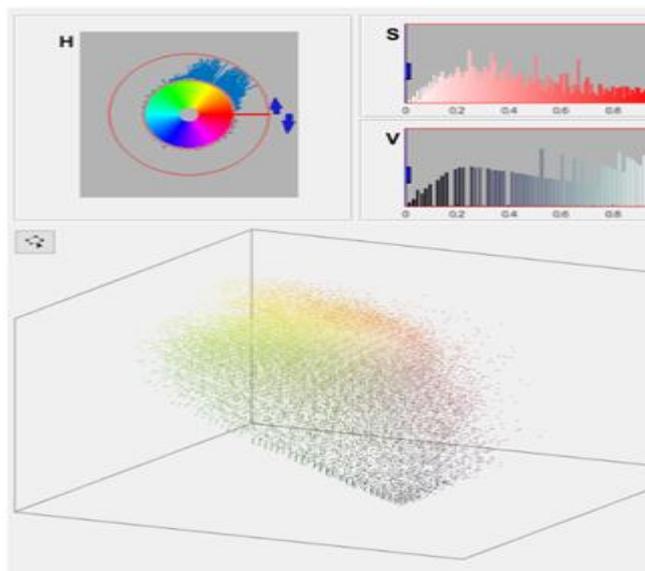

Figure-9. RBG color space.
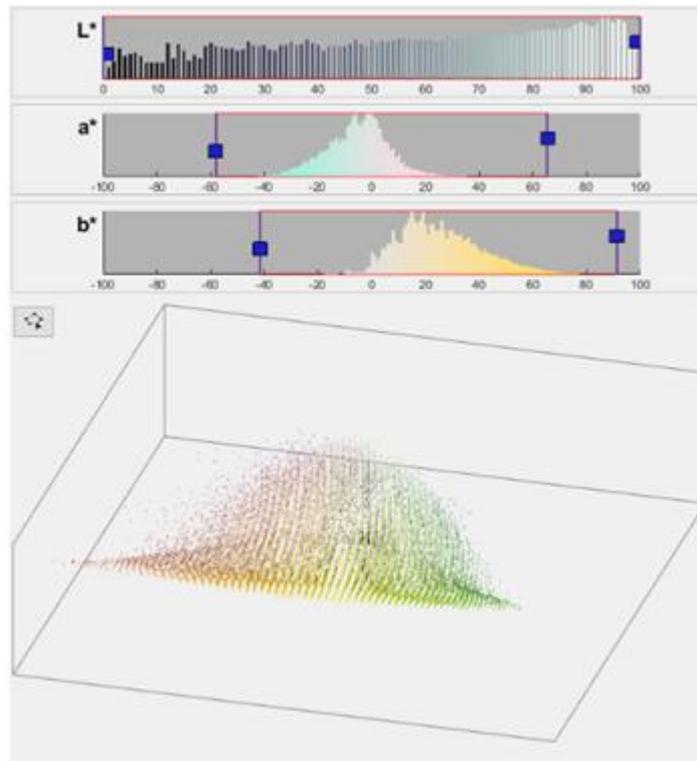

Figure-10. HSV color space.

Figure-11. L*a*b* color space.

According to Wang, et al. [7] segmentation based on the L*a*b* color space showed a strong tendency to lead good results on slightly higher performance or similar quality comparing with other color spaces. The L*a*b* color space consists of a luminosity layer 'L*', chromaticity-layer 'a*' indicating where color falls along the red-green axis, and chromaticity-layer 'b*' indicating where the color falls along the blue-yellow axis. All of the color information is in the 'a*' and 'b*' layers. The L*a*b* color space enables not only to distinguish these colors from one another visually, but also to quantify the difference between two colors using the Euclidean distance metric [8].

In this study, therefore, the L*a*b* color space was selected for color-based forest cover type image segmentation purpose. Since the color information exists in the 'a*b*' color space, the targeted objects are pixels with 'a*' and 'b*' values. In MATLAB, "imsegkmeans" can be used to cluster the objects into assigned clusters. For every object in the input, "imsegkmeans" returns an index, or a label, corresponding to a cluster [2]. Label every pixel in the image with its pixel label Figure 12.
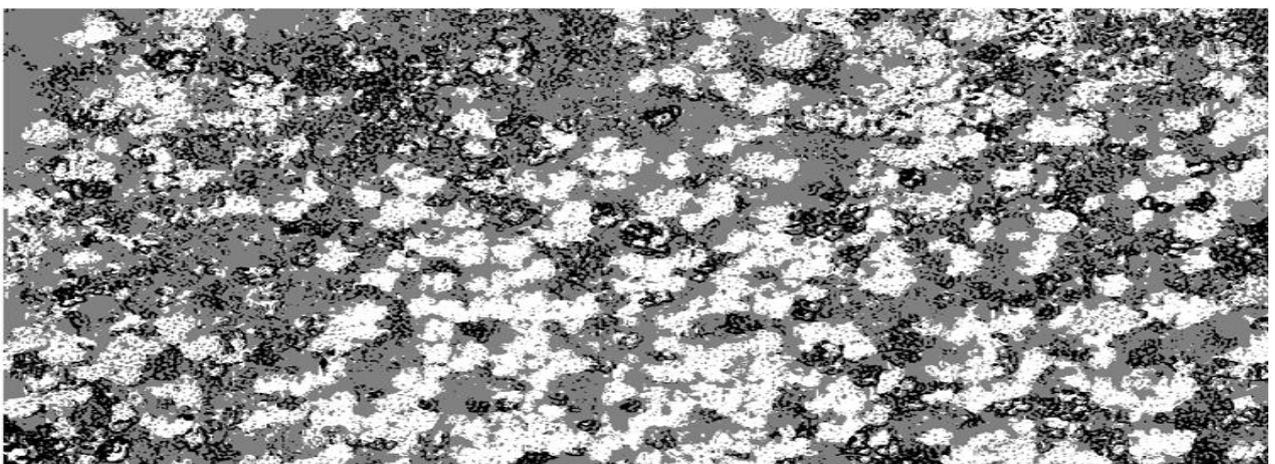

Figure-12. Image labeled by cluster index.

The K-Means clustering analysis was used to identify a solution with the specified number of clusters. Consequently, a three-cluster solution was developed upon the distance, computed using simple Euclidean distance, from the cluster centers to every object with the shortest distance to the cluster center. The clusters were labeled as *Hardwoods (Yellow Color) Cover Type*, *Hardwoods (Gray Color) Cover Type*, and *Loblolly Pines Cover Type* clusters. There was a significant difference among three different forest cover type clusters in terms of histograms and L*a*b* color space features visually.

Cluster 1 named *Hardwoods (Yellow Color) Cover Type*: Yellow color image represented as the trees in the Cluster 1 Figure 13, and also shown in the Black-White image Figure 14. The histogram of Cluster 1 image Figure 15 was flat and closed to *ZERO*, the color information in the 'a*' layer (red ~ green) was about between -5 ~ 65, and in the 'b*' layer (blue ~ yellow) was about between -5 ~ 85 Figure 16.
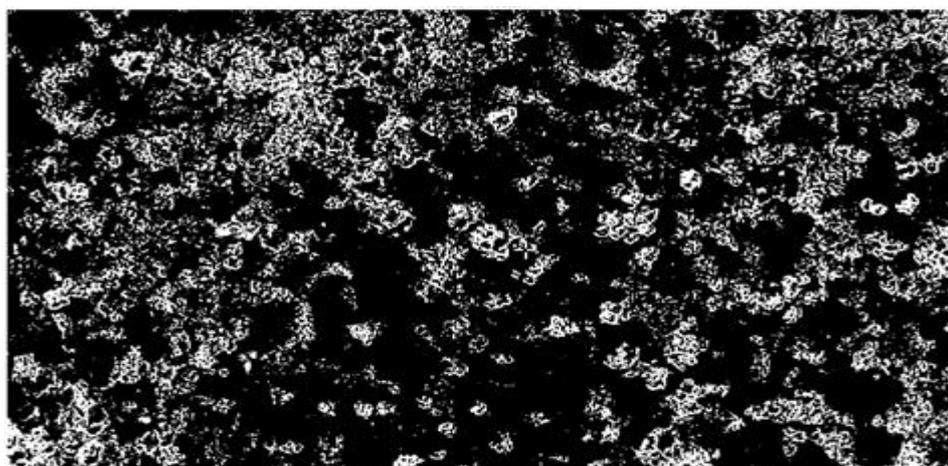
Cluster 2 named *Hardwoods (Gray Color) Cover Type*: Gray color image represented as the trees in the Cluster 2 Figure 17, and also shown in the Black-White image Figure 18. The histogram of Cluster 2 image Figure 19 was waved and between 2 ~ 4, the color information in the 'a*' layer (red ~ green) was about between -25 ~ 45, and in the 'b*' layer (blue ~ yellow) was about between -40 ~ 30 Figure 20.

Cluster 3 named *Loblolly Pines Cover Type*: Green color image represented as the trees in the Cluster 3 Figure 21, and also shown in the Black-White image Figure 22. The histogram of Cluster 2 image Figure 23 was flat and between 0 ~ 0.5, the color information in the 'a*' layer (red ~ green) was about between -60 ~ 0, and in the 'b*' layer (blue ~ yellow) was about between 0 ~ 90 Figure 24.

Cluster 1: *Hardwoods (Yellow Color) Cover Type* Figure 13, Figure 14, Figure 15, Figure 16
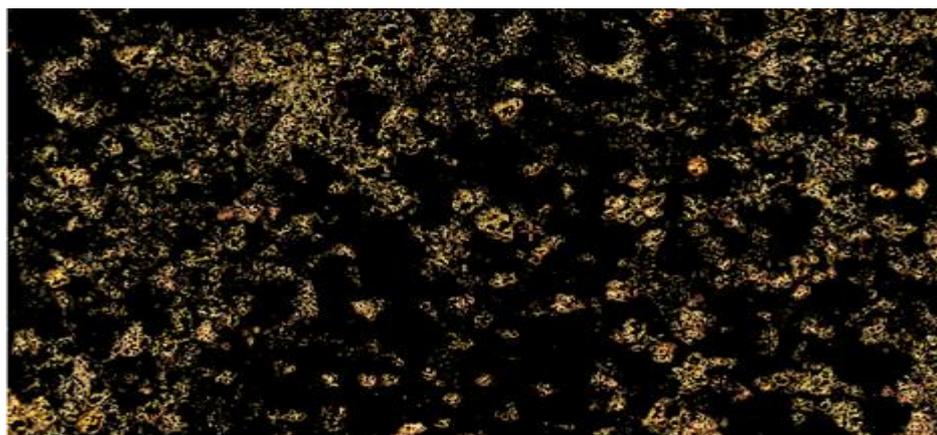
Figure-13. Trees in cluster 1 image.

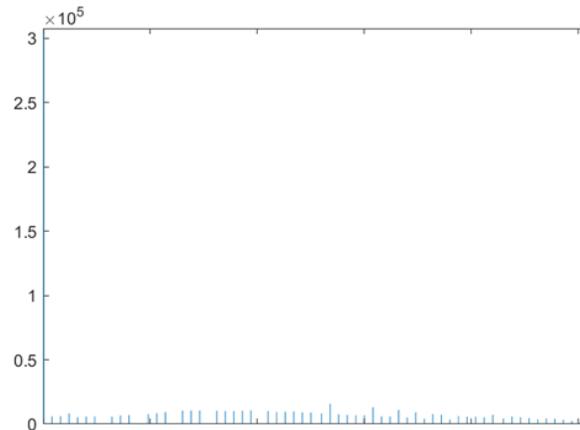**Figure-14.** Trees in cluster 1 BW image.

26
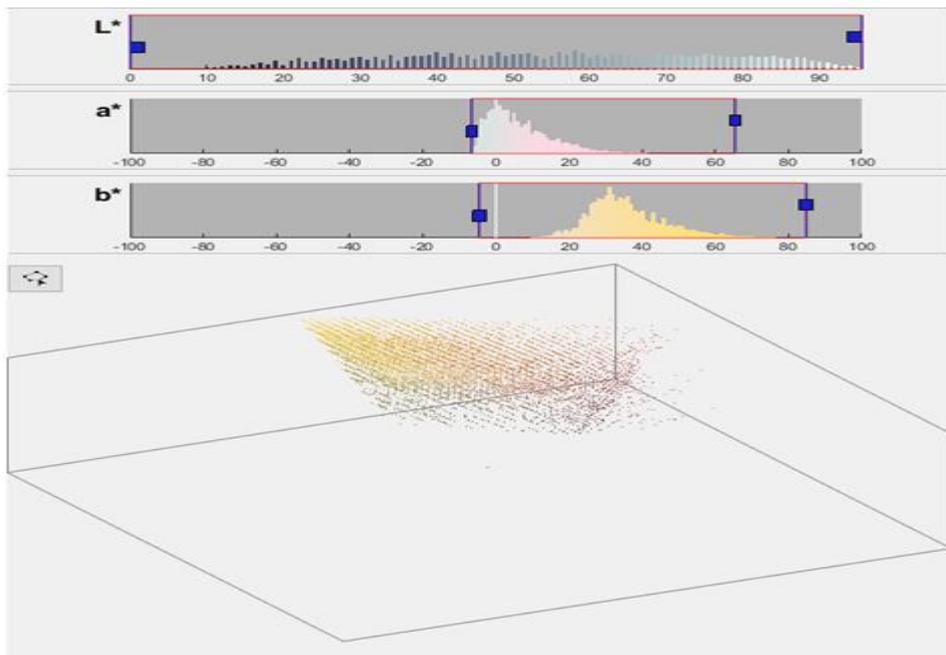
Figure-15. Histogram of cluster 1 image.

Figure-16. L*a*b* Color Space of Cluster 1 Image.

Cluster-2. Hardwoods (Gray Color) Cover Type Figure 17, Figure 18, Figure 19, Figure 20
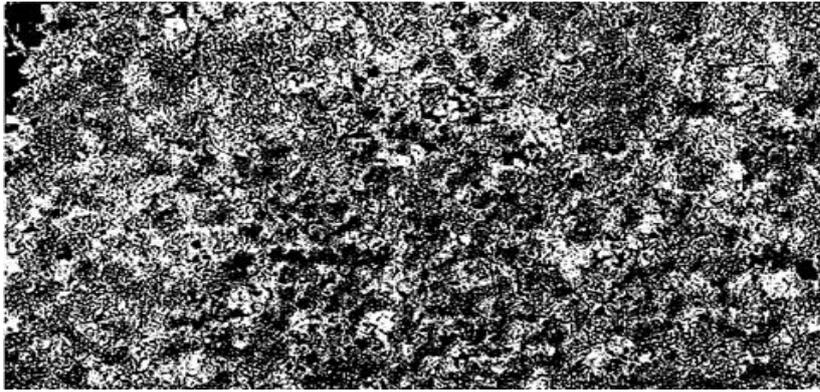

Figure-17. Trees in cluster 2 image

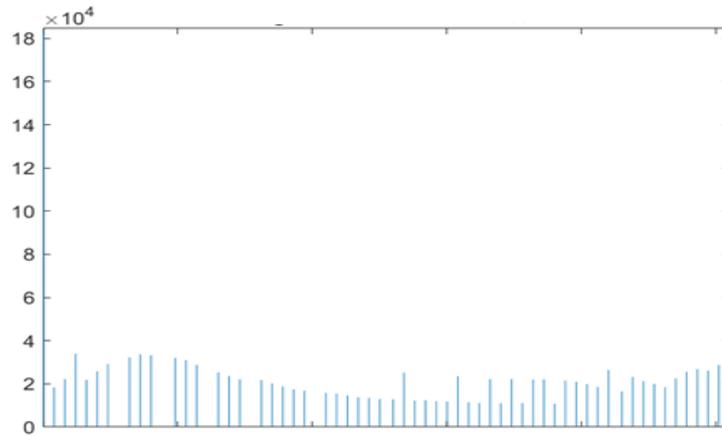**Figure-18.** Trees in cluster 2 BW image.

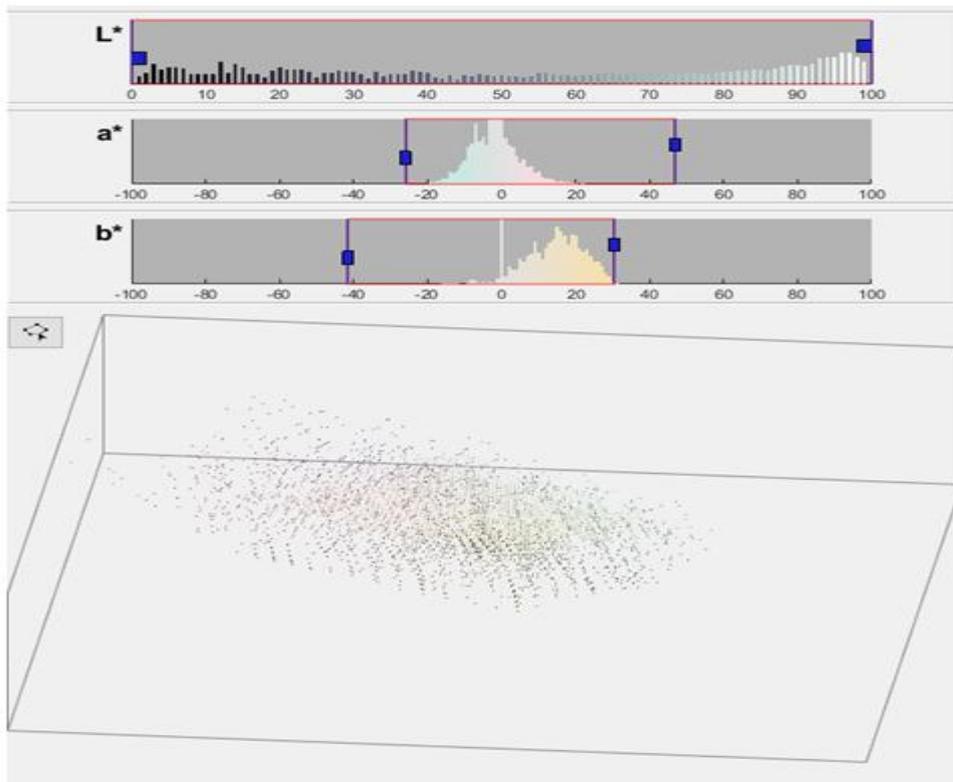**Figure-19.** Histogram of cluster 2 image.

Figure-20. L*a*b* color space of cluster 2 image.

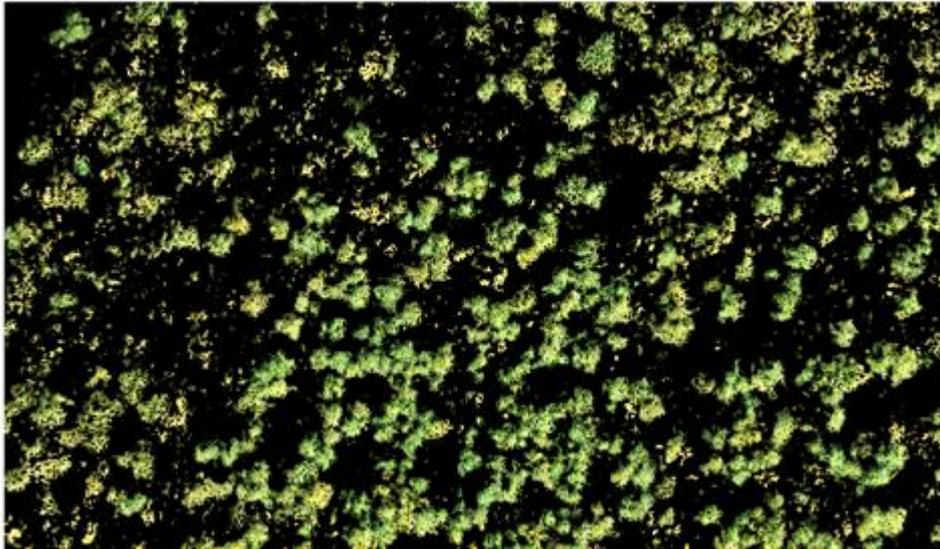Cluster 3: *Loblolly Pines Cover Type* Figure 21, Figure 22, Figure 23, Figure 24
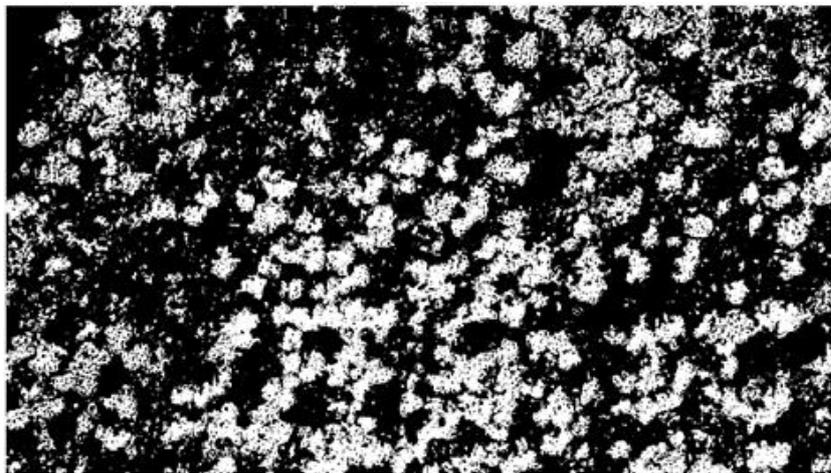
Figure-21. Trees in cluster 3 image.

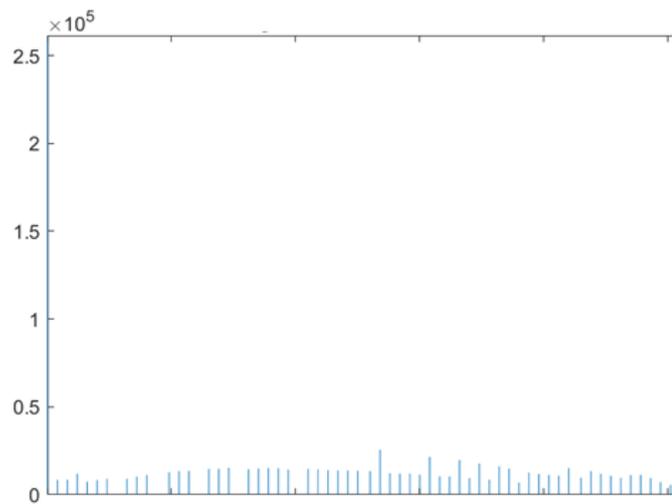**Figure-22.** Trees in cluster 3 BW image.

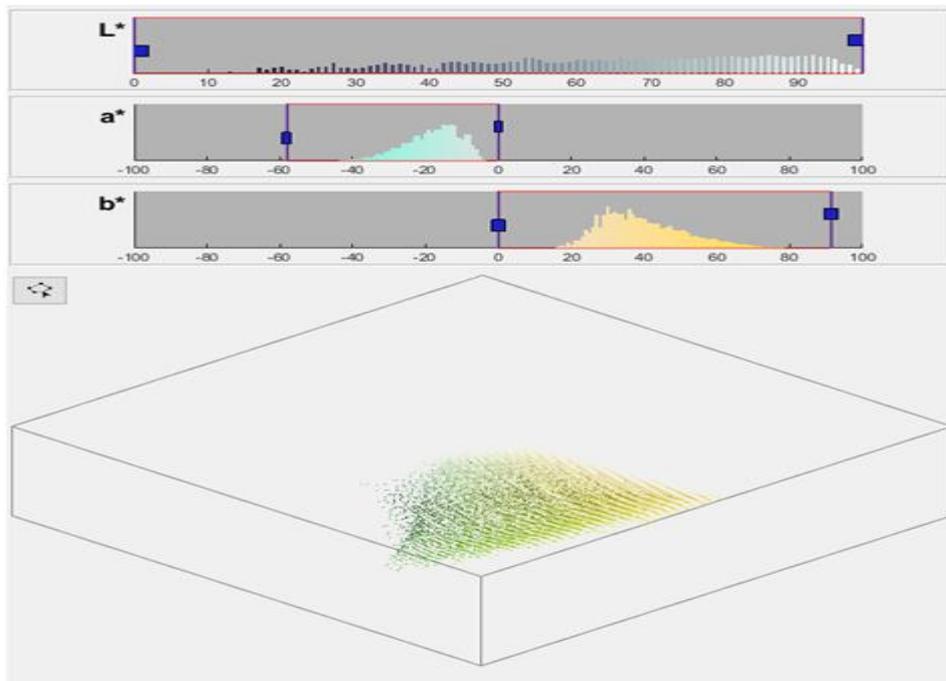**Figure-23.** Histogram of cluster 3 image.

Figure-24. L*a*b* Color space of cluster 3 image.

6. CONCLUSION

Using a mixed pine-hardwood forest cover type color image, this study provided a hands-on practice to classify forest cover type image using the image processing tool (e.g. MATLAB) to understand forest composition for further research regarding forest resilience, carbon sequestration, and climate change concerns. Forest shows large reflectance variation because of phenology, which complicates forest cover type classification. Thus, one further task should be to investigate how forest cover type classification accuracy depends on the seasonal variation (i.e. leaf-on and leaf-off season). For future research work, however, how to gain insights into the higher performance of forest cover type classification using satellite images would be a huge challenge. Furthermore, how to implement a convolutional neural network to determine forest cover type classification more precise and accurate would be another level of task.

REFERENCES

[1]   Global Forest Observations Initiative (GFOI), "Integration of remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests: Methods and Guidance (Edition 2.0). Food and Agriculture   Organization,   Rome.   Retrieved   from: https://unfccc.int/files/land_use_and_climate_change/redd/submissions/application/pdf/redd_20140218_mgd_repo rt_gfoi.pdf," 2016.

[2]   MathWorks, *MATLAB (R2019b) Image processing toolbox™ user's guide.* Natick, MA: The MathWorks, Inc, 2019.

[3]   L. G. Shapiro and G. C. Stockman, *Computer vision.* New Jersey: Prentice-Hall, 2001.

[4]   F. Schroff, A. Criminisi, and A. Zisserman, "Single-histogram class models for image segmentation. In: Kalra, P. K. and Peleg, S. (eds.) Computer Vision, Graphics and Image Processing. Lecture Notes in Computer Science," ed Berlin, Heidelbergb: Springer, 2006, p. 4338.

[5] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silvermank, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 881-892, 2002. Available at: https://doi.org/10.1109/TPAMI.2002.1017616.

[6] Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values," *Data Mining and Knowledge Discovery*, vol. 2, pp. 283-304, 1998.

[7] X. Wang, R. Hänsch, L. Ma, and O. Hellwich, "Comparison of different color spaces for image segmentation using graph-cut," in *Proceedings of the 9th International Conference on Computer Vision Theory and Applications (VISAPP-2014)*, 2014, pp. 301-308. Retrieved from: https://www.cv.tu-berlin.de/fileadmin/fg140/VISAPP_2014_127_CR.pdf.

[8] P. J. Baldevbhai and R. Anand, "Color image segmentation for medical images using L* a* b* color space," *IOSR Journal of Electronics and Communication Engineering*, vol. 1, pp. 24-45, 2012. Available at: https://doi.org/10.9790/2834-0122445.

Appendix

MATLAB Code (MathWorks, 2019) for this study was programmed as follows:

```
>> img = imread('ForestTypeImage');
>> imshow(img), title('Forest Type Image');
>> imhist(img), title('Histogram of Forest Type Image');
>> img1 = histeq(img);
>> imshow(img1), title('Forest Type Enhanced Contrast Image');
>> imhist(img1), title('Histogram of Forest Type Enhanced Contrast Image');
>> text(size(img1,2),size(img1,1)+15,...
        'Forest Type Image in the Stewart Neck Farm, UMES', ...
        'FontSize',7,'HorizontalAlignment','right');
>> lab = rgb2lab(img1);
>> ab = lab(:, :, 2:3);
>> ab = im2single(ab);
>> nColors = 3;
>> labels = imsegkmeans(ab, nColors, 'NumAttempts', 3);
>> imshow(labels,[]), title('Image Labeled by Cluster Index');
>> cover1 = labels==1;
>> cluster1 = img1 .* uint8(cover1);
>> imshow(cluster1), title('Trees in Cluster 1');
>> imhist(cluster1), title('Histogram of Trees in Cluster 1');
>> cover2 = labels==2;
>> cluster2 = img1 .* uint8(cover2);
>> imshow(cluster2), title('Trees in Cluster 2');
>> imhist(cluster2), title('Histogram of Trees in Cluster 2');
>> cover3 = labels==3;
>> cluster3 = img1 .* uint8(cover3);
>> imshow(cluster3), title('Trees in Cluster 3');
>> imhist(cluster3), title('Histogram of Trees in Cluster 3');
```