**TEKNOKOM : Jurnal Teknologi dan Rekayasa Sistem Komputer**
Vol. 6, No. 1, March 2023, pp. 8 - 13
ISSN: 2686-3219, DOI: 10.31943/teknokom. v6i1.90

# MONITORING VEGETATION HARVEST OF COFFEE TREES USING KNN-CLUSTERING ALGORITHM

**Dwi Handoko[1*], Nizamiyati[1], Herlini Oktaria[1], Agus Mulyanto[1], Muhamad Brilliant[1]**
[1]Institut Teknologi dan Bisnis Diniyyah Lampung

**A R T I C L E I N F O**
***History of the article:***
Received November 29, 2022
Revised December 21, 2022
Accepted December 25, 2022
Published March 3, 2023

**Keywords:**
Coffee Land
Coffee Tree
KNN-Clustering
Monitoring
Segmentation

**ABSTRACT**

Coffee is one of the plantation commodities spread throughout Indonesia. Coffee is the main commodity for export in Tanggamus Regency. The prediction of crop yields based on aerial photography is the main problem in this study, then there is no dataset of aerial imagery of coffee plantations that are specifically used for the purpose of determining coffee tree vegetation on coffee plantations so that farmers can find out which land is still overgrown by other plants. in addition to coffee trees and the possibility of making predictions for crop yields from aerial imagery of the coffee plantations, this research is also another urgency. This study is intended to build an intelligent model to detect the amount of coffee tree vegetation in a plantation using the KNN-Clustering segmentation algorithm. The image of the coffee tree was taken using a drone with a height of 50 m and an area of 0.25 ha. Preprocessing was carried out. The preprocessed image is called a dataset. After that, the segmentation process is carried out using the Region Growing method to form a black and white image. After Region Growing is done, then the image in Clustering uses the KNN-Clustering method to determine the color pattern of the image in the coffee plantation to distinguish the types of vegetation in the coffee plantation. From the results of KNN-Clustering, the area of coffee tree vegetation is obtained from a total of 0.25 ha of coffee plantation images.

**Correspondece:**
Dwi Handoko,
Institut Teknologi dan Bisnis Diniyyah Lampung,
Email : dwihandoko2401@gmail.com



## INTRODUCTION

Indonesia is known as an agricultural country with most of the population working in the plantation sector with vast land and various natural resources [1]. Plantations have a very important role in the sector of fulfilling basic needs, the social sector, the economic and trade sectors, especially coffee plantations [2]. However, efforts to realize this is still done manually by relying on very limited human capabilities. Therefore, the successful development of coffee requires the development of sustainable plantations with the application of good coffee cultivation techniques following the Regulation of the Minister of Agriculture of the Republic of Indonesia Number 49/Permentan/OT.140/4/2014, Concerning Technical Guidelines for Good Coffee Cultivation[3]. Abundant natural resources should be viewed in depth so that their utilization can be very optimal.

Coffee is one of the plantation commodities that are quite developed and spread throughout Indonesia, in Lampung itself coffee is the main commodity on the slopes of Mount Tanggamus, one of which is in Ulubelu District, Tanggamus Regency. Based on the above policies, it is necessary to carry out research on image segmentation of coffee plantations which aims to determine the amount of coffee tree vegetation in a plantation with the parameters of the plantation area. So it will be able to calculate the amount of harvest from the coffee plantation.

Aerial photography has been widely recognized and widely used for spatial information analysis to map the situation or condition of the earth's surface in general and especially the land surface [4]. Aerial photography can also be used for making topographic maps, altitude determination models, and other spatial applications. Aerial photography is also often used to detect changes in objects or vegetation on the surface of the earth. However, photography is quite expensive if it is needed with high frequency.

## RESEARCH METHOD

The research will be carried out according to the following stages:

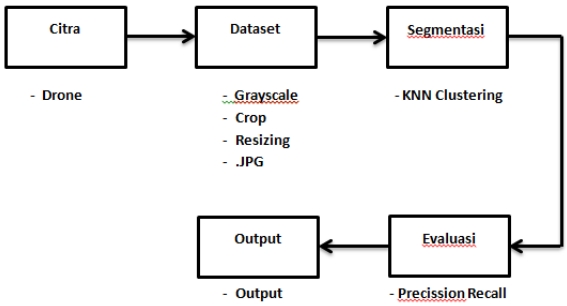



Figure 1. Research Method

In Figure 1 there are steps in the research which are explained in detail as follows:

**Image.**

The image to be analyzed is taken through aerial imagery or aerial photography using a drone. The images that will be collected as a dataset are 100 aerial images of coffee plantations with different altitude levels. In this step, an aerial image of the coffee plantation land is produced.

**Dataset.**

After the images have been collected, the next step is to create an aerial image dataset of coffee plantations using preprocessing techniques such as cropping, grayscaling (gray images), resizing, and changing the format to .jpg. The reason researchers use the .jpg format is because of standard considerations, storage, and process simplicity.

**Segmentation**

Segmentation uses the KNN-Clustering algorithm. This smart model recognizes the amount of coffee tree vegetation from aerial imagery of coffee plantations by determining the K2 parameter. After the K2 point is found, the Euclidean distance calculation will be determined. After this distance is known, the closest distance from point K2 is determined. The points with the closest distance are then grouped into clusters to be evaluated.

**Evaluate**

Evaluation will be carried out to calculate accuracy using the ROC method to calculate precision-recall. This step is to determine the percentage accuracy of aerial imagery of coffee plantation land according to suitability.

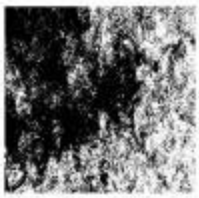
**Output**

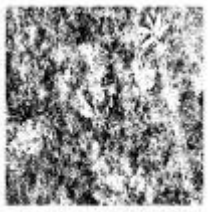
The resulting image output is an aerial image of coffee plantation land that has been segmented using the KNN-Clustering Algorithm and the evaluation results.

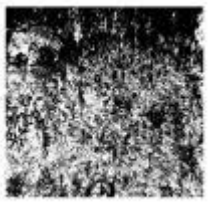
**Methods**

In this research, an aerial photo dataset with no less than 100 pieces will be built consisting of coffee plantation land images and aerial photography images. To simplify the algorithm and reduce the complexity of the segmentation and detection processes, a preprocessing step is carried out which changes the format of aerial photography and images of coffee plantations to images in .jpg format. The reason the research will be carried out using .jpg images is due to standard considerations, storage, and process simplicity [13].

Tabel 1. Example of *Aerial Photography*

| No | Citra | *Grayscale* |
|---|---|---|
| 1 |  |  |
| 2 |  |  |
| 3 |  |  |
| 4 |  | |



5  

**Determination / Identification of Individual Vegetation**

In this study, the K-Nearest Neighbor (KNN) algorithm was used. The K-Nearest Neighbor (K-NN) algorithm is a classification algorithm for a set of data based on learning data that has been previously classified [14]. Included in supervised learning, where the results of the new query instance are classified based on the majority of the proximity of the categories in the K-NN.

This algorithm works based on the shortest distance from the test sample to the training sample to determine its KNN. After collecting the KNN, then the majority of the KNN is taken to be used as a prediction from the test sample. Near or far neighbors are usually calculated based on Euclidean distance.

The KNN algorithm is carried out by comparing test data with training data. The stages that will be used in segmentation are: i. determine the number of K values, ii. calculating the distance of the dataset to the training data, iii. sort the data with the smallest euclidean distance, iv. determine the dataset group based on the majority label on the value of K.

**ROC**

ROC is a testing technique to visualize, organize and select classifiers based on their performance. ROC states the probability of occurrence of errors and the correctness of the matching on the system. By comparing the results of manual segmentation with the results of system segmentation, several values are obtained. True Positive has the same area which means the truth between the results of manual segmentation which only belongs to the results of system segmentation which means the uncertainty value between the results of manual segmentation and system segmentation. True Negative where the area that is not owned by the two objects means the truth value is outside the results of the system's manual segmentation or also called background. Meanwhile, False Negative is an area where only the original image has it, which means the value of inaccuracy between the results of manual segmentation and the background of the segmentation results. From these values, measurements are made by calculating the sum of the TP, FP, FN and TN values of the object being measured with the reference object. Measurements are made to measure accuracy, sensitivity and specificity which are described by the equation:

$$\text{Accuration} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (2)$$

**RESULTS AND DISCUSSION**

**Data**

Before carrying out the segmentation process on the image of coffee plantations, data collection is first carried out. The data that will be processed in this discussion is image data of coffee plantations obtained from direct image capture using drones.

Figure 2. Image Results of Coffee Plantation Land Obtained from Drones

In the picture above is an image of a coffee plantation located in Ulubelu, Tanggamus district taken from a height of 50 meters, where the image looks clear.

**Preprosessing**

At this preprocessing stage, the image size will be equalized to 1028 kb x 578 kb. This size matching aims to speed up the running process of the program system.

Figure 3. Preprocessing Image of Coffee Plantation Land

**Grayscale**

In segmenting the image of coffee plantation land, the first process that is carried out is to convert the RGB image into a grayscale image. In this process, the first step is that the input image (RGB image) will be converted into a gray image, the gray image itself is a digital image that only has one channel value for each pixel, in other words, the value of the RED=GREEN=BLUE section. The. A JPG image used as input is an image that has RGB color, so it must be converted to a gray image or also called a grayscale image. To convert a color image that has a matrix value of r, g, and b respectively to a grayscale image with a value of s, the conversion can be done by taking the average of the values of r, g, and b. RGB image can be seen in Figures 3 and 4 where the input image is an image obtained using a drone and is still in the RGB color pattern.

Figure 4. *Grayscale* Image of Coffee Plantation Land

**Region Growing**

Region growth is segmentation based on interrelated regions. The first thing to do in a region-growing segmentation is to determine the seed point. In determining the Seed Point, it is better to use the Maximum Gray Level value. The maximum gray level value is 255, then S = 255. [RGAW]. Next, choose a threshold. In this study a threshold of 120 was used, then T = 120. The image below is the result of the growing region segmentation which is a continuation of the input data on the image of coffee plantation land in Figure 5.

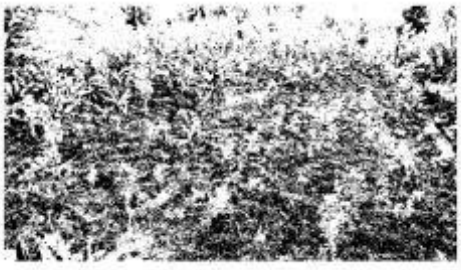


Figure 5. *Region Growing* Image of Coffee Plantation Land

From the results of research observations, there are 2 colors resulting from region growing segmentation, namely black and white. The black color represents an image with a higher height, while the white color represents an empty area or a lower image, or in the researchers' observations, the white image represents more coffee tree spacing. From the results of region growing, clustering is then carried out using KNN-Clustering so that the image is not only represented in black and white.

**KNN-Clustering**

The results of region growing segmentation only have black and white colors, so clustering is necessary to group regions with similar pixels based on colors that are more detailed and represent vegetation types more clearly. This research uses Neighbor connectivity 2 to 4 connections because the amount of image vegetation on coffee plantations is not too diverse.

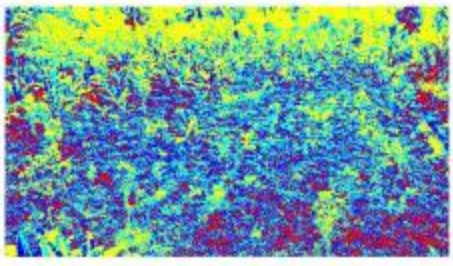


Figure 6. Image *Clustering* of Coffee Plantation Land

From the results of observations on the clustering process that has been added to the labeling, where the results of this labeling will later be used as a ground truth image to measure accuracy. The label is a certain color that can distinguish an image based on its pixels. Where the light blue color describes the surface of the land that is not planted with coffee tree vegetation

or also the spacing between one coffee tree vegetation and another, the dark blue color describes the land with coffee tree vegetation, the red color describes lower vegetation and the yellow color describes the surface. land was overgrown with vegetation taller than coffee trees. The pixel area of the coffee tree vegetation in this study is 246591 pixels.

**Pecission Recall**

This research has results in the form of a histogram of images of coffee plantations, where the image has 2 intensities towards 0 black images and 1 white image with a number of black images of more than 1*105. Here we can see changes in the intensity of areas covered with coffee tree vegetation.

Figure 7. *Pecission Recall* Histogram

**Output**

The resulting output is a system that can display grayscale imagery, region growing imagery, and clustering imagery results, and there are KNN-output calculation values, threshold, the number of coffee tree vegetation, and the number of neighbors, from the original coffee tree vegetation image taken using a drone.


Figure 8. Smart System for Image Segmentation of Coffee Plantation Land

smart software model for detecting the amount of coffee tree vegetation in a plantation using the KNN-Clustering segmentation algorithm. The coffee tree image was taken using a drone with a height of 50m and an area of 0.25 Ha and preprocessing was carried out. The resulting preprocessing image is called a dataset. After that, a segmentation process is carried out using the Region Growing method to form a black and white image. After doing Region Growing, then the image is clustered using the KNN-Clustering method to determine the color pattern of the image on the coffee plantation to distinguish the types of vegetation on the coffee plantation.

**CONCLUSION**

The results of this study are a model that can classify coffee tree vegetation on a coffee plantation and calculate the total area of coffee tree vegetation, so that with this model farmer are expected to know the amount of land area covered by other vegetation, so that crop yields can increase in the future by minimizing other vegetation besides coffee trees on their plantation land. the results of color classification using the KNN clustering method obtain a fairly high level of accuracy, namely 80.3516% so that this model can be used for classification of coffee vegetation in coffee plantation images.

**ACKNOWLEDGMENT**

This research was supported/partially supported by Institut Teknologi dan Bisnis Diniyyah Lampung. We thank our colleagues from Institut Teknologi dan Bisnis Diniyyah Lampung who provided insight and expertise that greatly assisted the research, although they may not agree with all of the interpretations.

**REFERENCES**

[1] M. FADLY, “Faktor-faktor yang Mempengaruhi Produksi Kopi di Kabupaten Barru (Studi Kasus di Desa Harapan Kecamatan Tanete RiajaKabupaten Barru),” Apr. 2021.

[2] A. Sektor Basis Sub Sektor Pertanian Terhadap Tenaga Kerja Sektor Pertanian Di Kabupaten Kudus, J. Arinda, and R. Retno Sugiharti, “Analisis Sektor Basis Sub Sektor Pertanian Terhadap Tenaga Kerja Sektor Pertanian Di Kabupaten Kudus,” *J. Paradig. Multidisipliner*, vol. 2, no. 1, May 2021, doi: 10.1210/.V2I1.59.

[3] F. Fachruddin, R. Fadhil, Syafriandi, and D. Dahlan, “Suitability analysis of scrubland for arabica and robusta coffee plants in Aceh Besar Regency,” *IOP Conf. Ser. Earth Environ. Sci.*, vol. 644, no. 1, p. 012011, Jan. 2021, doi: 10.1088/1755-1315/644/1/012011.

[4] Brilliant, Muhamad, et al. "Land Cover Changes Detection Using Region Growing Segmentation." *Journal of Physics: Conference Series*. Vol. 1529. No. 2. IOP

Publishing, 2020.
[5] A. Padmo, A. Masa, and H. Hamdani, “Klasifikasi Motif Citra Batik Menggunakan Convolutional Neural Network Berdasarkan K-means Clustering,” *J. MEDIA Inform. BUDIDARMA*, vol. 5, no. 4, pp. 1292–1301, Oct. 2021, doi: 10.30865/MIB.V5I4.3246.
[6] M. Simsek, B. Kantarci, and A. Boukerche, “Utility-Aware Legitimacy Detection of Mobile Crowdsensing Tasks via Knowledge-Based Self Organizing Feature Map,” *IEEE Trans. Mob. Comput.*, 2021, doi: 10.1109/TMC.2021.3136236.
[7] N. Kaur, N. Nazir, and Manik, “A Review of Local Binary Pattern Based texture feature extraction,” pp. 1–4, Nov. 2021, doi: 10.1109/ICRITO51393.2021.9596485.
[8] “Utility-Aware Legitimacy Detection of Mobile Crowdsensing Tasks via Knowledge-Based Self Organizing Feature Map | IEEE Journals & Magazine | IEEE Xplore.” https://ieeexplore.ieee.org/abstract/document/9655485 (accessed Feb. 12, 2022).
[9] T. Tommy, R. Siregar, A. M. Elhanafi, and I. Lubis, “Implementasi Color Quantization pada Kompresi Citra Digital dengan Menggunakan Model Clustering Berdasarkan Nilai Max Variance pada Ruang Warna RGB,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 6, pp. 1099–1106, Nov. 2021, doi: 10.25126/JTIIK.2021863490.
[10] Kurniawan, Yosefani, Lily Puspa Dewi, and Silvia Rostianingsih. "Implementasi Locally Adaptive K-Nearest Neighbor Algorithm based on Discrimination Class (DC-LAKNN) pada Kasus Deteksi Fake Account Instagram." *Jurnal Infra* 10.1 (2022): 98-103.
[11] D. Cahyanti, A. Rahmayani, and S. A. Husniar, “Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara,” *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 39–43, Jul. 2020, doi: 10.33096/IJODAS.V1I2.13.
[12] Handoko, Dwi, Suhendro Y. Irianto, and Sri Karnila. "Landcover Quality Detection Using Segmentation and Content Base Image Retrieval Methods." *Proceeding International Conference on Information Technology and Business*. 2019.
[13] R. WAHYUDI, A. Mursalin, and I. Sagala, “Implementasi E-government di Kejaksaan Tinggi Provinsi Jambi,” Nov. 2020.
[14] Aliabad, Fahime Arabi, et al. "Investigating the Ability to Identify New Constructions in Urban Areas Using Images from Unmanned Aerial Vehicles, Google Earth, and Sentinel-2." *Remote Sensing* 14.13 (2022): 3227.
[15] Diao, Qi, et al. "Superpixel-based attention graph neural network for semantic segmentation in aerial images." *Remote Sensing* 14.2 (2022): 305.