ISSN : 2460 – 7223

Prediction of Coffee Bean Quality Using Segmentation Methods And K-Nearest Neighbor

1st Agung Pradana
Department of Informatics Engineering
Institute of Informatics and Business (IIB) Darmajaya
Bandar Lampung, Indonesia
agung.pradana.1821210002@mail.darmajaya.ac.id

2nd Suhendro Yusuf Iranto
Department of Informatics Engineering
Institute of Informatics and Business (IIB) Darmajaya
Bandar Lampung, Indonesia
suhendro@darmajaya.ac.id

3rd Sri Karnila
Department of Informatics System
Institute of Informatics and Business (IIB) Darmajaya
Bandar Lampung, Indonesia
srikarnila_dj@darmajaya.ac.id

4th Hendra Kurniawan
Department of Informatics System
Institute of Informatics and Business (IIB) Darmajaya
Bandar Lampung, Indonesia
hendra.kurniawan@darmajaya.ac.id

***Abstract*—The condition of people's coffee farming management is relatively poor when compared to large state-owned plantations. The main problem in smallholder plantations is the quality of the results that do not meet standardization. This study designs a system that is able to identify the quality of coffee beans using Segmentation, K-Nearest Neighbor and Gray Level Co-occurrence Matrix methods. Based on the test results using texture feature extraction, the highest accuracy was obtained at K-5 of 85%. It is possible that if the K value used is too small, there will be a lot of noise which reduces the level of accuracy in data classification, but if the K value is too large it can cause errors in the range of values taken, which will indirectly affect the level of accuracy. The results of the study were the identification of coffee beans with good quality or poor quality. It is hoped that this research can contribute to improving the quality of people's coffee so that it can increase the production of people's coffee that is able to compete in the market.**

Keywords—Gray Level Co-occurrence Matrix, K-Nearest Neighbor, Segmentation

I. INTRODUCTION

Coffee is one of the plantation commodities in Indonesia which has an important role in economic activities in Indonesia. Coffee plants are also one of Indonesia's leading export commodities which contribute to the country's foreign exchange in addition to coming from oil and gas. On the other hand, the potential for coffee exports is very promising, there is also a potential domestic coffee market opportunity that can excite coffee farmers [1]. In determining the quality of really good coffee, there are many obstacles due to the subjective nature in the selection of coffee beans or a lack of understanding of science so that it requires someone who is really an expert or expert, so it can be ascertained that the coffee beans are in the good, good category [2].

Indonesia has implemented quality standards for coffee beans based on physical tests on the basis of the number of defects since 1990. This quality standard has been revised several times and is currently contained in the Indonesian National Standard (SNI) Number 01-2907-2008. Revision of quality standards is carried out to respond to the dynamics of the demands of the growing domestic and global market. Thus, the quality criteria in SNI must always refer to the international requirements issued by the ICO (International Coffee Organization) [3]. In accordance with the decision of the ICO (International Coffee Organization) since October 1, 1983 until now, to determine the quality of coffee in Indonesia using the Defects Value System. By using this system, the more defects there are, the lower the coffee quality will be, and the fewer defects, the better the coffee quality [4]

The problem of identifying the quality of coffee beans can be overcome by detecting the texture of the coffee beans using digital image processing. The method used in this research is image segmentation, Gray Level Co-occurrence Matrix and K-Nearest Neighbor with good and bad seeds classification.

Based on the test results, this system obtained the highest level of accuracy in the K3 test of 85% to identify the image of coffee beans using the image segmentation method, Gray Level Co-occurrence Matrix and K-Nearest Neighbor.

II. METHODS

A. *Coffee*

The coffee plant is a genus of Coffea which is included in the Rubiaceae family and has about 100 species. The Coffea genus is one of the important genera that has high economic value and is developed commercially, especially Coffea Arabica, Coffea Liberica, Coffea Kanephora including Robusta coffee. The coffee plant is a tropical plant originating from Africa.

Arabica coffee can grow at an altitude of 700-1,400m above sea level with temperatures ranging from 15-24˚C and a soil pH of 5.3-6.0 and an average rainfall of 2000-4000mm/ year and the number of dry months 1-3 months. / yr. Robusta coffee can grow at an altitude of 300-600 m above sea level with rainfall of 1500-3000 mm/ year with a temperature of 24-30˚C and a soil pH of 5.5-6.0 [5].

B. Preproccessing

Preprocessing is the first step taken after getting from image acquisition. This process is done to make it easier for the system to recognize objects. For the steps carried out in preprocessing as shown in Figure 1.

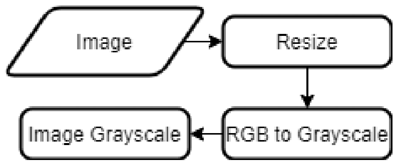

Fig. 1. Pre-Processing

Based on Fig.1, resizing is a process to reduce the image size to 200×200 aiming to ease the work of the system in image processing. The next resized image will be converted to a grayscale image. This process serves to convert a three-dimensional image into a one-dimensional one with the same intensity, so that the computational process does not require a long time. For converting RGB images to grayscale images such as equation 1[6].

$$grayscale = 0{,}299R + 0{,}587G + 0{,}114\,B \quad (1)$$

C. Region Growing Segmentation

Image segmentation is the process of dividing a digital image into several regions or groups, where each area consists of a set of pixels. Image segmentation simplifies and changes the image representation to something more meaningful and easier to analyze. Image segmentation is used to find the object you want to find and the boundaries of object shapes such as lines, curves in the image [7].

Region growing is a simple area-based image segmentation method. This method is also classified as a pixel-based image segmentation method because this method includes a selection process from the starting points[8]. This segmentation approach examines neighboring pixels from the starting points and determines whether neighboring pixels should be merged into regions. This process is iterative (iteration), with the same behavior as the usual data clustering algorithm [9].

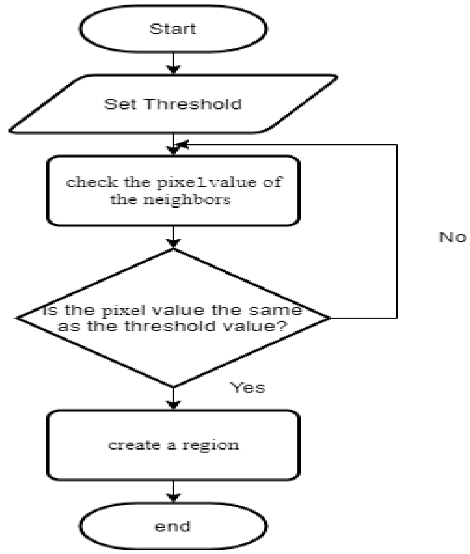

Fig. 2. Flowchart Region Growing

The stages in the region growing segmentation begin by determining the Seed Point Image. It then examines the neighbors of the selected Seed Point Seed point, selecting eight surrounding neighbors. Then process the neighboring pixels after getting eight neighboring pixels, then the next process is selecting pixels. Each neighboring pixel will be checked according to the gray color with reference to the seed point and threshold. Then specify the foreground and background. The segmented image is shown in Fig. 3.

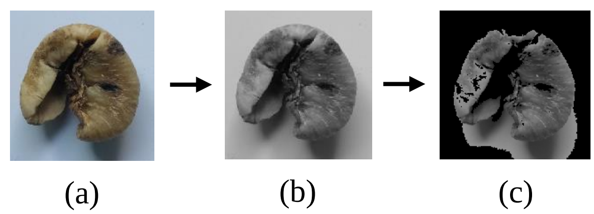

Fig. 3. (a)Citra RGB, (b)Citra Grayscale, (c)Citra Segmentasi

D. Feature Extraction

Feature extraction is the process of taking the characteristics of an object that can describe the characteristics of the object [10]. Gray Level Co-occurrence Matrix is an image feature extraction method that is widely used for image classification which is quite effective in classifying because it is able to provide detailed texture information of an image. Texture extraction is done to retrieve basic information from an image before it is used for the next process[11]. The GLCM method uses several features of a statistical approach such as energy, entropy, contrast, and so on [12].

The following are the steps used in taking texture features from an image [13]:

1) The color image is converted to a grayscale image
2) Each value of the RGB image is converted to gray using the equation (1)
3) New pixel = setPixel(255, gray value, gray value, gray value)
4) Segmentation of color values into 16 bin

5) Calculate the co-occurance matrix values in four directions of 0, 45, 90, and 135 . respectively
6) Calculate information on texture characteristics, namely contrast, correlation, homogeneity, and energy.

E. K-Nearest Neighbor Classification

K-Nearest Neighbor (K-NN) is a method that uses a supervised algorithm where the results of the new test sample are classified based on the majority of the categories in K-NN. The purpose of this algorithm is to classify new objects based on attributes and training samples [14]. According to[15] the equation used for the knn algorithm:

$$d(xi, xj) = \sqrt{\sum_{r=1}^{n}(a_r(x_j) - a_r(x_j))^2} \quad (2)$$

$d(xi, xj)$	: euclidean distance
xi	: record to i
xj	: record to j
a_r	: data to r
i,j	: 1,2,3…n

III. RESULT AND DISCUSSION

Testing this system is done by preparing the image that will be used as training data. The data used as training data in the application that was built amounted to 480 data with details of 240 good seed data and 240 bad seed data. The test data consisted of 20 images with details of 10 good seed data and 10 bad seed data. The training phase begins with image preprocessing. Furthermore, image segmentation serves to separate the background and image objects. Then get the texture feature extraction from the training image. The texture characteristics that have been obtained can then be trained using KNN. After that the training data from the results of the KNN training have been defined as data as shown in Figure 7.

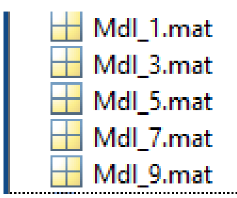

Fig. 5. Training Data

The next process is testing the new test image. This process is carried out to determine the accuracy of the model built in the training process, then test or testing data is used to predict the classes. Furthermore, the KNN function is to determine the class of test data with the classification results parameters obtained at the data training stage. The system will display the results of the coffee bean classification according to the system's predictions. Predictions of good quality coffee beans are shown in Figure 6 and predictions of poor/poor quality coffee beans are shown in Figure 7.

Fig. 6. examples of good coffee beans

Fig. 7. example of bad coffee beans

In Experiment 1, the test was carried out with the parameter K worth 1. The accuracy obtained was 75% as shown in Table I.

TABLE I. TESTING WITH K-1

Class	Class Positive	Classified Negative
+	TP (10)	FN (0)
-	FP (5)	TN (5)
$accuracy = \frac{TP + TN}{(TP + TN + FP + FN)}$ $= \frac{15}{20} \times 100 = 75\%$		

In experiment 2, the test was carried out with the parameter K value 3. The results of the experiment 2 got an accuracy of 65% as shown in Table II.

TABLE II. TESTING WITH K-3

Class	Class Positive	Classified Negative
+	TP (8)	FN (2)
-	FP (5)	TN (5)
$accuracy = \frac{TP + TN}{(TP + TN + FP + FN)}$ $= \frac{13}{20} \times 100 = 65\%$		

Experiment 3 tests were carried out with a feasible K parameter 5. The results of test 3 got an accuracy of 85% as seen in Experiment 3, the test was carried out with a feasible K parameter 5. The results of trial 3 got an accuracy of 85% as shown in Table III.

TABLE III. TESTING WITH K-5

Class	Class Positive	Classified Negative
+	TP (10)	FN (0)
-	FP (3)	TN (7)
$accuracy = \frac{TP + TN}{(TP + TN + FP + FN)}$ $= \frac{17}{20} \times 100 = 85\%$		

In experiment 4, the test was carried out with the parameter K worth 7. The results of the experiment 4 got an accuracy of 70% as shown in Table IV.

TABLE IV. TESTING WITH K-7

Class	Class Positive	Classified Negative
+	TP (8)	FN (2)
-	FP (4)	TN (6)

$$accuracy = \frac{TP + TN}{(TP + TN + FP + FN)}$$
$$= \frac{14}{20} \times 100 = 70\%$$

In experiment 5, the test was carried out with the parameter K worth 9. The results of the experiment 5 got an accuracy of 65% as shown in Table V.

TABLE V.TESTING WITH K-9

Class	Class Positive	Classified Negative
+	TP (8)	FN (2)
-	FP (5)	TN (5)

$$accuracy = \frac{TP + TN}{(TP + TN + FP + FN)}$$
$$= \frac{17}{20} \times 100 = 65\%$$

Where True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). From the test results of all experiments, the maximum accuracy at K-5 is 85% while the minimum accuracy at K-3 and K-9 is 65% as shown in Table VI.

According to the book [16], the level of accuracy can be diagnosed as follows:

Accuracy 0.90 – 1.00 = *Excellent Classification*
Accuracy 0.80 – 0.90 = *Good Classification*
Accuracy 0.70 – 0.80 = *Fair Classification*
Accuracy 0.60 – 0.70 = *Poor Classification*
Accuracy 0.50 – 0.60 = *Failure*

TABLE VI. TEST RESULT

No	Testing	K	Accuracy
1	Testing 1	1	75%
2	Testing 2	3	65%
3	Testing 3	5	85%
4	Testing 4	7	70%
5	Testing 5	9	65%

IV. CONCLUSION

Based on the results of the design, implementation and testing that has been done, the coffee bean quality prediction system with Segmentation and KNN methods is able to predict the quality of coffee beans with indicators of good beans and bad beans. The results of the accuracy test generated using the Confusion Matrix based on the K value used and the diagnosis of the accuracy level can be concluded that the system is a good classification with the success rate of identifying the quality of coffee beans getting the highest accuracy value at K-5, which is 85%.

For further development this system can be developed using other methods or it can be combined with different methods so that the accuracy results can be compared. Types of coffee can be added, such as arabica, gayo, liberika, or perhaps it can be compared with fermented coffee beans from civet/civet animals.

REFERENCES

[1] E. D. Martauli, “April 2018 Volume 01 No 02 ANALYSIS OF COFFEE PRODUCTION IN INDONESIA Elvin Desi Martauli,” vol. 01, no. 02, pp. 112–120, 2018.

[2] N. Firmansyah, “KUALITAS KOPI BERBASIS WEB DENGAN,” vol. 5, no. 3, pp. 298–306, 2017.

[3] Standar Nasional Indonesia, “Standar Nasional Indonesia : Biji kopi ; Biji kakao ; dan Rumput laut SNI Biji kopi,” *Badan Standarisasi Nasional, Jakarta*, pp. 1–60, 2017.

[4] B. Raharjo and F. Agustini, “Metode Forward Chaining pada Sistem Pakar Penilaian Kualitas Biji Kopi Berbasis Web,” vol. 4, pp. 73–82, 2020.

[5] D. I. Sulawesi, B. Selatan, and P. Abad, “Abstrak,” vol. 12, no. 1, pp. 13–26, 2017.

[6] F. Zikra, K. Usman, and R. Patmasari, “Deteksi Penyakit Cabai Berdasarkan Citra Daun Menggunakan Metode Gray Level Co-Occurence Matrix Dan Support Vector Machine,” 2021.

[7] L. Angriani, “Segmentasi Citra dengan Metode Threshold pada Citra Digital Tanaman Penyelenggara : Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Muslim Indonesia Makassar , Indonesia Editor : Tim Editor SNRIK 2015 Penerbit : Fakultas Ilmu Komputer,” no. July, 2018.

[8] S. Karnila, S. Irianto, and R. Kurniawan, “Face Recognition using Content Based Image Retrieval for Intelligent Security,” *Int. J. Adv. Eng. Res. Sci.*, vol. 6, no. 1, 2019, doi: 10.22161/ijaers.6.1.13.

[9] A. M. Simarmata, S. Kom, M. Kom, T. P. Sinaga, and O. William, “DETEKSI PENYAKIT LIMFOMA DENGAN MENGGUNAKAN,” no. 2, 2019.

[10] D. Satria and M. Mushthofa, “Perbandingan Metode Ekstraksi Ciri Histogram dan PCA untuk Mendeteksi Stoma pada Citra Penampang Daun Freycinetia,” *J. Ilmu Komput. dan Agri-Informatika*, vol. 2, no. 1, p. 20, 2013, doi: 10.29244/jika.2.1.20-28.

[11] M. Brilliant, S. Y. Irianto, S. Karnila, and R. Z. Abdul Aziz, “Land Cover Changes Detection Using Region Growing Segmentation,” in *Journal of Physics: Conference Series*, 2020, vol. 1529, no. 2, doi: 10.1088/1742-6596/1529/2/022066.

[12] M. Ramadhani, “Klasifikasi Jenis Jerawat Berdasarkan Tekstur dengan Menggunakan Metode GLCM,” *e-Proceding of Enggineering*, vol. 5, no. 1, pp. 870–876, 2018.

[13] I. G. Rai *et al.*, “Ekstraksi Fitur Warna , Tekstur dan Bentuk untuk Clustered-Based Retrieval of Images (CLUE),” vol. 16, no. April, 2017.

[14] W. P. Hidayanti, “Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok El ektrik) pada ‘ Lombok Vape On ,’” vol. 3, no. 2, 2020.

[15] S. Nurjanah, A. M. Siregar, and D. S. Kusumaningrum, “Penerapan Algoritma K – Nearest Neighbor (Knn) Untuk Klasifikasi Pencemaran Udara Di Kota Jakarta,” *Sci. Student J. Information, Technol. Sci.*, vol. 1, no. 2, pp. 71–76, 2020.

[16] Florin Gorunescu, *Data Mining (Concepts, Models and Techniques)*. springer, 2011.