Identifying Cocoa ripeness using K-Nearest Neighbor (KNN) Method

by Demiawan Rachamatta

Submission date: 26-Dec-2022 11:01AM (UTC+0700)
Submission ID: 1986619623
File name: ntifying_Cocoa_ripeness_using_K-Nearest_Neighbor_KNN_Method.docx (93.7K)
Word count: 2100
Character count: 10877

2018 International Conference on Applied Science and Technology (ICAST)

Identifying Cocoa ripeness using K-Nearest Neighbor (KNN) Method

Hendra Yufit Riskiawan
Department of Information Technology
Politeknik Negeri Jember
Jember, Indonesia
yufit@polije.ac.id

Trismayanti Dwi Puspitasari
Department of Information Technology
Politeknik Negeri Jember
Jember, Indonesia
trismayantidwipuspitasari@gmail.com

Faik Inayatul Hasanah
Department of Information Technology
Politeknik Negeri Jember
Jember, Indonesia
faikinayatul@gmail.com

Nanang Dwi Wahyono
Department of Plantation Production
Politeknik Negeri Jember
Jember, Indonesia
nanang@polije.ac.id

M Fatoni Kurnianto
Department of Agricultural Technology
Politeknik Negeri Jember
Jember, Indonesia
ftnpolije@gmail.com

***Abstract*— There are three level maturity of cocoa namely mature, medium and raw. Nowadays for identification the process quality of maturity by manually, allowing errors to occur and take a time. So technology is needed to help identify cocoa maturity. One technology that can be used is digital image processing techniques by utilizing several techniques such as image segmentation and color extraction which aims to improve image quality. The process of identifying the quality of cocoa maturity using the K-Nearest Neighbor method (KNN). KNN is a method uses a supervised algorithm that stores all available cases and classifies new cases based on a similarity measure. The purpose of this algorithm is to identify new objects based on attributes and training data. The accuracy of this application is 100%.**

Keywords*— *cocoa, image processing, KNN, segmentation

I. INTRODUCTION

Cocoa (Theobroma cacaoL) is one of the plantation crops that has good economic value and a large market opportunity. The demand for cocoa has increased along with the increasing demand for raw materials for dry cocoa beans [1]. Based on International Coffee Organization(ICO) Indonesia is produced the fourth largest coffee in the world [2]. Market opportunities for these commodities are also increasingly open in line with the decline of the amount production experienced by other producing countries [3].

There are three level maturity of Cocoa namely mature, medium and raw. Nowadays for identification the process quality of maturity by manually, allowing errors to occur and take a time. So technology is needed to help identify cocoa maturity. One technology that can be used is digital image processing techniques by utilizing several techniques such as image segmentation and color extraction which aims to improve image quality.

Similar research was actually made, in his thesis entitled identification of the quality of maturity of cocoa using image processing and Fuzzy logic method but still has weaknesses, among others, the determination of the inference model must be appropriate [4]. To overcome the weakness the process of identifying the quality of cocoa maturity using the K-Nearest Neighbor method (KNN). KNN is a method uses a supervised algorithm that stores all available cases and classifies new cases based on a similarity measure. The purpose of this algorithm is to identify new objects based on attributes and training data. The purpose of designing and identifying the quality of ripeness of cocoa using image processing is to help cocoa farmers recognize the quality of maturity of cocoa.

II. LITERATURE REVIEW

Literature reviewthat used for this research are:

A. *Cocoa*

Cacao pods are the fruits of the cacao tree. Successfully pollinated flowers mature into this ribbed and oval fruit. The ripening process takes about five months. It is quite common to see both flowers and pods together on the same tree throughout the year. The thick shelled cacao pod contains "mucilage", a sweet white pulp that surrounds the bitter cocoa beans. Each tree produces about 20 pods and each pod contains between 20 and 60 seeds. In order to produce one kilograms of cocoa paste about 10 pods are required. Healthy mature pods will eventually rot on the tree unless picked by an animal or farmer. When the pods ripen they turn from green or yellowto orange or red [5]. Judging ripeness isn't as simple as looking at the color because there are so many different types of cacao, and the way they express their ripeness varies. They are different colors but all of them might be ripe, and the same is true of cacao.

B. *Image Processing*

Image segmentation is the separation of one object with another object in an image or between objects with a background contained in an image. With this segmentation position, each object in the image can be taken individually so that it can be used as input for other processes. Image segmentation is an important step in the image processing process. After the object is successfully segmented, the results of this segmentation will be used for further processes that can be carried out on an image, for example the process of image identification, the process of identifying objects or feature extractions in the image. The principle of image segmentation can be applied to static and dynamic images. Hue Saturation Lightness (HSL) make it easier to compare objects with similar hue but varying lighting conditions. the hue is a continuous angular scale of Hue where 0 and 360 degrees meet at red, so that red (360 degree) has maximum hue while orange has minimum [6]. Before we start processing images, we need to talk about how computers represent colors as numbers. RGB is an additive model works like light. For variables of class uint8 (and 8-bit images) [7].

978-1-5386-7548-9/18/$31.00 ©2018 IEEE

a. 0-255 are the possible integer values (same for the uint8 class!)

b. 0 is minimum for any RGB color

c. 255 is max for any RGB color

d. To define any color, you must specify the Red (R), Green (G), and Blue (B)

e. Grayscale images only need one value (0=black, 255=white)

C. K – nearest neighbor Method

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). A case is classified by a majority vote of its neighbors, with the case being assigned to the class most common amongst its K nearest neighbors measured by a distance function. If K = 1, then the case is simply assigned to the class of its nearest neighbor.

$$d_{ij} = \sqrt{\sum_{k=1}^{n}(x_{ik} - x_{jk})^2} \quad (1)$$

Where,

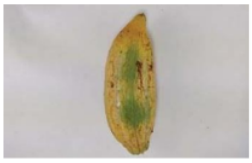
d_{ij} = distance between two vector I dan j; x_{ik} = testing data; x_{jk} = sampel data.

III. METHOD

A. Identification

Sampling of cocoa was carried out in three places, namely, Suci village, Peranti, Jember Regency, and Kaliwelang and Dampar village, Bades, Lumajang Regency. The cocoa obtained in all three places were 38 pieces, of which 23 were for training data consisting of 10 raw cocoa, 5 medium cocoa and 8 mature cocoa fruits. The samples obtained were not too many due to the non-season cocoa. The figure of cocoa

Raw cocoa



Medium cocoa

Mature cocoa

Fig.1. Cocoa image

B. Data Analysis

This application is designed to identify the maturity quality of cocoa objects based on RGB color (Red, Green, Blue). The results obtained from the application are in the form of information whether the cocoa fruit is raw, medium, or mature. Figure 2 is flowchart system

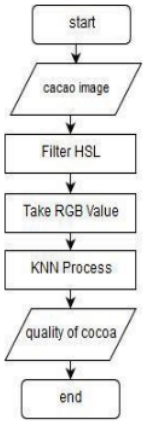


Fig. 2 Flowchart system

a. *Input Cocoa Image,* the inputted image is in the form of JPEG / JPG format with a size of 416 x 312. At the time of image sampling, a single cocoa fruit is placed in a box made of cardboard where the back and bottom are coated with white paper and the top, side right, and the left side is left open / perforated, because the part will be used as a place for lights where each lamp is 15 watts. The distance of image taking is around 25-30 cm. The scenario process for cocoa can be seen in Figure 3.

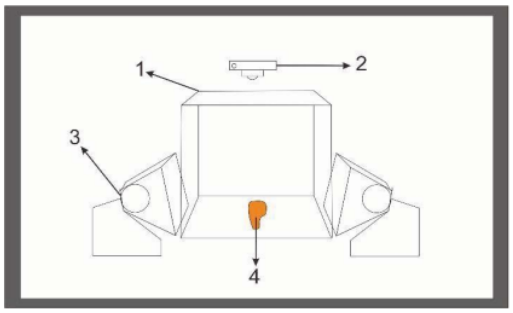


Fig. 3. Scenario for taking cocoa pictures

b. *HSL* feature, this process is to change the background color as a result of previous shots, which is the background color which was previously used as black.

c. *RGB,* this process is carried out to determine the color levels of Red, Green, Blue contained in the image that has been in the previous HSL filter.

KNN, there are four stages in the KNN process. Determining the value of k, the value of k here is the closest

value to be taken, such as k = 3 means that here it will sort 3 values of the nearest neighbor. Then measuring the distance of the training data using Euclidean Distance, the result of calculating the distance is still irregular so the next step is to sort the results of the shortest distance, when the distance is sorted then it will conclude whether the fruit is ripe, raw or half cooked.

IV. Result and Discussion

K - nearest neighbor identifies by taking the closest distance to the object that becomes the data test. This algorithm is used to determine the quality of cocoa fruit handling based on RGB values [8]. The advantages of this KNN algorithm are more effective against larger training data and can produce more accurate data. In the identification first determine the value of k, then calculate the datatest with the training data using the euclidien distance formula [9]. After that, do the data sorting from the calculation between training data and test data. The final step is to do the most voting on identification between the identification of the object's k value. The following are the final results of the program created in figure 4*:*

The test results on the quality program for the maturity of

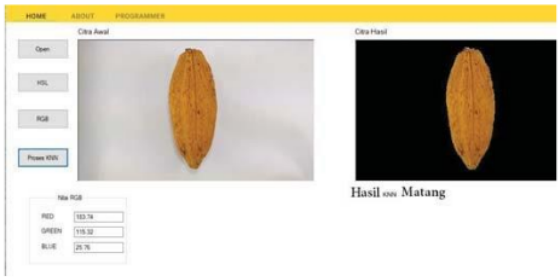


Fig. 4. Interface program

the cocoa fruit using test data of 15 samples with 5 raw samples, 5 medium samples and 5 mature samples. The training data used were 23 samples. The first test is testing raw samples of cocoa fruit using RGB color value parameters table I. The second test is testing medium samples table II and the last is testing mature samples table III.

Table I. The First Testing Raw Cocoa

Sample	Class Data	Result
1	Raw	Raw
2	Raw	Raw
3	Raw	Raw
4	Raw	Raw
5	Raw	Raw

In Table 1 shows that the testing process in the program can identify all test data totaling 5 samples. When calculated in percent, the following are:

$$Percentage\ raw = \frac{right\ data}{all\ data}\ x\ 100\% \quad (2)$$

$$= \frac{5}{5}\ x\ 100\%$$

$$= 100\%$$

Table II. The Second Testing Medium Cocoa

Sample	Class Data	Result
1	Medium	Medium
2	Medium	Medium
3	Medium	Medium
4	Medium	Medium
5	Medium	Medium

In table II shows that the testing process in the program can identify all test data totaling 5 samples. When calculated in percent, the following are:

$$Percentage\ medium = \frac{right\ data}{all\ data}\ x\ 100\% \quad (3)$$

$$= \frac{5}{5}\ x\ 100\%$$

$$= 100\%$$

Table III. The Third Testing

Sample	Class Data	Result
1	Mature	Mature
2	Mature	Mature
3	Mature	Mature
4	Mature	Mature
5	Mature	Mature

In table III shows that the testing process in the program can identify all test data totaling 5 samples. When calculated in percent, the following are:

$$Percentage\ mature = \frac{right\ data}{all\ data}\ x\ 100\% \quad (4)$$

$$= \frac{5}{5}\ x\ 100\%$$

$$= 100\%$$

V. Conclusion

Based on the problems that exist in the identification of the quality of cocoa fruit maturity, we can get conclusions: Identification of the quality of the maturity of the cocoa fruit using digital image processing is done by determining the RGB value of the image already in the previous HSL filter; Accuracy results from the identification of the quality of maturity of cocoa can identify all test data. 100% raw, 100% medium and 100% mature.

ACKNOWLEDGMENT

The researcher would like to express the gratitude and appreciation to Department of Information Technology, Politeknik Negeri Jember, who has helped support for this research.

REFERENCES

[1] Danial, D. et al. 2015. Increasing production and quality of cocoa through the Gernas movement in East Kalimantan. Proceeding. Vol 1. No 5. ISSN: 2407-8050. DOI:10.13057/psnmbi/m010542

[2] International Coffee Organization.2018. Total Production by All Exporting Countries. http://www.ico.org/. Access [August 24 2018]

[3] Amran. 2009. Evaluation Study of the National Movement for Increasing Production and Cocoa Quality (GERNAS KAKAO).

[4] Yogi, A. 2015. Application of Identification of Cocoa Fruit Maturity Quality Using Image Processing and Fuzzy Logic Method. Thesis, Univ. of Jember, Department Of Information System, Jember.

[5] ChocoMuseo. 2018. The World of chocolate. https://worldstandards.eu/chocolate%20-%20cacao.html Access [August 24 2018].

[6] Strand, R.-. Color and Compression. Center for Image Analysis. Swedish University of Agricultural Sciences Uppsala University.

[7] Glacier, M.-.Image Processing. http://www.appstate.edu/~marshallst/GLY3455/lectures/9_Image_Processing.pdf. Access [August 24 2018]

[8] Puspitasari, T.D., Sari, E.O., Destarianto, P. and Riskiawan, H.Y., 2018, January. Decision Support System for Determining Scholarship Selection using an Analytical Hierarchy Process. In *Journal of Physics: Conference Series* (Vol. 953, No. 1, p. 012119). IOP Publishing.

[9] Riskiawan, H.Y., Rizaldi, T., Setyohadi, D.P.S. and Leksono, T., 2017, September. Nitrogen (N) fertilizer measuring instrument on maize-based plant microcontroller. In *Electrical Engineering, Computer Science and Informatics (EECSI), 2017 4th International Conference on* (pp. 1-4). IEEE.

Identifying Cocoa ripeness using K-Nearest Neighbor (KNN) Method

ORIGINALITY REPORT

23%	20%	15%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ijtech.eng.ui.ac.id Internet Source	4%
2	slideplayer.com Internet Source	3%
3	www.slideshare.net Internet Source	3%
4	infor.seaninstitute.org Internet Source	3%
5	www.dandelionchocolate.com Internet Source	2%
6	www.it.uu.se Internet Source	2%
7	Submitted to Georgia Institute of Technology Main Campus Student Paper	2%
8	sipora.polije.ac.id Internet Source	2%

Submitted to Sekolah Tinggi Sandi Negara

9	Student Paper	2%
10	Kelvin Herwanda Tandrio, Samgadi Palgunadi Yohanes, Esti Suryani. "Prediction of Glucose Content and Level of Tomato Maturity with Backpropagation Based on Image Vector Feature", 2018 International Seminar on Application for Technology of Information and Communication, 2018 Publication	2%

Exclude quotes	On	Exclude matches	< 2%
Exclude bibliography	On		