Study on Physical Feature Extraction of Fermented Cocoa Bean with Digital Image Processing

Radi¹*, Anggoro C. Sukartiko², M. Prasetya Kurniawan², Agus Pamudji R.³, Gabriel C. Saragi¹, Bayu P. Bangun⁴, and Dena P. Putri²

¹Department of Agricultural and Biosystems Engineering, Faculty of Agricultural Technology, Universitas Gadjah Mada, Yogyakarta, Indonesia
²Department of Agro-Industrial Technology, Faculty of Agricultural Technology, Universitas Gadjah Mada, Yogyakarta, Indonesia
³Department of Food and Agricultural Products Technology, Faculty of Agricultural Technology, Universitas Gadjah Mada, Yogyakarta, Indonesia
Corresponding author:
E-mail: radi-tep@ugm.ac.id

Abstract. The study aimed to develop software for estimating the size of the fermented cocoa bean. The size of cocoa bean is one of the quality attributes needed in secondary cocoa processing, especially those related to the activities of sorting and preparing raw materials. This system was developed by adopting digital image processing techniques, starting from sample preparation and followed by capturing the image of cocoa bean samples, converting the color image into monochrome, segmentation, improving image results with a combination of dilation and erosion processes, to the feature extraction as the final stage. The extracted features of the beans’ size are length, width, circumference, and area of the samples, all measured in pixel unit. The features, called as estimated size, were then regressed to the real samples’ size, called as physical features, measured physically in the laboratory. The obtained regression formula showed a strong relationship between the estimated and physical features. The strong correlation between the two features enables the development of a rapid estimation of the physical features of cocoa beans based on digital image processing.

Keywords—fermented cocoa bean, physical parameter, estimation, digital image processing

1. Background
Different from coffee or tea, which are consumed in the form of extracts, cocoa is consumed in the form of a solution (suspense). Beside stimulating alkaloids, especially theobromine, cocoa products also contain several nutrients, such as fats, carbohydrates, and proteins. Unlike coffee and tea, cocoa must be consumed in large quantities in order to experience the stimulating effects [1]. Because of its distinctive flavor, the cocoa powder and its derivatives are very popular.

Ready-to-serve cocoa products are obtained from the processing of dried cocoa beans through a series of stages, including beans sorting, roasting, milling, and packaging. The sorting stage is the initial stage in the processing of secondary cocoa, which aims to select good beans to produce high quality of cocoa products. At the farmer level, this sorting stage is generally conducted manually, as depicted in Figure 1, illustrating the workers were sorting the fermented beans into several classes
according to the specified quality criteria. Generally, the fermented cocoa beans are divided into four categories, large, medium, small, and bad classes. The sorting is based on the visual appearance of the beans, which is undoubtedly difficult to be quantified, the sorted can vary due to elements of subjectivity, and can take long processing time.

Figure 1. Manual Sorting of Fermented Cocoa Beans

In contrast to the sorting system at the farm level, the cocoa processing industry generally uses standard sorting. In this case, a number of producers implemented a quality classification system normally used in its trading. In Indonesia, the classification and quality standards of cocoa beans are regulated by Indonesian National Standard SNI 2323-2008, which includes definitions, classifications, quality requirements, methods of sampling, and testing methods. Based on these standards, cocoa grains are grouped based on two approaches. The first approach uses bean size, where the beans are grouped into five classes, i.e., AA, A, B, C, S, ranged from showing the biggest/superior to the smallest classes [2]. The second approach uses the basis of seed defect rates, in which cocoa beans are classified into three classes. The standard has set quality attributes as a basis for sorting standard to reduce the risk of non-uniformity in the previous process. Although farmers and industries use different standards, both apply physical parameters for classification.

Because the manual sorting process has several weaknesses, this study aims to estimate the physical parameters of cocoa beans that are useful in developing a method of sorting cocoa beans quickly. The digital image processing method approach was used to develop the sorting system. This study focused on the initial study to see the potential for digital image analysis in estimating the physical quality parameters of cocoa beans, especially those related to size quality parameters.

Various efforts to quickly detect the quality of cocoa beans have been carried out. Some quality parameters used as a basis include physical parameters of seeds and fermentation index. One study to classify the quality of cocoa beans has been done by [3]. The study used the basis of image processing to identify the quality of cocoa beans, consisting of two parts, namely the image processing module and classification module. The image processing module extracts the image of cocoa beans to get the shape and color characteristics such as area, circumference, length, segment width, roundness, R, G, B, color values, hue, saturation, and intensity, while the classification module includes two artificial neural network structures to classify cocoa beans into damage categories and fermentation status. The test results showed that the classification level for damage criteria reached accuracy of 79.25%, which consisted of 84.38% accuracy for intact seed, 51.72% accuracy for broken seeds, and 98.29% for seed fraction, and 20% for damaged seeds. While the classification of fermentation status has an accuracy of 88.54%, which consists of 95.62% for fermented seeds, and 71.43% for unfermented cocoa beans.

Meanwhile, [4] investigated a quantitative procedure based on simple image analysis. The analysis applied color measurements and artificial neural networks (ANN) to predict the fermentation index of cocoa beans. The ANN model based on color measurement was tested to predict fermentation index (FI) of fermented cocoa beans. The RGB values were measured from the surface and middle area of fermented seeds on images obtained with a camera and desktop scanner. FI is defined as the ratio of
total free amino acids in fermented and non-fermented samples. The ANN model process inputs of the RGB value of cocoa surface and R/G ratio of wet cocoa beans to predict the FI. The model was able to predict the FI without statistical differences compared to the experimental value. The performance of the ANN model was evaluated by the coefficient of determination, Bland-Altman analysis and Passing-Bablok regression analysis.

Another study was carried out by [5], which examined nondestructive detection using FT-NIR spectroscopy. In this study, cocoa beans were prepared with fermentation variations for 0, 7, 14, and 21 days. The cocoa seeds are fermented by the heap method. The fermentation index of all treatments ranged from 0.535 and 1.242, the pH value between 4.26 to 6.13, and the total polyphenol levels varied between 6.48 and 15.58 mg/g. Samples were scanned with the NIR apparatus for spectral analysis. The results obtained from the conventionally analysis of cocoa quality was correlated with the wave spectrum for each treatment. The partial squared regression technique was then used to develop a calibration model. The scanned spectrum was then processed by vector normalization, multiplication correction scatter, and the first derivative technique. The calibration model turned out to be able to predict the total fermentation index and total polyphenols (R2>0.80), while for pH obtained the value R2<0.80. With this good result, this method can be used as an alternative of FI determination of fermented cocoa grains with advantages of detection time (<1 minute), more quickly than conventional methods (28 hours).

2. Material and Method

2.1. Sample Preparation and Image Capture
The sample was taken from the local farmers association at Gunungkidul, Yogyakarta. The measurements were carried out at the Energy and Agricultural Machinery Laboratory, Universitas Gadjah Mada, from April to November 2018. The sample used for this study was fermented dry cocoa beans. The beans consist of three categories, according to the sorting system at the farmer level, i.e., large, medium and small classes. The image was captured in a closed container, equipped with constant lighting, and a camera mounted on a sample board with a distance of 10 cm. The cocoa bean scanning box was made similar to [6]. To simplify the image processing stage, cocoa beans were placed on a sample board with a blue background. The image scanner wall was coated with a dark (black) cloth to minimize the amount of reflected light to the camera, as described in [7]. Other equipments used were vernier calipers and digital scales.

2.2. Research Procedures
In this study, several samples were taken randomly (each category as many as 60 beans) from the farmers. The cocoa beans were then taken to the laboratory for further analysis. Each cocoa bean was then taken into its image. Every bean was scanned once to produce an image with a bitmap format. The collection of images was then analyzed using a program that has been designed. Other equipment used was vernier caliper and digital scales which were used to measure the physical quality parameters of cocoa beans, i.e., length, width, circumference, area, and weight. Detail description is shown below, and the illustration is depicted in Figure 2.

Figure 2. Physical Parameter of Cocoa Bean
Grain length = $a$ \quad \text{[1]}
Grain width = $b$ \quad \text{[2]}
Grain circumference = $\pi \cdot a \cdot b$ \quad \text{[3]}
Grain area = $\frac{1}{4} \pi \cdot a \cdot b$ \quad \text{[4]}

Image analysis of cocoa beans aims to determine these physical parameters automatically. To get
the features in question, the image of cocoa beans was analyzed by a series of stages, including the
conversion of color images into gray images (monochroma), image segmentation, image
improvement, feature extraction. The conversion of color images into gray images was intended to
simplify the segmentation process. In this case, the gray image is formed from the blue channel (B).
While the segmentation process is carried out by the method of thresholding. The results of
thresholding generally still contain a number of unwanted noise. For this reason, the closing process
with a combination of dilation-erosion processes is applied to improve the segmented image. Image
features are then extracted from the repaired image. The stages of image processing are presented in
Figure 3. The image features are then analyzed for their relationship to the measurement parameters.

3. Results and Discussion

3.1. The physical feature extraction program of cocoa beans
The program for estimating the physical parameters of cocoa beans is presented in Figure 4. The
program was developed to analyse the image of cocoa beans offline. The program interface consists of
three part, namely the main panel which contains buttons for uploading images, setting parameter for
analysis and saving the result; the second part is image panel which includes three spaces (original
image, monochrome image, and segmented image); and a display panel to show extracted features.
Figure 4. The Interface of The Program for Extracting Physical Features of Cocoa Beans

3.2. Features of length and width of cocoa beans

The main physical parameters of cocoa beans are length and width. Based on the algorithm, those parameters measured in pixels were then compared to the real measurement feature. Both data of those features are plotted in a two-dimensional graph. The plotting result of the length parameter is shown in Figure 5, while data of cocoa width are scattered in Figure 6. The result shows that linear regression analysis obtains relation of length feature with equation of $y = 0.0047x + 0.5178$, while relation of length feature with equation of $y = 0.0033x + 0.56$. The result also shows that the regression coefficient of length feature ($R^2 = 0.5992$) is higher than the width feature ($R^2 = 0.2669$). It means that the proposed methods have the ability to extract length feature better than width feature. From the image data, noise of object shadow appears on the side of cocoa bean. Its may be caused due to the position of lighting of the chamber. From the Figure 5, we can see that a number of points deviates from other data. It may cause the low value of the regression coefficient.

Figure 5. Plotting Data of The Length Parameter of Cocoa Beans
3.3. Features of circumference and area of cocoa beans

Circumference and area of cocoa bean are other useful physical features. Those features are directly related to the size of cocoa beans, which is one parameter for the classification. In this case, circumference and area features are calculated by assuming that the grain is an ellipse (Figure 2). The relationship between the circumference data from measurement and the circumference data obtained from segmentation process is given in Figure 7. While relationship between area data from real measurement and from object segmentation is provided in Figure 8. Based on these results, it can be inferred that the circumference parameters provide a weak relationship indicated from the small regression coefficient. The low regression coefficient may be caused by different measurement method. In the y-axis component (Figure 7), data was obtained through circumference calculations based on eq. 3, while measurement of the feature on image processing approaches the number of outermost points of segmented object. For this reason, it is necessary to investigated more the both methods. Compare to the circumference feature, area parameter provides better distribution scatter. It has strong relation indicated from the value of the regression coefficient (\( R^2 = 0.6148 \)). It means that the number of pixels counted on the object strongly correlates with the area of cocoa beans.
4. Conclusion
The cocoa bean physical feature extraction program has been developed on the basis of digital image processing. These physical features include seed length, seed width, seed circumference, and seed area. The regression results between features extracted with feature measurements show a fairly varied distribution which causes low regression coefficients obtained. Based on the regression results, the feature area provides the best results with a regression coefficient of 0.615. The distribution of area data that is still varied is very likely caused by imperfect segmentation results. For this reason, further studies related to adaptive segmentation techniques will be developed.

5. Acknowledgements
This work was supported by the Ministry of Research, Technology and Higher Education of the Republic of Indonesia [Schema of PTUPT with no. of agreement: 1882/UN1/DITLIT/DIT-LIT/LT/2018]. We also would like to grateful to Universitas Gadjah Mada and Farmer Association of Ngudi Raharjo II, Plosokerep, Bunder, Patuk, Gunungkidul, Yogyakarta for their support.

References

[1] H.-D. Belitz, W. Grosch, and P. Schieberle, *Food Chemistry*. Springer Science & Business Media, 2009.
[2] Anonim, “Biji Kakao (SNI 2323:2008).” Badan Standarisasi Nasional, 2008.
[3] I. W. Astika, M. Solahudin, A. Kurniawan, and Y. Wulandari, “Determination of Cocoa Bean Quality with Image Processing and Artificial Neural Network,” *Computer Based Data Acquisition and Control in Agriculture*, 2010.
[4] N. León-Roque, M. Abderrahim, L. Nuñez-Alejos, S. M. Arribas, and L. Condezo-Hoyos, “Prediction of fermentation index of cocoa beans (Theobroma cacao L.) based on color measurement and artificial neural networks,” *Talanta*, vol. 161, pp. 31–39, Dec. 2016.
[5] S. Sunoj, C. Igathinathane, and R. Visvanathan, “Nondestructive determination of cocoa bean quality using FT-NIR spectroscopy,” *Comput. Electron. Agric.*, vol. 124, pp. 234–242, Jun. 2016.
[6] Radi, Muhammad Rivai, and Mauridhi Hery Purnomo, “Combination of First and Second Order Statistical Features of Bulk Grain Image for Quality Grade Estimation of Green Coffee Bean,” *ResearchGate*, vol. 10, no. 18, pp. 8165–8174, Oct. 2015.
[7] F. Faridah, G. O. F. Parikesit, and F. Ferdiansjah, “Coffee Bean Grade Determination Based on Image Parameter,” *TELKOMNIKA Telecommun. Comput. Electron. Control*, vol. 9, no. 3, pp. 547–554, Dec. 2011.