Detection of Coffee Bean Damage in The Roasting Process Based on Shape Features Analysis

I G A Gunadi¹, I P M K Artha², I G P Christyaditama², G A S Wicaksana², and I M Martina²

¹ Departement of Physics and Natural Science teaching - Universitas Pendidikan Ganesha, ² Magister program of Computer Science - Universitas Pendidikan Ganesha

Abstract. Many studies raise the issue of analysis on the quality of coffee beans (green beans). Green beans are the term used to indicate coffee beans before roasting. The importance of analyzing the quality of coffee beans not only in coffee beans but also in rice seeds. Rice seeds are used as a mixture in the roasting process of coffee. Analysis of the quality of roasted coffee is very important. In roasted coffee beans can be damaged and changes to the coffee beans. Some types of damage to coffee beans include, blackened seeds, partially blackened seeds, browned seeds, broken seeds, perforated seeds, epidermis seeds, horned skin seeds, too young seeds, spotted seeds, and dirty seeds. This study focuses on one of the damage factors, namely broken coffee beans after broken beans. In this study, analysis of roasted coffee quality was determined by shape analysis. The shape analysis uses elongation values. There are two categories of classification, the first class of coffee is of good quality, and the second is coffee with poor grade. Classification is good if the elongation value is more than 0.85, for other conditions it is classified as bad. The results of this study have the following accuracy values, at coffee with a classification of good accuracy of 93.4%, while those at coffee with a bad classification of accuracy are 78.5%.

Keywords. digital image analysis, damage to coffee beans, shape features

1. Introduction

Indonesia is the fourth of biggest coffee producer after Brazil, Vietnam, Columbia, and one of the largest robust coffee bean producer in the world [1]. At present, the Ministry of Trade and Trade of the Republic of Indonesia uses the Defective System stipulated in the Indonesian National Standard (SNI) for coffee beans, No: 01-2907-1999. The renewal of standards for the quality of coffee beans by applying SNI aims to make quality assurance in the coffee bean industry more in line with quality standards in many countries where products will be exported [2].

Digital image analysis is an investigation of a digital image through certain processes to find out certain features of an image or in other words called separation of features [3]. There are several features that can be used in digital image processing, including: shape, size, geometry, texture, and color. After going through the image separation and feature extraction from an image, the process is usually continued for the pattern recognition process where the process is a process that has an output
description of an image, but does not rule out the process of stopping the process until the separation of features as digital image output only. In the pattern recognition process, feature extraction can be referred to as preprocessing and the feature separator is called low level image processing. The pattern recognition process really requires the feature extraction process because it is not possible to use pattern recognition methods with a single digital image that is intact and raw, although it allows it to take up a lot of memory and it is difficult to recognize patterns from an image, therefore feature extraction is needed.

Research in the field of digital image analysis has increasingly developed in recent years. Image analysis about the quality and damage of coffee beans is one that has been developed by several researchers, such as Faridah, et al [4] who have examined the quality of coffee beans determination based on image parameters. The research was carried out by extracting data parameters of rice coffee image in the form of texture (energy, entropy, contrast, homogeneity) and color (RGB) and determining the quality of coffee beans based on image parameters by applying artificial neural network algorithms, with accuracy reaching 100%. Afonso C., et al [5] have examined the classification of coffee beans based on the type of damage using the Support Vector Machine (SVM) method. The research uses image input of processed coffee beans using image processing and machine learning techniques namely SVM to classify coffee beans based on the type of damage: size, color, shape, etc. with an accuracy of 97%. In the research of Faridah, et al and Afonso C., et al, there is an equation that is the data used is the image of rice coffee beans (green beans). However, in the industrial world that is not enough. The importance of analyzing the quality of coffee beans is not only on the rice coffee beans, but also important to re-analysis at the roasting stage. It aims to help the roasters (the designation for the work of people roasting coffee beans) in the process before packaging the product for sale. In another sense, this research will focus on contributing to the problem above.

This study aims to develop a digital image analysis system that can determine the quality of roasted coffee beans, in terms of detecting coffee beans that are feasible and not suitable for consumption and sold as specialty coffee as contained in the Standards of classification of coffee beans provided by SNI No. 01-2907-1999. This research focuses on one of the damage factors, namely broken coffee beans after the roasting process (broken beans). At the end of this study, it is expected to produce a precise and accurate analysis in detecting damage to roasted coffee beans based on images. Good picture of one roasted coffee beans, and pictures with many roasted coffee beans at once.

2. Theoretical Review

2.1. Digital Image Processing
Digital image processing is a process that aims to manipulate and analyze images with the help of computers. Digital image processing can be grouped into two types of activities, namely: improving the quality of an image, so that it can be more easily interpreted by the human eye and processing information contained in an image for the purpose of recognizing objects automatically. The second field of application is very closely related to pattern science (pattern recognition) which generally aims to recognize an object by extracting important information contained in an image. When pattern recognition is associated with image processing, it is hoped that a system will be able to process the input image so that the image can be identified. This process is called image recognition. This image recognition process is often applied in everyday life. Image processing and pattern recognition become part of the process of image recognition. Both of these applications will complement each other to get the characteristics of an image to be recognized. In general, the stages of digital image processing include image acquisition, image quality improvement, image segmentation, representation and description, introduction and interpretation.
2.2. **Image Acquisition**
Data retrieval can be done using various media such as analog cameras, digital cameras, camcorders, scanners, optical readers, and so on. The resulting image is not necessarily digital data, so it needs to be digitized.

2.3. **Image Quality Improvement**
At this stage, it is known as pre-processing which in improving image quality can increase the likelihood of success at the next digital image processing stage.

2.4. **Image Segmentation**
Segmentation aims to select and isolate (separate) an object from the whole image. Segmentation consists of down sampling, filtering, and edge detection. The down sampling stage is a process to reduce the number of pixels and eliminate some information from the image. With fixed image resolution, down sampling results in smaller image sizes. The next segmentation stage is filtering with a median filter, this is done to eliminate the noise that usually appears at high frequencies in the image spectrum. In filtering with filtering, the gray level of the image in each pixel is replaced by the median value of the gray level in the pixels contained in the filter window. The last step in the segmentation process is edge detection. The Canny algorithm approach is based on the convolution of the image function with the Gaussian operator and its derivatives. This edge detector is designed to represent an ideal edge, with the desired thickness. In general, the segmentation process is very important and direct.

3. **Research Methodology**
In this study, image analysis of roasted coffee bean damage detection uses feature extraction. Before the image is detected, the image to be processed applies as test data to go through several stages including: gray scaling, thresholding, feature extraction, analysis, and accuracy testing as shown in Figure 1.

![Figure 1. Architectural Analysis of Digital Image Damage in Roasted Coffee Beans](image)

3.1. **Data Collection and Acquisition**
Data input is a 2D digital image of roasted coffee beans. Coffee bean samples were from the Arabica type. The first was named Halu Pink Banana, which originated from Mount Halu, West Java, with a honey process and roasted medium roast profile. The second named Solok comes from West Sumatra, with a semi washed process and a roasted medium roast profile. Roasted coffee beans are placed on white paper and images are taken with an iPhone 6 mobile camera. The camera is placed 30 cm above the surface of the object. The front and back sides of the seeds are taken. The brightness of the lighting
is assisted by using natural sunlight and camera flash. Data collection and acquisition of roasted coffee beans image used as training data, a number of eight images that represent roasted good coffee beans “OK” (Figure 2) and four images that represent the damaged roasted coffee bean “BAD” (Figure 3).

Based on the training data of a good roasted coffee bean “OK” (Figure 2), can be made in tabular form as in Table 1 below. Where the lowest metric value is 0.85 which will be the lowest standard.

**Table 1. Data training of good roasted beans “OK”**

| Title                          | Metric |
|--------------------------------|--------|
| Halu_JawaBarat_07.jpg          | 0.85   |
| Halu_JawaBarat_06.jpg          | 0.91   |
| Solok_SumateraBarat_09.jpg     | 0.85   |
| Halu_JawaBarat_04.jpg          | 0.89   |
| Halu_JawaBarat_03.jpg          | 0.91   |
| Halu_JawaBarat_02.jpg          | 0.87   |
| Halu_JawaBarat_01.jpg          | 0.89   |
| Solok_SumateraBarat_15.jpg     | 0.86   |

Based on the training data on the damaged coffee beans “BAD” (Figure 3), can be made in tabular form as in Table 2 below.

**Table 2. Data training of damage roasted beans “BAD”**

| Name                           | Metric |
|--------------------------------|--------|
| Solok_SumateraBarat_10.jpg     | 0.73   |
| Solok_SumateraBarat_17.jpg     | 0.72   |
| Solok_SumateraBarat_03.jpg     | 0.83   |
The training data will be used as a comparison value during the image analysis process of the test data. Test data are obtained in the same way, then entered into the system. Test data can be in the form of one roasted coffee beans or many roasted seeds in one picture. The process of digital image analysis to detect damage to the results of roasted coffee beans can be seen in Figure 1 Architecture of Digital Image Analysis of Damage to Roasted Coffee Beans.

### 3.2. Grayscale
The original image as test data is entered into the system. The original color of the image will be converted from Red, Green, Blue (RGB) to black and white (grayscale). Grayscale image is a digital image that only has one known value in each pixel. This value is used to indicate the level of intensity. Colors that are owned are colors from black, gray and white. Gray level here is the color of ash with various levels from black to near white. The purpose of calculating gray level is to facilitate the next process, namely the thresholding process. The results of this process can be seen in Figure 4.

![RGB to grayscale conversion](image)

**Figure 4.** RGB to grayscale conversion

### 3.3. Thresholding
Thresholding is changing an image into a binary image. Thresholding looks at each pixel then decides whether it is made white (255) or black (0). This decision is generally made by comparing the pixel numerical values with certain values called thresholds. If the pixel value is smaller than the threshold, then the pixel is changed to 0, otherwise the other is changed to 255. This can also be done vice versa. Thresholding is the process of simplifying images from gray levels into binary colors so that based on their gray levels the pixels are divided into backgrounds and objects of interest. The purpose of thresholding is to separate objects from the background. This is done by changing the intensity of the pixels of an existing image to only 2 intensities, namely black and white. The results of this process can be seen in Figure 5.

![Thresholding](image)

**Figure 5.** Thresholding

### 3.4. Extraction of Shape Features using Metric
Shape feature extraction is a feature used to distinguish one object form from another object [6]. This study uses one form feature extraction technique, namely metric. Metric is the value of the ratio between area and circumference of the object. Metric has a range of values between 0 to 1. Objects that are elongated or close to the shape of a straight line, their metric values are close to 0, while objects that are round or circular, their metric values are close to 1. The calculation of metrics can be seen in (1).

\[
M = \frac{4\pi A}{c^2}
\]  

(1)
Where $M$ is the metric value, $A$ is the area, and $C$ is the circumference or length of the object's side. This metric technique was chosen because the shape to be searched for is the shape of a circle [6]. The results of this process can be seen in Figure 6.

![Figure 6. Feature extraction results](image)

3.5. Analysis
In the process of analyzing digital images, a check on the lower threshold of the specified metric value is based on training data, which is 0.85. Based on the feature extraction process, the metric value obtained from the test data is 0.87. Because 0.87 is more than 0.85, the test data is declared “OK” as a good roasted coffee bean. Can be described into branches as follows: If (metric > 0.86) then “OK” Else “BAD”. The results of the digital image analysis process damage to coffee beans can be seen in Figure 7.

![Figure 7. Results of roasted bean damage analysis](image)

4. Results and Discussion
The application developed in this research will be implemented using Matlab 2016. This research produces an application that can perform digital image analysis to detect damage to the results of roasted coffee using shape feature extraction. Can be used to detect one coffee bean in an image and many coffee beans in one picture at a time. Objects that have been identified will be labeled as good roasted coffee beans “OK” or broken roasted coffee bean “BAD”. To get results, it must go through several stages, namely: entering test data, converting the color of the test data using grayscale techniques, segmenting to separate objects from their background and detecting the number of objects using thresholding techniques, and finally analyzing the results of feature extraction using metric values for determine whether the roasted coffee bean results include “OK” or “BAD”. Figure 8 shows the results of digital image analysis to detect damage to the results of roasted coffee beans using the extraction of shape features on 1 object. The test data used is the image of roasted Arabica Halu Pink Banana coffee beans from West Java.
Figure 8. Results of digital image analysis for detecting good roasted bean “OK” using shape feature extraction on 1 object

In Figure 8, the application successfully detects and analyzes the roasted coffee bean correctly that is “OK”. That is because the metric value of the object is 0.87 (more than 0.85). Subsequent experiments used the image of the broken Solok West Sumatra coffee bean as a test data to detect damage to the roasted coffee bean results using the extraction of shape features on 1 object as shown in Figure 9.

Figure 9. Results of digital image analysis for detecting damaged roasted bean “BAD” using shape feature extraction on 1 object

In Figure 9, the application successfully detects and analyzes the roasted coffee bean correctly namely “BAD”. That is because the metric value of the object is 0.72 (less than 0.85). Subsequent experiments used the image of heterogeneous Solra West Sumatra coffee beans as test data to detect damage to roasted coffee bean results using the extraction of shape features to 9 objects with regular spacing as shown in Figure 10.

Figure 10. Result of digital image analysis for detecting multiple objects with regular distance
In Figure 10, the application successfully detected and analyzed the roasted coffee bean correctly namely “OK” as many as 5 objects because the metric value is more than 0.85 and “BAD” as many as 4 objects because the metric value is less than 0.85. Subsequent experiments used the image of heterogeneous and randomly random Solok coffee beans in West Sumatra as test data to detect damage to roasted coffee bean results using the extraction of shape features to 97 objects with irregular spacing as shown in Figure 11.

Figure 11. Result of digital image analysis for detecting multiple objects with irregular distance

Results in Figure 11, the application still cannot detect and analyze coffee beans with 100% correct. Of the supposedly 90 good roasted coffee beans only 83 were labeled “OK” and out of 7 broken or damaged roasted coffee bean there were 10 that were labeled “BAD”. The accuracy of the results of digital image analysis to detect coffee beans are measured in several stages. The level of accuracy is done to measure the level of success of the analysis conducted. Based on 107 objects of the 3 test images used there are several objects in the image that have not been detected correctly. This accuracy test includes the accuracy of the segmentation stage using thresholding to find out how accurately the system can recognize the number of objects in an image, the accuracy test of the “OK” roasted coffee bean analysis, and the accuracy test of the “BAD” roasted coffee bean analysis so that in each image the object is detected can be labeled to indicate which coffee beans are good and which are damaged.

Table 3. Segmentation results

| Data Testing                  | Total Object | Identified Object | Segmentation Result |
|-------------------------------|--------------|-------------------|---------------------|
| Halu_JawaBarat_05.jpg         | 1            | 1                 | True                |
| Solok_SumateraBarat_01.jpg   | 9            | 9                 | True                |
| Solok_SumateraBarat_02.jpg   | 97           | 93                | False               |

Based on the results of segmentation accuracy testing of the division of objects in each image or image in Table 3 produces an accuracy of 96.3%. Of the 3 images (data testing) processed there are 107 objects, consisting of 103 objects that can be recognized and 4 objects that cannot be recognized.
Unrecognizable objects are in test data number 3 because there are many objects that are irregularly spaced and in contact with each other.

Table 4. Accuracy analysis of good roasted beans “OK”

| Data Testing                | “OK” | “OK” Identified | Analysis Result |
|-----------------------------|------|-----------------|-----------------|
| Halu_JawaBarat_05.jpg       | 1    | 1               | True            |
| Solok_SumateraBarat_01.jpg | 5    | 5               | True            |
| Solok_SumateraBarat_02.jpg | 90   | 83              | False           |

Table 4 shows the results of the accuracy testing of roasted coffee bean analysis labeled “OK”. From the image test data 1 and image 2 produces 100% accuracy, while the image test data 3 there are some objects that are not detected. Overall, of 105 “OK” roasted coffee bean objects resulted in an accuracy of 93.4%, the system still incorrectly analyzed 7 objects and 98 other objects were successfully analyzed with the label “OK”. As for the analysis of roasted coffee beans with the label “BAD” or damaged roasted beans can be seen in Table 5.

Table 5. Accuracy analysis of damage roasted beans “BAD”

| Data Testing                | “BAD” | “BAD” Identified | Analysis Result |
|-----------------------------|-------|-----------------|-----------------|
| Halu_JawaBarat_05.jpg       | 0     | 0               | True            |
| Solok_SumateraBarat_01.jpg | 4     | 4               | True            |
| Solok_SumateraBarat_02.jpg | 7     | 10              | False           |

Based on Table 5, the accuracy test of roasted coffee beans “BAD” yields an accuracy of 78.5%. Not because of inaccurate training data, but because there are several coffee beans that touch each other are detected as one object so that they are detected as damaged coffee beans “BAD”, as in the image test data 3. Overall accuracy test results can be seen in Table 6.

Table 6. Accuracy results

| Testing Object | Accuracy   |
|----------------|------------|
| Thresholding Segmentation | 96.3%     |
| Roasted beans as “OK”      | 93.4%     |
| Roasted beans as “BAD”     | 78.5%     |

5. Conclusions

The results of this research succeeded in detecting damage to coffee beans in the form of broken beans based on test data in the form of roasted coffee beans or images. Can be used to detect one coffee bean in an image and many coffee beans in one picture at a time. The results of this assessment have an accuracy value of detecting the number of objects in an image test data of 96.3%, an accuracy value of “OK” roasted coffee bean analysis of 93.4%, and an accuracy value of “BAD” roasted coffee bean analysis of 78.5%. This research still has weaknesses in the segmentation process using thresholding. Because of the image used cannot detect images of objects that touch each other. In the future, this research can be developed by adding extraction of other features such as texture, color, and size to produce higher accuracy values in classifying the quality of coffee beans.

References

[1] International Coffee Organization 2011 Trade Statistics of Total Production of Exporting Countries www.ico.org
[2] Mulato S 2006 Primer and Secunder Coffee Product Processing (Jember: Pusat Penelitian Kopi
dan Kakao Indonesia)
[3] Darma P 2010 Digital Image Processing (Yogyakarta: ANDI OFFSET)
[4] Faridah, Gea O F P, Ferdiansjah 2011 Coffee Bean Grade Determination Based on Image Parameter TELKOMNIKA 9(3) pp 547–554
[5] Carmelita A, Hidekazu F U K A I 2018 Classification of Coffee Bean Images Based on Defect Types using SVM TAJST 1 pp 181–187
[6] Adi P 2017 Image Characterisitcs Extraction https://pemrogramanmatlab.com/pengolahan-citra-digital/ekstraksiciri-citra-digital