Semi-automatic Ground Truth Image Construction for Coffee Bean Defects Classification Based on SNI 01-2907-2008

Made Windu Antara Kesiman[1](✉) and Ismail Sulaiman[2]

[1] Virtual, Vision, Image, and Pattern Research Group, Faculty of Engineering and Vocational, Universitas Pendidikan Ganesha, Bali, Indonesia
antara.kesiman@undiksha.ac.id

[2] Program Studi Teknologi Hasil Pertanian, Faculty of Agriculture, Universitas Syiah Kuala, Banda Aceh, Indonesia

Abstract. In the coffee bean test procedure for determining the value of defects, physical separation of defective beans is carried out. This physical test is carried out using human senses or using assistive devices. It is necessary to have a system that can perform the identification and selection process of defective coffee beans automatically. To develop an automation system for determining the value of coffee bean defects based on SNI 01-2907-2008, a semi-automatic ground truth image construction is needed. In this study, a coffee bean storage device was designed, then a digital image is taken using a standard mobile phone camera. The digital image processing steps will be carried out. The results show that the proposed method is able to construct and to extract optimal number of the image samples of coffee bean for each type of defects. All extracted image samples of coffee bean defects will serve as a new dataset for the future automation system for determining the value of coffee bean defects based on SNI 01-2907-2008.

Keywords: Dataset · Coffee bean · Defect · Classification

1 Introduction

Coffee is one of the prima donna commodities that contribute to the export value of Indonesian products to foreign countries. In addition to bringing high economic value, coffee farming in Indonesia is also inseparable from the content of the cultural values of each farmer and coffee-producing area in Indonesia. To maintain and improve the quality of Indonesian coffee beans, the government through the National Standardization Agency (BSN) has issued the Indonesian National Standard (SNI) 01-2907-2008 Coffee Beans [1]. This SNI for Coffee Beans is a revision of SNI 01-2907-1999 Coffee Beans, which aims to harmonize Indonesian coffee standards with world coffee quality standards. The main improvement made in this revision is in terms of coffee quality requirements. This standard stipulates the classification and quality requirements, methods of testing, marking, and packaging of robusta and arabica coffee beans. One of the quality requirements for coffee beans stipulated in the SNI for Coffee Beans is a special

J. Sumantyo et al. (Eds.): ICoSIA 2022, ABSR 29, pp. 453–463, 2023.
https://doi.org/10.2991/978-94-6463-122-7_43

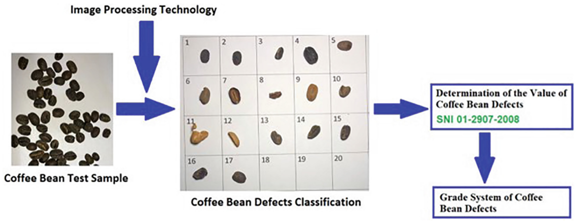

Fig. 1. Automation system for determining the value of coffee bean defects based on SNI 01-2907-2008.

quality requirement based on a defective value system. Based on this system, the quality of the coffee beans is classified, from Quality 1 to Quality 6 with the requirements for the range of defects in the sample of coffee beans being tested. Determination of the number of defects is carried out by identifying the type of coffee bean defect found and calculating the amount of the defect value according to the type of defect in the coffee bean. In SNI Coffee Beans, a table for determining the value of coffee bean defects has been determined, accompanied by an equivalent description of the type of defect and the value of the defect.

The SNI for Coffee Beans stipulates 20 (twenty) types of coffee bean defects, with a description of the types of defects that include the color of black, brown beans, the discovery of coffee logs, coffee husks, horn-skinned beans, broken beans, holes, and the discovery of twigs, soil, or seeds. Stones on the tested coffee bean sample. The process of identifying the type of coffee bean defect is usually carried out by the coffee industry's quality control team. Because this process is done manually by humans, of course the results are strongly influenced by human subjectivity factors which are highly dependent on the level of expertise, work experience and cannot be separated from the level of physical and mental fatigue at work. To overcome these problems, it is necessary to develop a system that can determine the value of coffee bean defects based on SNI 01-2907-2008 automatically (Fig. 1). Because the process of identifying the type of coffee bean defect is based on several visual characteristics of the coffee bean such as color, hole, shape and the presence or absence of foreign objects, digital image processing technology or computer vision in general will be the main backbone of the system to be built.

This work is under the scheme of Project COFFECTION (Coffee Defect Identification) by Virtual, Vision, Image, and Pattern Research Group (VVIP-RG) (https://research.undiksha.ac.id/vvip-rg/), Universitas Pendidikan Ganesha. This project aims to develop an automation system for determining the value of coffee bean defects based on SNI 01-2907-2008. The general objective of this research is to explore digital image processing methods to identify types of seed defects. The specific objective of this research is to design a prototype of an automation system for determining the value of coffee bean defects based on SNI 01-2907-2008. One part of this research will explore the semi-automatic ground truth image construction for coffee bean defects classification based on SNI 01-2907-2008. This research is expected to provide benefits for the development of digital image processing science and technology in general, and of course it is hoped that it can improve the work quality of the Indonesian coffee bean industry for world quality. This paper will be organized as following. Section II briefly describe the corpus of coffee bean. The explored method for this research is presented in Section III. Section IV shows the experimental results and findings. And finally, Section V conclude this research results.

2 Coffee Bean: Corpus and Previous Works

Indonesia is known as one of the world's coffee producers. To participate in the development of the global market, improving the quality of Indonesian coffee must go through the implementation of quality standards. The Indonesian National Standard for Coffee Beans was then compiled. The scope of the SNI stipulates the classification and quality requirements, methods of testing, marking, and packaging of robusta and arabica coffee beans [2].

There are several criteria for classifying coffee [2]. Based on the type of coffee, coffee is classified into Robusta and Arabica coffee. Based on the method of processing, it is classified into dry and wet processing. Based on the size, dry-processed robusta coffee is classified into large and small, wet-processed robusta coffee is large, medium, and small, while Arabica coffee is large, medium, and small. Based on the number of pieces of seeds, classified into Peaberry and Polyembryoni. Based on the value of the defect, it is classified into six quality levels, where specifically for robusta coffee, quality 4 is divided into sub-levels of quality 4a and 4b. In this study, the focus of the analysis and design of the system development is the classification based on the value of the defect. The defect value test is carried out on Ready coffee beans or ready for Export to determine the quality or grade of the coffee.

Based on national quality standards, coffee beans are classified into 6 different grades according to the value of defects in the coffee beans (Table 1 and Table 2).

In the coffee bean test procedure for determining the value of defects, physical separation of defective beans is carried out. This physical test is carried out using human senses or using assistive devices. At this stage, it is necessary to have a system that can perform the identification and selection process of defective coffee beans automatically. Because it involves the identification of the visual features of defective coffee beans, digital image processing technology is a very appropriate technology to be used as the basis for the support for the automated determination of coffee bean defect value system to be developed.

Several studies on the identification and classification of coffee bean quality and defects have been carried out previously. Ilhamsyah, et al. (2021) [3] have proposed a method for classifying the quality of coffee beans using a multilayer perceptron based on LCH color features. The study used 12 grades/quality of coffee beans. However, the definition of the 12 grades of coffee beans has not been adjusted to the SNI grade for coffee beans. In this study, four machine learning methods were tested, namely Naïve Bayes, Decision Tree, SVM, Multilayer Perceptron, and using seven color features. Aramiko, et al. (2020) [4] proposed a method of classifying damage to coffee beans using the Naïve Bayes Classifier method. However, there were only 4 types of defect classes tested in the study, namely broken seeds, full black seeds, partially black seeds, and broken black seeds. Saputra, et al. (2020) [5] conducted a process of identifying

Table 1. Grade System of Coffee Bean Defects (SNI 01-2907-2008)

Grade	Defect value
Grade 1	0–11
Grade 2	12–25
Grade 3	26–44
Grade 4a	45–60
Grade 4b	61–80
Grade 5	81–150
Grade 6	151–225

Table 2. Determination of the Value of Coffee Bean Defects (SNI 01-2907-2008).

Type of Defect	Value of Defects
1 (one) full black bean	1 (one)
1 (one) partial black bean	1/2 (half)
1 (one) broken black bean	1/2 (half)
1 (one) pod/cherry	1 (one)
1 (one) brown bean	1/4 (quarter)
1 (one) large shell skin	1 (one)
1 (one) medium shell skin	1/2 (half)
1 (one) small shell skin	1/5 (fifth)
1 (one) parchment bean	1/2 (half)
1 (one) large husk	1/2 (half)
1 (one) medium husk	1/5 (fifth)
1 (one) small husk	1/10 (tenth)
1 (one) broken/chipped bean	1/5 (fifth)
1 (one) immature bean	1/5 (fifth)
1 (one) bean with slight insect damage	1/10 (tenth)
1 (one) bean with severe insect damage	1/5 (fifth)
1 (one) bean with fungus damage	1/10 (tenth)
1 (one) foreign matter (large size branch or stone)	5 (five)
1 (one) foreign matter (medium size branch or stone)	2 (two)
1 (one) foreign matter (small size branch or stone)	1 (one)

the quality of Arabica coffee beans based on defects using the Convolutional Neural Network technique. However, the test was only carried out using a 2-class and 4-class model. Wallelign, et al. (2019) [6] also used CNN to determine 12 grades of coffee beans from Ethiopia. Arboleda, et al. (2018) [7] specifically propose an image processing technique with color features to identify coffee beans that are classified as black beans. Pizzaia, et al. (2018) [8] also used the MLP classifier to classify good beans and bad beans in Arabica coffee. Portugal-Zambrano, et al. (2016) [9] built a special device to detect 13 types of defects in coffee beans. Huang, et al. (2019) [10] proposed the process of classifying coffee beans in real time using CNN. Sanchez-Aguiar, et al. (2019) [11] proposed a technique to separate good and young seeds using the HSV and YCbCr color spaces.

For this research, the coffee bean sample came from Kintamani Arabica coffee, taken from Belantih Mangani coffee producer at an altitude of 1200 m above sea level, with type of Arabica copiol which was processed by wet processing.

3 Method: Semi-automatic Ground Truth Image Construction

In this study, a coffee bean storage device was designed in the form of a tray-like container that has holes the size of one coffee bean on average. The device was then implemented using the Creality CR Max 10 three-dimensional (3D) printing tool (Fig. 2). This special device is deliberately designed to simplify and speed up the process of optimally sampling the image of a large number of coffee beans. Besides that, it is also expected to facilitate the digital image processing process to extract and identify the shape and size of coffee beans.

Samples of coffee beans will be placed evenly on the device, then a digital image is taken using a standard mobile phone camera (Fig. 3). In the digital image from the coffee bean sampling, the following digital image processing steps will be carried out: the process of extracting the blue channel from the RGB colour space in the digital image, the grey level thresholding process with a parameter value of 127 from the blue channel image, the median filter process with a size filter 11, the process of masking the original image with the binary image of the median filter, and the process of forming a bounding rectangle/box to extract per coffee bean.

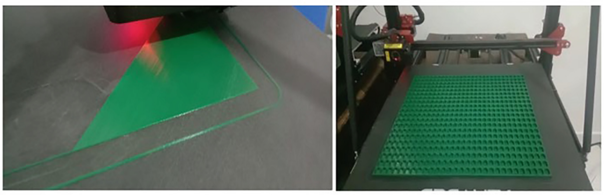

Fig. 2. The initial process of making a device with 3D printing.

Fig. 3. Sample images of coffee bean with defects capturing process.

4 Experimental Results and Findings

The capturing process for one type of defect was performed twice with different configuration of the coffee bean on our tray-like container (Fig. 4).

4.1 Sequence of Image Processing

In Fig. 5 and 6, the results of digital image processing from each processing stage are shown on the digital image of the coffee bean sample. The blue channel of the RGB color space provides an excellent grayscale image for separating the color of the coffee bean from a green set that was originally designed (Fig. 5b). This is proven by the results of the next thresholding process which can very well separate the existing coffee bean area. Several small points/areas of noise in the image resulting from the thresholding process are formed from the point/area of the device that receives higher illumination, due to the reflection of the device material on the side of the coffee bean hole (Fig. 5c). However, in the next process, namely the median filter, these noises can be removed to produce a very clean binary image of coffee beans (Fig. 6a). By producing this excellent binary image of the coffee bean area, the masking process (Fig. 6b) for extracting the coffee bean area with a bounding rectangle/box becomes very optimal (6c).

4.2 Coffee Bean Sample Images Extraction

The final results show that most coffee beans can be extracted well (Fig. 7 and 8). Number of extracted image samples of coffee bean defects is presented in Table 3. From the results of the automatic extraction of coffee bean images, there are still some unsuitable image extraction results. Then the validation process was carried out again manually by the coffee bean expert to determine and select the appropriate coffee bean images. The coffee bean image that has been validated by the expert will be the image of the coffee bean sample for the coffee bean defects dataset.

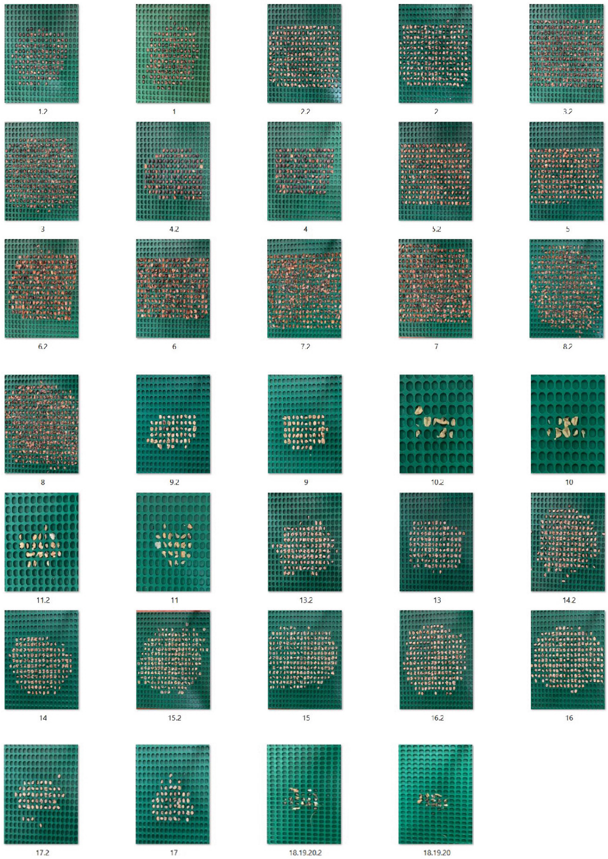

Fig. 4. Captured image for each type of coffee bean defect.

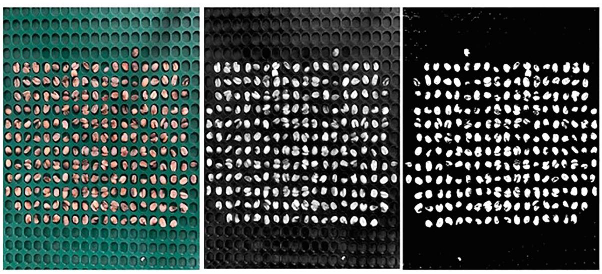

Fig. 5. **a**) The original input image, **b**) The image resulting from the blue channel extraction process from the RGB color space in the digital image, **c**) The image resulting from the gray level thresholding process with parameter value 127 from the blue channel image.

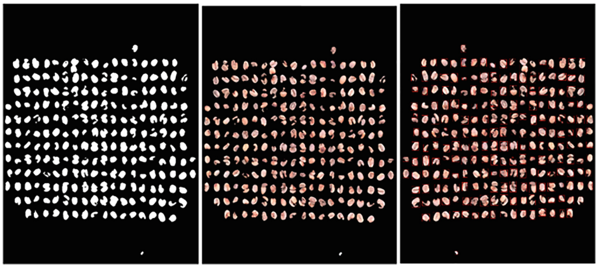

Fig. 6. **a**) Image resulting from the median filter process with a filter size of 11, **b**) Image from masking the original image with a binary image from the median filter, **c**) Image from the process of forming bounding rectangle/box to extract per coffee bean.

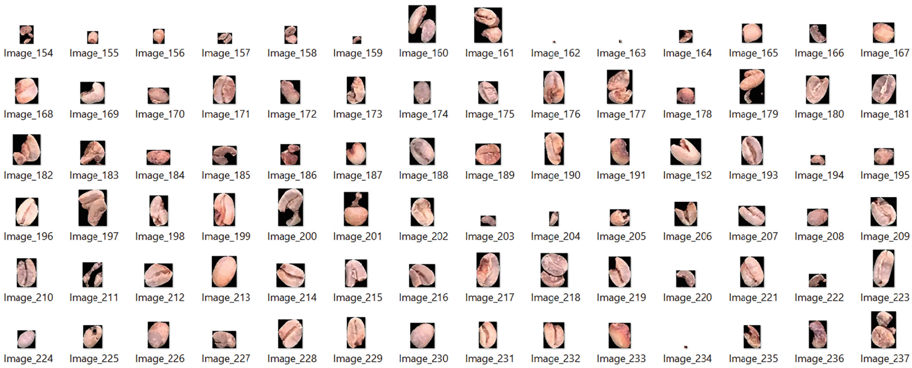

Fig. 7. Image Extracted image per coffee bean for defect type 2.

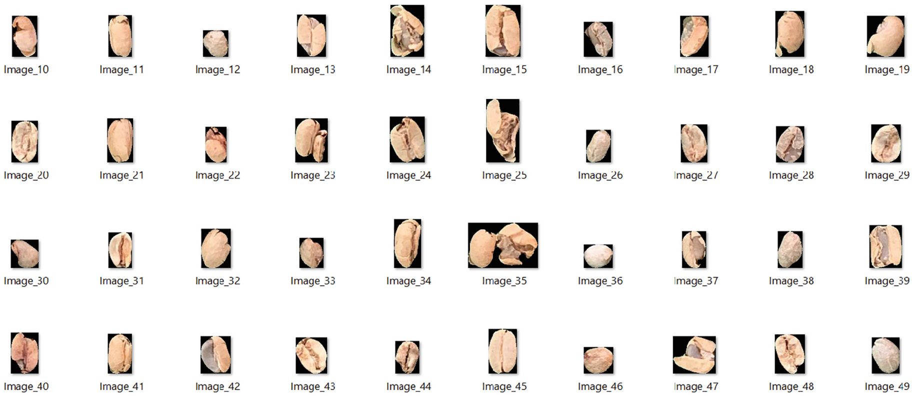

Fig. 8. Extracted image per coffee bean for defect type 9.

Table 3. Number of Extracted Image Samples of Coffee Bean Defects (SNI 01-2907-2008)

No	Type of Defect	Nb Image Capture 1		Nb Image Capture 2		Nb Total Image	
		Extracted	Valid	Extracted	Valid	Extracted	Valid
1	1 (one) full black bean	174	115	169	96	343	211
2	1 (one) partial black bean	243	208	258	216	501	424
3	1 (one) broken black bean	386	238	377	244	763	482
4	1 (one) pod/cherry	173	83	172	79	345	162
5	1 (one) brown bean	207	196	234	208	441	404
6	1 (one) large shell skin	283	184	259	187	542	371
7	1 (one) medium shell skin	403	252	384	213	797	465
8	1 (one) small shell skin	469	258	463	231	932	489
9	1 (one) parchment bean	56	52	56	49	112	101
10	1 (one) large husk	8	7	9	7	17	14
11	1 (one) medium husk	24	21	22	20	46	41
12	1 (one) small husk	-	-	-	-	-	-
13	1 (one) broken/chipped bean	120	106	112	99	232	205
14	1 (one) immature bean	150	128	165	150	315	278
15	1 (one) bean with slight insect damage	230	198	224	200	454	398
16	1 (one) bean with severe insect damage	204	183	200	151	404	334
17	1 (one) bean with fungus damage	54	48	54	47	108	95
18	1 (one) foreign matter (large size branch of stone)	29	15	37	18	66	33
19	1 (one) foreign matter (medium size branch of stone)	*join with no 18*		*join with no 18*		*join with no 18*	*join with no 18*
20	1 (one) foreign matter (small size branch of stone)	*join with no 18*		*join with no 18*		*join with no 18*	*join with no 18*

5 Conclusion

In this study, a semi-automatic ground truth image construction for coffee bean defects classification based on SNI 01-2907-2008 was presented. The results show that the proposed method is able to construct and to extract optimal number of the image samples of coffee bean for each type of defects. All extracted image samples of coffee bean defects will serve as a new dataset for the future automation system for determining the value of coffee bean defects based on SNI 01-2907-2008. In the next research phase, this new dataset will be used to build a deep learning-based training model to identify the types of defects in coffee beans. The resulting training model will later become the main engine of software development to determine the overall grade of coffee bean defects.

Acknowledgments. The authors thank the CV Kaki Lima and the Belantih Coffee Kintamani, in providing the coffee bean samples for this work.

Authors' Contributions. Made Windu Antara Kesiman designed and developed the algorithm for the extraction and classification of coffee bean defects. Ismail Sulaiman manually classifies coffee bean defect samples and validates the extracted and segmented images of coffee bean defects.

References

1. Badan Standardisasi Nasional (BSN), Standar Nasional Indonesia (SNI) 01-2907-2008 Biji Kopi, 2008
2. Hendro Kusumo, Sekilas Tentang Standar Nasional Indonesia: Biji kopi; Biji kakao; dan Rumput laut, Badan Standardisasi Nasional, 25 Oktober 2017
3. Ilhamsyah, Aviv Yuniar Rahman, Istiadi, Klasifikasi Kualitas Biji Kopi Menggunakan Multilayer Perceptron Berbasis Fitur Warna LCH, Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi, Vol. 5, No. 6, p.1009–1017, 2021
4. Juni Aramiko, Klasifikasi Kerusakan Biji Kopi Menggunakan Metode Naïve Bayes Classifier, Skripsi, Jurusan Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim, Malang, 2020
5. Mahmuda Saputra, Kusrini, Mei P Kurniawan, Identifikasi Mutu Biji Kopi Arabika Berdasarkan Cacat Dengan Teknik Convolutional Neural Network, Inspiration Jurnal Teknologi Informasi dan Komunikasi, Juni, 2020
6. Serawork Wallelign, Mihai Polceanu, Towfik Jemal, Cedric Buche, Coffee Grading with Convolutional Neural Networks Using Small Datasets with High Variance, Journal of WSCG, Vol. 27, No. 2, 2019, https://doi.org/10.24132/JWSCG.2019.27.2.4
7. Edwin R. Arboleda, Arnel C. Fajardo, Ruji P. Medina, An Image Processing Technique for Coffee Black Beans Identification, IEEE International Conference on Innovative Research and Development (ICIRD), 11–12 May 2018, Bangkok, Thailand
8. Joao Paulo Lebarck Pizzaia, Igor Rodrigues Salcides, Gustavo Maia de Almeida, Arabica coffee samples classification using a Multilayer Perceptron neural network, 13th IEEE International Conference on Industry Application, 2018
9. Christian E. Portugal-Zambrano, Juan C. Gutierrez-Caceres, Juan Ramirez-Ticona, Cesar A. Beltran-Castanon, Computer vision grading system for physical quality evaluation of green coffee beans, IEEE XLII Latin American Computing Conference (CLEI), 2016, DOI: https://doi.org/10.1109/CLEI.2016.7833383
10. Nen-Fu Huang, Dong-Lin Chou, Chia-An Lee, Real-Time Classification of Green Coffee Beans by Using a Convolutional Neural Network, IEEE 3rd International Conference on Imaging, Signal Processing and Communication, 2019
11. Andres Felipe Sanchez-Aguiar, Alberto Mario Ceballos-Arroyo, Albeiro Espinosa-Bedoya, Toward the recognition of non-defective coffee beans by means of digital image processing, IEEE XXII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), 2019