Available online www.jsaer.com

Journal of Scientific and Engineering Research, 2023, 10(9):18-22

Research Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Grading of Green Coffee Beans for Specialty Coffee using Image Processing Techniques

Chamika Kuruppuarachchi*, Kahawattage S.P. Amaratunga

Department of Agricultural Engineering, Faculty of Agriculture, University of Peradeniya.
Email: chamikakuruppu@gmail.com

Abstract Specialty coffee is considered the highest-grade coffee available in the market, and the grading process plays a crucial role in its production. Across the globe, manual selection is widely practiced for grading green coffee beans, and it is a labor-intensive task. However, image processing technology offers a potential solution for effective grading of green coffee beans based on their unique physical characteristics. This study aimed to train an algorithm based on image processing technology, in combination with a Raspberry Pi4 computer and a Pi camera, to identify specialty-grade beans of Coffea arabica. The dataset comprised 1,200 images of coffee beans and the algorithm was specifically trained utilizing OpenCV library (Open-Source Computer Vision Library – Version 4.5.5, developed by Intel) to grade individual beans under controlled artificial lighting conditions. The algorithm successfully rejected deformed beans, inert materials, blackened beans, broken beans, dried cherries, and undersized beans, apart from specialty-grade beans. The overall accuracy of the trained algorithm was 86.5% and Matthews Correlation Coefficient (MCC) value was 0.738.

Keywords Coffee grading, image processing, specialty coffee grading

1. Introduction

Specialty-grade coffee is considered the premium grade in the coffee industry. The quality of specialty-grade coffee mainly relies on its grading process. According to the International Coffee Organization, 37% of the total coffee production is specialty-grade coffee [1]. Compared to the total coffee production only a limited portion of specialty-grade coffee is produced due to the nature of the grading process.

Grading coffee beans for the specialty grade is a crucial operation. It is carried out in various stages in the coffee production process, but the critical grading is assessed after the milling process prior to roasting. Specialty Coffee Association of America has declared standards for bean characteristics and cup characteristics of specialty-grade coffee. Among those the bean characteristics are (1) no primary defects in coffee beans are allowed (full black, full sour, pod/cherry, large stones, medium stones, large sticks, medium sticks), (2) coffee beans should not have more than 5 secondary defects (parchment, hull/husk, broken/chipped, insect damage, partial black, partial sour, floater, shell, small stones, small sticks, water damage) in a sample of 300 g, and (3) coffee beans above or below the maximum of 5% screen size indicated is tolerated (screens 14, 15, 16, 17, and 18, are considered acceptable). Coffee beans which are fulfilling these requirements are considered specialty-grade coffee beans [2].

Throughout the history of the coffee industry, manual selection has been practiced for the grading of specialty-grade green coffee. The manual selection which is the process of selective hand-picking is a time-consuming and labor-intensive task. This leads to economic inefficiency in the coffee industry. To make this selection process economical and faster, image processing approaches can be utilized.

In recent years, image processing has emerged as a promising technology. This technology has found successful applications in various agricultural processes like grading, harvesting, weed detection, and disease identification [3], [4]. But this technology is not yet commercially developed for green coffee grading in the coffee industry. Currently, industrial-scale operations rely on conventional color sorters, which are only capable of identifying the colors of the coffee beans [5]. However, image processing has the potential to provide a more comprehensive assessment of coffee beans such as size, shape, presence of primary and secondary damages. Therefore, the objective of this study was to develop an automated coffee bean selection mechanism based on image processing techniques.

2. Materials and Methods

Collection of Coffee Samples

The study relied upon the images of coffee beans as the primary data source. To capture the images of coffee beans with relevant features, pre-graded coffee beans were purchased from a reputable commercial coffee vendor. All the coffee beans utilized for this study belonged to the variety of *Coffea arabica*, since this variety is preferred to produce premium grade specialty coffee. The physical appearance of specialty-grade beans as well as primary and secondary defects were the main consideration for the study. Representative samples (600 beans from each) from specialty grade and the beans with defects (primary and secondary) were used for image acquisition.

Image Acquisition

Captured images of real coffee beans were utilized for the study as the size is a crucial factor in the selection criteria of beans. The process of capturing coffee beans was carried out using a Raspberry Pi camera (Raspberry Pi foundation-UK, Version 1.2) connected to a Raspberry Pi computer (Raspberry Pi foundation-UK, Pi4 model B).

Lighting is a crucial factor in computer vision technology. Therefore, as illustrated in figure 1 a box equipped with artificial lights was utilized to maintain optimal illumination conditions during the image acquisition process. To facilitate the diffusion and reflection of light, the inner walls of the light box were completely covered with white colored sheets. The electrical setup of the lightbox included 5 W, 12 V warm white LED strips.

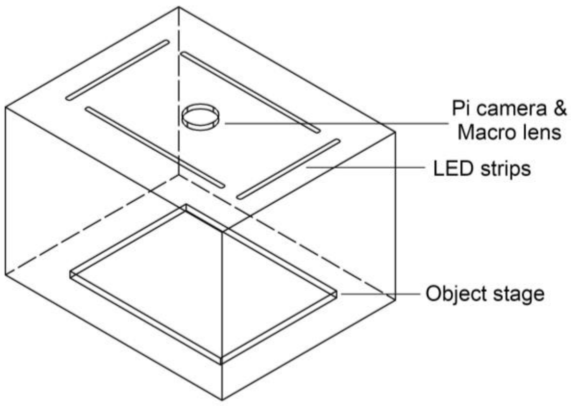

Figure 1: Artificial lightbox

The Pi camera was fixed with a macro lens at the top center of the artificial light box. The pi camera utilized for this study was a fixed focus module with a focal distance of 0.5 m. The focal length of the camera was effectively reduced up to 0.076 m by the macro lens In order to facilitate the acquisition of macroscopic images of coffee beans.

The object stage for keeping the beans for capturing the image consisted of a black surface. The intention of using a black surface was to minimize reflection of light and to contrast the coffee beans in photos. Throughout the image acquisition process, coffee beans were kept individually on the stage, and the relative position of the camera, and the stage were kept invariant.

A singular photograph of each coffee bean was captured as illustrated in figure 2, and the photos were in JPEG format with a resolution of 640x480 pixels. The complete set of images with 600 photos of specialty-grade coffee beans classified as positive images, and 600 photos of defects classified as negative images were generated using this setup.

As the individual coffee beans were captured under controlled artificial lighting conditions, specific pre-processing steps such as background removal, noise reduction, image segmentation, and color channel adjustments were not performed for the photos.

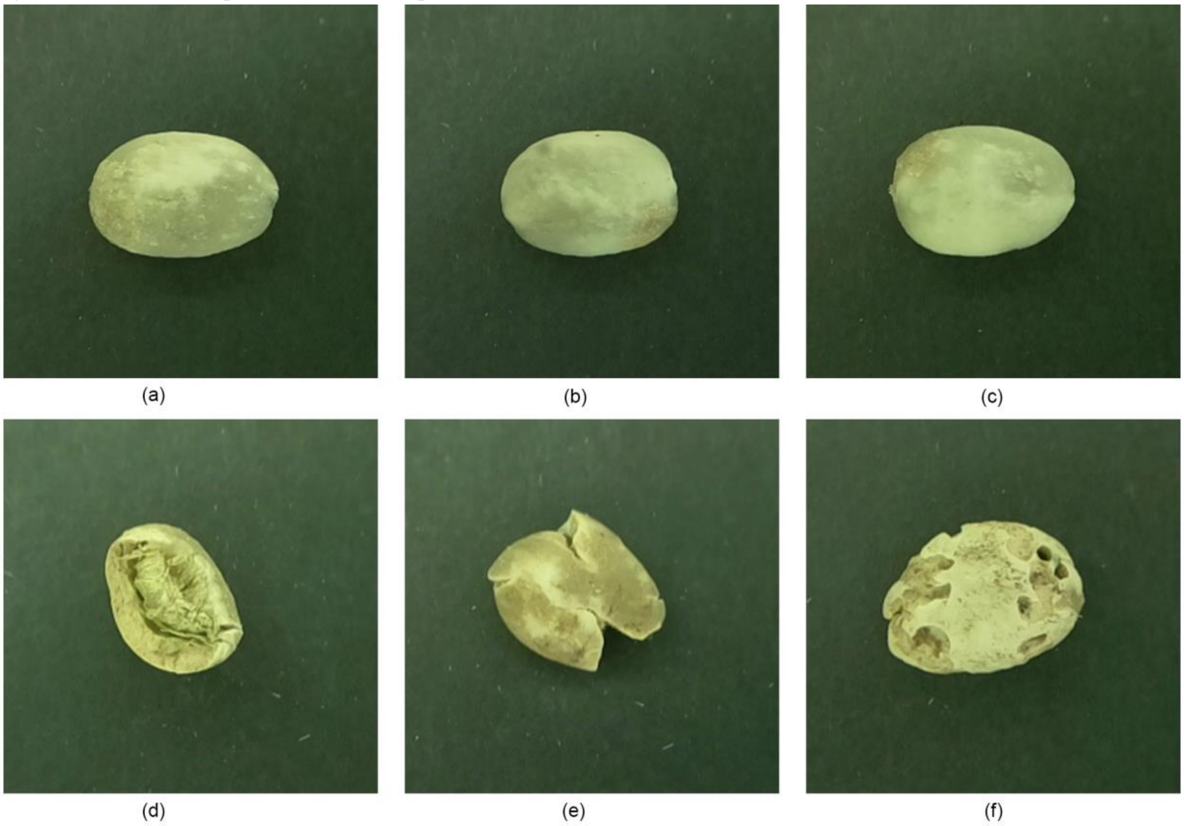

Figure 2: Specialty-grade beans: (a), (b), (c), and defective beans: (d), (e), (f)

Extraction of Features

The extraction of features of the positive images was carried out manually utilizing the "Annotation tool" offered by the OpenCV library (Open-Source Computer Vision Library – Version 4.5.5, developed by Intel). The annotation process demarcated the region of interest in a photo which contained the interested objects or features by drawing a bounding box on the image. Thereby it recorded the location of the specified objects or features in a photo. Bounding boxes were drawn for all the 600 positive images resulting in the creation of a description file saved in the "TXT" format. From the positive text file, a vector file was generated utilizing the "Create sample tool" provided by the OpenCV library for further processing. The description file for the negative images was created manually without the inclusion of annotations, also resulting in a "TXT" file. The text file generated from the negative images and the vector file generated from the positive images were utilized to train the model.

Training of the Algorithm

OpenCV library was selected as the primary tool for training the algorithm due to its advanced capabilities in image processing. OpenCV offers a wide range of functionalities and features that facilitate the training process. Specifically, OpenCV provides dedicated tools to simplify and optimize the training procedure.

The HAAR (Haar-like features – introduced by Alfréd Haar) Classifier principle was followed for training the algorithm to identify specialty-grade coffee beans [6]. The algorithm was trained utilizing the "Train cascade tool" provided by the OpenCV library, with the previously generated image data from the positive and negative samples. The training process is supervised learning consisting of multiple learning stages, and each stage represents a weak learner. The boosting technology which is "Adaboost Training" in the training process takes the weighted average of decisions made by each learning stage and develops a highly accurate binary classification (two possible classes for each observation as specialty grade or non-specialty grade) algorithm in "XML" (Extensible Markup Language) format [7]. The training process was conducted on a Core i3 computer with a 2.0 GHz processor and 8 GB RAM because of the high processing power compared to the Raspberry Pi computer.

Testing of the Algorithm

The previously generated "XML" file was transferred to the Raspberry Pi computer for testing purposes. Testing was carried out in the same artificial light box which was used to capture the images of coffee beans. For the testing a combination of 100 specialty-graded coffee beans and 100 defective beans was utilized. When the algorithm successfully identifies a specialty-grade coffee bean, a green color bounding box was drawn as illustrated in figure 3 with a caption of "Good bean".

Various iterations were performed by adjusting key parameters within the training process, including the number of stages, the number of positive samples, the number of negative samples, and other relevant parameters until the accuracy reached above 80%.

Figure 1: Identification of specialty grade bean

The performance of the trained algorithm was evaluated (eq 1 through 6) based on overall accuracy and Matthew's Correlation Coefficient (MCC) using true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) of the test results. The true positive rate, false positive rate, false negative rate, overall accuracy, and MCC were employed to determine the most optimal trained algorithm. MCC is a commonly used parameter that assesses the quality of binary classification models and provides a balanced evaluation of the algorithm's performance. The MCC ranges from -1 to +1, with +1 indicating a perfect prediction, 0 indicating no better than a random prediction, and -1 indicating complete disagreement between the algorithm's predictions and the actual outcomes [8].

$$True\ positive\ rate = \frac{TP}{TP+FN} \tag{1}$$

$$True\ negative\ rate = \frac{TN}{TN+FP} \tag{2}$$

$$False\ negative\ rate = \frac{FN}{FN+TP} \tag{3}$$

$$False\ positive\ rate = \frac{FP}{FP+TN} \tag{4}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \tag{5}$$

$$MCC = \frac{(TP\times TN)-(FP\times FN)}{\sqrt{(TP+FP)\times(TP+FN)\times(TN+FP)\times(TN+FN)}} \tag{6}$$

3. Results and Discussion

The test result of the optimal trained algorithm derived from a combination of 100 specialty-graded beans and 100 defective beans is presented in table 1. Based on the results, the algorithm's performance was evaluated using the true positive rate, false positive rate, false negative rate, overall accuracy, and MCC value as illustrated in table 2. The true positive rate denotes the algorithm's capability to identify the desired specialty-grade beans. Since the true positive rate was closer to 1, it indicated the algorithm's effectiveness in identifying specialty-grade beans. The true negative rate indicates the ability of the algorithm to correctly identify defective beans, obtaining a 0.94 true negative rate indicated that the trained algorithm can successfully reject defective coffee beans. Achieving a 0.06 false positive rate was considered favorable, as it indicated the results were towards specialty-grade standards and a minimal probability of including defective beans in the graded batch. The false negative rate quantifies the probability of incorrectly classifying specialty-grade beans as defective beans. With a false negative rate of 0.21, there was a probability of rejecting specialty-grade beans within a coffee bean sample. The overall accuracy measures how well the trained algorithm distinguishes specialty-grade beans and defective beans. Based on the results the optimal trained algorithm had an accuracy of 86.5%.

Table 1: Readings of the optimal trained algorithm

Number of True Positive Beans (TP)	Number of False Positive Beans (FP)	Number of True Negative Beans (TN)	Number of False Negative Beans (FN)
79	6	94	21

Table 2: Performance evaluations for the optimal trained algorithm

True Positive Rate	**True Negative Rate**	**False Positive Rate**	**False Negative Rate**	**Accuracy**	**MCC**
0.79	0.94	0.06	0.21	86.5%	0.738

The overall accuracy measures the ratio between the correctly identified samples and the overall sample. There is a possibility of over-optimistic estimation of the trained algorithm when the data set is unbalanced. But the MCC is unaffected by an unbalanced dataset, therefore the MCC value provides a comprehensive assessment of the trained algorithm compared to the overall accuracy [8]. Matthew's correlation coefficient value of 0.738 for the trained algorithm indicated a relatively good level of performance, and it performed significantly better than random guessing. It indicated that the algorithm's predictions were consistent and meaningful, with more true positives and true negatives than false positives and false negatives.

4. Conclusion

This research was conducted to develop an algorithm based on the identification of unique physical features exhibited by specialty-grade green coffee beans. The algorithm exhibited a true positive rate of 0.79, a true negative rate of 0.94, a false negative rate of 0.21, and a false positive rate of 0.06 with an overall accuracy of 86.5% and MCC value of 0.738. The performance of the trained algorithm demonstrated the potential for employing image processing techniques within the coffee industry to grade specialty-grade coffee beans. Further, this sorting technique can be improved to grade green coffee beans from Coffea canephora and Coffea liberica by training separate image processing algorithms.

References

[1]. CoffeeResearch, "Coffee History - CoffeeResearch.org," 1999. http://www.coffeeresearch.org/coffee/history.htm (accessed Sep. 12, 2022).

[2]. SCA, "Specialty Coffee Association," 2017. https://sca.coffee/ (accessed Sep. 12, 2022).

[3]. Y. He et al., "Method for Grade Identification of Tobacco Based on Machine Vision," Trans. ASABE, vol. 61, no. 5, pp. 1487–1495, 2018, doi: 10.13031/trans.12627.

[4]. S. Shajahan et al., "Identification and Counting of Soybean Aphids from Digital Images Using Shape Classification," Trans. ASABE, vol. 60, no. 5, pp. 1467–1477, 2017, doi: 10.13031/trans.12105.

[5]. E. R. Arboleda, A. C. Fajardo, and R. P. Medina, "Classification of coffee bean species using image processing, artificial neural network and K nearest neighbors," 2018 IEEE Int. Conf. Innov. Res. Dev. ICIRD 2018, pp. 1–5, Jun. 2018, doi: 10.1109/ICIRD.2018.8376326.

[6]. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 2001, vol. 1, pp. I-511-I–518. doi: 10.1109/CVPR.2001.990517.

[7]. S. Wu and H. Nagahashi, "Parameterized AdaBoost: Introducing a Parameter to Speed Up the Training of Real AdaBoost," IEEE Signal Process. Lett., vol. 21, no. 6, pp. 687–691, Jun. 2014, doi: 10.1109/LSP.2014.2313570.

[8]. D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," BMC Genomics, vol. 21, no. 1, p. 6, Dec. 2020, doi: 10.1186/s12864-019-6413-7.